SULFOLOBUS SOLFATARICUS AS SOURCE OF GLYCOSYL HYDROLASES WITH BIOTECHNOLOGICAL POTENTIAL

Michele Girfoglio

Dottorato in Scienze Biotecnologiche – XXI ciclo
Indirizzo Biotecnologie Industriali
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XXI ciclo
Indirizzo Biotecnologie Industriali
Università di Napoli Federico II

SULFOLOBUS SOLFATARICUS AS SOURCE OF GLYCOSYL HYDROLASES WITH BIOTECHNOLOGICAL POTENTIAL

Michele Girfoglio

Dottorando: Michele Girfoglio

Relatore: Prof. Mosè Rossi

Coordinatore: Prof. Giovanni Sannia

A tutte le persone a me care,
che mi hanno sempre sostenuto
ed hanno rappresentato un
fondamentale supporto durante
il percorso di cui questa tesi
rappresenta il compimento.
A Elia P.

INDEX

ABBREVIATIONS

bp Base pairs

BSA Bovine serum albuine

CAPS N-cyclohexyl-3-aminopropanesulfonic acid

CBD Cellulose binding domain

CPs Conjugative plasmids

Da dalton

DNA Deoxyribonucleic acid

dNTP Deoxyribonucleotide triphosphate

DSM Deutsche Sammlung von Mikroorganismen

EDTA Ethylenediaminetetraacetic acid

h Hours

HGT Horizontal gene transfer

hph Hygromycin B phosphotransferase

IPTG Isopropyl β-D-1-thiogalactopyranoside

IS Insertion sequence

LB Luria Bertani medium

min Minutes

MITEs Miniature inverted transposable elements

OD Optical density

ORF Open reading frame

PCR Polimerase chain reaction

PMSF Phenyl methane sulphonyl fluoride

PVDF Polyvinylidene difluoride

RNA Ribonucleic acid

SBP Sugar binding protein

SDS Sodium Dodecyl Sulphate

SDS-PAGE Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis

TBE 45 mM Tris-borate, 1mM EDTA pH 8.0

TBP TATA-binding protein

TLC Thin Layer Chromatography

Tris tris(hydroxymethyl)aminomethane

TYS tryptone yeast sucrose (medium)

UTR untraslated region

X-gal 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

YPG Yeast extract Peptone Galactose

Riassunto

Introduzione

Uno dei principali ambiti di interesse delle Biotecnologie Industriali riguarda gli enzimi sintetizzati dagli organismi estremofili, in particolar modo ipertermofili; l'attenzione con cui il mondo scientifico guarda a tali proteine è motivata dalle doti di stabilità e resistenza alle alte temperature che mostrano caratteristiche peculiari che le distinguono dalle controparti da organismi mesofili. In tale contesto un ruolo di primo piano è rivestito dagli enzimi che prendono parte al processo degradativo di carboidrati, i quali possiedono un ampio spettro di campi applicativi che vanno dall'industria di produzione della carta a quella alimentare, da quella tessile a quella farmaceutica; in particolare, grande risalto negli ultimi anni sta assumendo l'applicazione che vede tali enzimi impiegati nel processo di produzione del bioetanolo, carburante il cui crescente utilizzo sta assumendo ritmi esponenziali. La produzione e l'impiego di tale carburante risulterebbe particolarmente vantaggiosa se questo venisse ottenuto utilizzando come materia prima scarti industriali ad alto contenuto cellulosico, evitando di impiegare a tal scopo cereali appositamente ed intensivamente coltivati; si potrebbe in tal modo superare uno dei maggiori problemi legati alla produzione di biocarburanti, che riguarda l'aumento dei prezzi delle materie prime alimentari ed il conseguente aggravamento delle condizioni di vita nelle aree in via di sviluppo.
A tale riguardo, diversi microrganismi ipertermofili eterotrofi sono stati studiati rispetto alla loro capacità di crescere su carboidrati complessi come unica fonte di carbonio e di energia. Tale caratteristica può essere particolarmente vantaggiosa per questi organismi, in quanto spesso il loro habitat risiede in nicchie ecologiche dove le fonti di carbonio, di origine vegetale, sono scarse.
La maggior parte degli Archaea mostra un eccezionale adattamento alla vita in condizioni estreme, tra cui l'alta temperatura e i valori di pH estremi; alla luce di ciò, essi rappresentano una fonte di biocatalizzatori particolarmente adatti ad essere impiegati per applicazioni industriali nell'idrolisi di matrici complesse di zuccheri. Fino a pochi decenni fa gli Archaea non erano riconosciuti dal punto di vista della classificazione filogenetica degli esseri viventi. Guardando un "albero della vita" infatti, avremmo potuto notare esclusivamente due domini, quello dei Procarioti (identificati con i Batteri) e quello degli Eucarioti. Una fondamentale innovazione in questo contesto è scaturita dagli studi di Carl Woese che, alla fine degli anni '70, dopo aver condotto approfondite analisi comparative di sequenze di rRNA 16S procariotiche allo scopo di delineare un più chiaro quadro delle correlazioni intergenere nel mondo batterico, giunse alla conclusione che il mondo dei viventi fosse da suddividere in tre domini. Nacque così la denominazione di "ArchaeoBatteri", creata per definire il nuovo dominio rispetto agli EuBatteri, poi modificata in Archaea per affermarne definitivamente la separazione dal dominio dei Batteri. Facendo quindi cadere l'idea dell'unicità procariotica, Woese distinse i Procarioti in due gruppi, gli Archaea appunto, ed i Bacteria. I suoi studi avevano messo in evidenza che questi microrganismi si differenziano talmente dai Batteri da meritare appieno una classificazione filogenetica indipendente. Gli Archaea risultano attualmente ulteriormente suddivisi in tre regni: Crenarchaeota (ipertermofili e termoacidofili); Euryarchaeota (ipertermofili, metanogeni ed alofili); Korarchaeota (poco conosciuti a livello biochimico e fisiologico).

In questo contesto *S. solfataricus*, isolato originariamente dalle pozze solfatariche di Pisciarelli a Napoli, e che cresce ad un pH acido (in un intervallo tra 2.0 e 5.0) e ad alte temperature (75-90°C), rappresenta un organismo modello per lo studio dei Crenarchaeota, in quanto risulta essere facilmente propagabile in coltura liquida così come in mezzo solido. Inoltre, risulta rilevante segnalare che il genoma di *S. solfataricus* P2 è stato uno dei primi genomi crenarchaeali ad essere completamente sequenziato, e ciò ha aperto nuove prospettive di indagine riguardo tali organismi, dando un forte impulso allo studio delle loro caratteristiche.
Le specie appartenenti al genere *Sulfolobus*, come pure *S. solfataricus*, sono provviste di un'ampia gamma di diversi elementi genetici, comprendenti virus, plasmidi, sequenze d'inserzione, che costituiscono un potente mezzo di indagine filogenetica per gli organismi nei quali si ritrovano. Inoltre, tali elementi, in particolar modo virus e plasmidi, vengono sempre con maggiore interesse studiati nell'ottica di poterli adoperare come base nella messa a punto di sistemi stabili di clonaggio/espressione, i quali, per gli organismi archaeali, risultano essere ancora ad uno stato incipiente di sviluppo.
Inoltre, *S. solfataricus* è una interessante fonte di attività enzimatiche di tipo cellulasico; tali enzimi potrebbero essere utilizzati in maniera vantaggiosa per la produzione di bioetanolo, grazie alla loro peculiare combinazione di termostabilità e resistenza a pH acidi. Infatti, nel processo di produzione del bioetanolo i materiali cellulosici vengono idrolizzati impiegando acidi ad alte temperature, ed in tale contesto operativo la conduzione di idrolisi enzimatica con enzimi resistenti ad alte temperature e bassi pH porta ad una semplificazione del processo produttivo, con un conseguente abbassamento dei costi.
Il sequenziamento del genoma di *S. solfataricus* ha mostrato l'esistenza di un completo gruppo di ventidue glicosil idrolasi capaci di processare carboidrati complessi quali l'amido, la cellulosa e le emicellulose, compresi mannani, xiloglucani e xilano. In particolare, sono state identificate tre sequenze codificanti per putative β-endoglucanasi (SSO2534, SSO1354 ed SSO1949), sulle quali si è concentrata l'attenzione nello svolgimento del presente progetto; inoltre in *S. solfataricus* è stata rilevata un'attività β -endoglucanasica extracellulare, sia liberamente rilasciata che adesa alle cellule.
Per quanto riguarda SSO2534, il prodotto proteico è stato ampiamente caratterizzato, anche se i tentativi di espressione in *E. coli* non hanno avuto successo; il gene *sso1949* invece è stato clonato ed espresso in *E. coli* , ma a livelli molto bassi.
Alla luce di ciò, al fine di ottenere la (sovra)espressione di sequenze codificanti per attività enzimatiche provenienti da *S. solfataricus*, risulta utile intraprendere la strada della sovra-espressione di geni nello stesso organismo di origine o in uno fisiologicamente più correlato, considerati i diversi fattori limitanti che si possono presentare esprimendo tali sequenze in maniera eterologa, in ospiti filogeneticamente troppo distanti. Si comprende quindi l'importanza di disporre di efficaci sistemi-vettore per l'espressione xenologa-eterologa in *S. solfataricus*, anche se il numero dei sistemi ad oggi sviluppati e che sono risultati efficaci risulta essere ancora scarso.
Nel laboratorio dove si svolge tale progetto di Dottorato è stato progettato e messo a punto un efficiente vettore navetta *E. coli-S. solfataricus*, definito pMSSV, costruito sulla base del plasmide archaeale pSSVx da *S. shibatae* REY 15/4, al quale sono state aggiunte sequenze derivanti dal plasmide pUC19.

Il pSSVx è un elemento genetico isolato, insieme al fusellovirus SSV2, dal ceppo REY 15/4 di *S. islandicus*. Il genoma è formato da 5705 bp e mostra regioni di alta similarità di sequenza con i plasmidi criptici pRN1, pRN2 e pDL10; per questo viene considerato appartenente alla famiglia dei pRNs. All'interno del genoma del pSSVx si trovano nove ORFs delle quali tutte, tranne una, sono orientate nella stessa direzione. Quattro di queste, insieme a due elementi di sequenza, mostrano similarità con gli altri plasmidi della famiglia. In particolare due delle quattro ORFs sono quelle che codificano per la proteina Rep (ORF892) e per la CopG (ORF60), mentre le due sequenze sono costituite dalle origini di replicazione a singolo e doppio filamento (sso e dso). In aggiunta a ciò, è presente una regione che non mostra similarità con i pRNs; questa contiene due ORFs adiacenti, 154 e 288, per le quali è possibile trovare regioni codificanti omologhe nei fusellovirus SSV1 (a153 e b251) e SSV2 (153 e 233); queste sequenze si strutturano tutte con simile arrangiamento (è presente una regione di sovrapposizione tra le due ORFs). La similarità risulta essere maggiore per la ORF 154, e le similarità mostrate dai virus tra loro sono maggiori di quelle tra virus e plasmide.
L'individuazione del pSSVx si è avuta durante una procedura di screening volta a identificare elementi genetici extracromosomali, specialmente del tipo SSV1. Dopo aver effettuato analisi di restrizione, volte ad accertare la presenza di DNA episomale, attraverso indagini di microscopia elettronica sui supernatanti cellulari sono state trovate particelle virali, dalla caratteristica forma a "limone", di due distinte dimensioni: le più grandi mostravano grandezza uguale alle particelle di SSV1, ovvero 80nm-55nm, mentre le più piccole si attestavano su dimensioni di 60nm-40nm. Il genoma di SSV2, una volta clonato, produceva soltanto le particelle di dimensioni maggiori, all'interno delle quali si trovava impacchettato il DNA virale. Dal confronto tra i pattern di restrizione del DNA del solo SSV2 e di quello derivante dalla miscela di particelle, e da analisi di cross-ibridazione, si poteva evincere che nella miscela era presente un ulteriore elemento genetico della grandezza di circa 5.7 kb. Ciò forniva evidenza del fatto che le particelle SSVx più piccole potessero contenere tale elemento, chiamato appunto pSSVx. Risulta interessante notare che il pSSVx risulta propagare efficacemente in mancanza di pressione selettiva solo se co-trasfettato con SSV2; tale plasmide non è capace infatti di impacchettarsi e dar luogo ad un processo infettivo autonomamente, come testimoniato dalla constatazione che risulta essere privo delle regioni codificanti per le proteine strutturali del virus. In quest'ottica, le due ORFs che il pSSVx mostra nella regione non conservata, e che hanno omologia con le sequenze degli SSV1 e SSV2, sembrano codificare per proteine che sono implicate nel processo di assemblaggio delle particelle virali, e con buona probabilità sono state acquisite dallo stesso genoma di SSV2. Risultano essere essenziali per l'impacchettamento e la propagazione del plasmide, in quanto gli elementi pRN1 e pRN2, che ne sono privi, non mostrano la capacità di propagarsi sfruttando come *helper* SSV1 o SSV2. Alla luce delle affermazioni fin qui riportate, risulta chiara la comprensione dell'appellativo dato a tale elemento, che viene sovente definito come un "ibrido" tra un plasmide ed un virus.
Il vettore pMSSV ricalca quindi le caratteristiche fondamentali del plasmide pSSVx da cui prende origine; si mantiene in alto numero di copie in entrambi gli ospiti, e si è dimostrato efficace nel condurre all'espressione del gene *lacS* di *S. solfataricus*, codificante per una β-glicosidasi, sotto il controllo del promotore del gene *tf55α*, codificante per una subunità del complesso-termosoma di *S. shibatae* e quindi inducibile da shock termico.

Risultati e discussione

Allo scopo di delucidarne le potenzialità applicative, è stata effettuata una approfondita caratterizzazione del vettore pMSSV. A tale scopo, la cassetta d'espressione costituita dalla sequenza codificante per una β-glicosidasi da *S. solfataricus* (*lacS*) sotto il controllo del promotore inducibile del gene *tf55α* è stata clonata all'interno del vettore, e con tale elemento sono state elettroporate cellule del ceppo GθW/SSV2 di *S. solfataricus*. L'utilizzo del gene *lacS* per il clonaggio è stato adottato in quanto comporta il vantaggio di permettere un facile monitoraggio dell'espressione; è possibile infatti effettuare la rilevazione dell'attività su centrifugati cellulari utilizzando il substrato cromogenico X-gal, ed avendo trasferito il costrutto in cellule del ceppo GθW (che presenta a livello del genoma un'estesa delezione che interessa il locus del gene *lacS*, e per tale motivo risulta essere difettivo dell'attività β-galattosidasica), la presenza di attività può essere senza incertezza attribuita all'espressione della sequenza contenuta nel vettore. Il vettore messo a punto è risultato capace di portare con successo all'espressione del gene inseritovi, riuscendo a complementare in questo caso la mutazione presente a livello del locus del gene *lacS*. Rispetto ai sistemi genetici *E. coli-S. solfataricus* finora sviluppati presenta diverse caratteristiche vantaggiose, tra le quali spiccano un'elevata efficienza di trasformazione, dovuta alla modalità propagativa di natura virale utilizzata dal vettore, e la stabilità nella propagazione, senza integrazione nel cromosoma dell'ospite e con un elevato numero di copie per cellula (fino a circa 130 in fase stazionaria, con accumulo in dipendenza dalla fase di crescita). Il pMSSV rappresenta quindi un potente strumento genetico, con forti potenzialità applicative tanto nello studio di base quanto nella produzione biotecnologica di proteine omologhe/eterologhe.

Il gene codificante per la putativa endoglucanasi extracellulare SSO1354, l'unico per il quale non sono presenti in letteratura studi di caratterizzazione sul prodotto proteico, è stato quindi impiegato in tentativi di espressione xenologa/eterologa. Inizialmente, la sequenza proteica corrispondente ad SSO1354 è stata sottoposta ad analisi bioinformatiche; da un allineamento con le sequenze presenti nelle banche dati, utilizzando il programma FASTA 3, si è potuto constatare che la più significativa similarità si evince ripetto alla sequenza della sopracitata SSO1949, con cui SSO1354 mostra una percentuale di identità dell'85%. Inoltre, SSO1354 mostra un allineamento significativo con altre endoglucanasi, la maggior parte delle quali risultano appartenere a Batteri termofili del genere *Thermotoga* (in particolare di specie *maritima* e *neapolitana*); escludendo SSO1949, l'allineamento con altre endoglucanasi è possibile però solo per il dominio catalitico, mentre per quello N-terminale non è possibile trovare riscontri positivi.

Particolarmente interessante risulta inoltre il raffronto della sequenza di SSO1354 con quelle delle *sugar binding proteins* (SBP) di *S. solfataricus* AraS, GlcS e TreS; queste proteine extracellulari, che legano rispettivamente l'arabinosio, il glucosio e il trealosio, sono delle subunità dei trasportatori *ATP binding cassette* (ABC), i quali sono responsabili per l'assimilazione degli zuccheri da parte di *S. solfataricus*. Dall'allineamento delle sequenze proteiche, che è particolarmente significativo all'estremità N-terminale, è stato possibile notare come SSO1354 abbia un'organizzazione in domini simile a quella delle SBPs; una regione N-terminale

idrofobica è seguita da una regione ricca in serine e treonine, con il dominio di catalitisi/legame posizionato all'estremità C-terminale. Diversamente da SSO1354, le SBPs possiedono un ulteriore dominio idrofobico all'estremità C-terminale. In AraS, GlcS e TreS la regione idrofobica serve per ancorare la proteina alla membrana, mentre quella ricca in Ser/Thr assume la funzione di linker flessibile tra il dominio di ancoraggio e quello di legame agli zuccheri. Considerata la significatività dell'allineamento nella regione N-terminale, è possibile ipotizzare che in SSO1354 tali domini assolvano alle stesse funzioni; SSO1354 potrebbe essere quindi con buona probabilità una proteina legata alla membrana. In quest'ottica, la co-localizzazione di una cellulasi e delle proteine coinvolte nell'assimilazione degli zuccheri potrebbe rappresentare un vantaggio in una strategia volta ad ottimizzare l'internalizzazione ed il metabolismo degli zuccheri, soprattutto per organismi che popolano habitat dove i carboidrati complessi rappresentano la maggiore, seppur scarsa, fonte di nutrimento.
Per quanto riguarda l'espressione del gene *sso1354*, questa è stata tentata in prima istanza per via xenologa, considerando le difficoltà riportate in letteratura nell'espressione delle cellulasi da *S. solfataricus* in un ospite mesofilo tradizionale come *E. coli*. La sequenza di SSO1354 è stata quindi utilizzata per la costruzione di una cassetta di espressione, nella quale tale gene è stato posto sotto il controllo del promotore del gene *araS*, forte ed inducibile da arabinosio. La cassetta di espressione *araS*1354 è stata inserita nel vettore pMSSV, con il quale poi sono state trasformate cellule di *S. solfataricus* Gθ-white. Tale ceppo è stato in questo caso scelto in quanto presenta il vantaggio di minori tempi di duplicazione rispetto al ceppo wild-type. I risultati dei tentativi di espressione hanno permesso di riscontrare la produzione della proteina ricombinante; questa però, inaspettatamente, è stata trovata nel mezzo di coltura in frazione significativamente maggioritaria. Comunque, il frazionamento delle proteine su gradiente di cesio ha consentito di localizzare la proteina esclusivamente in bande corrispondenti a complessi macromolecolari lipidi-proteine, tipicamente derivati dalla autolisi e quindi rottura delle membrane cellulari. Quindi il rilascio dell'enzima dalle cellule, che erano state tenute in crescita in fase stazionaria per tempo prolungato, è spiegabile con una lisi cellulare piuttosto che con un tipico "sorting" molecolare nel mezzo di coltura; la mancanza di attività cellulasica negli estratti di membrana può essere invece dovuta ai trattamenti di solubilizzazione di quest'ultima, che possono aver portato ad una inattivazione dell'enzima. La proteina è risultata attiva ad alta temperatura e pH basso (pH 2, 80°C); inoltre, un peso molecolare apparente di 45kDa su SDS-PAGE, contro un peso teorico di 37.2 kDa, indicava una probabile glicosilazione della proteina, dato coerente con quanto riscontrato dalle analisi bioinformatiche, che identificavano 11 siti ipotetici di glicosilazione. L'ottimizzazione della produzione xenologa di SSO1354 in *S. solfataricus* non ha raggiunto i risultati sperati; il costrutto pMSSV*araS*1354 si è rivelato instabile quando le cellule venivano cresciute in terreno minimo utilizzando l'arabinosio come fonte di carbonio, condizione necessaria per ottenere una completa risposta induttiva del promotore. Sono a tutt'oggi in corso esperimenti volti alla stabilizzazione genetica del vettore con risultati molto promettenti. La nuova strategia è basata sull'inserimento di un marcatore genetico di selezione nel vettore e l'uso di un diverso ospite, ovvero il ceppo parentale *S. islandicus*, curato, ovvero privo, di entrambi gli elementi SSV2 e pSSVX, o del solo pSSVX.
Nell'ottica di diversificare l'approccio all'espressione del gene *sso1354*, è stata successivamente intrapresa la strada dell'espressione in ospite mesofilo, inizialmente *E. coli*. I tentativi iniziali di esprimere la proteina nella sua intera

lunghezza sono risultati fallimentari; la proteina quindi è stata successivamente clonata nel vettore di espressione pET28c a partire dal 28° codone, in modo da eliminare la regione idrofobica N-terminale probabilmente responsabile dell'ancoraggio alla membrana. Il vettore è stato utilizzato per trasformare due ceppi differenti di *E. coli*: BL21(DE3) ed il ceppo denominato *arctic express*; quest'ultimo possiede la peculiarità di permettere la crescita a basse temperature, riducendo i rischi che la resa dell'espressione possa essere compromessa dall'aggregazione dalla proteina ricombinante, dalla formazione di corpi di inclusione e/o dalla degradazione da parte delle proteasi indotte in condizioni di overespressione. In tal modo, è stata ottenuta la produzione della SSO1354 in conformazione cataliticamente attiva in entrambi i ceppi, con risultati paragonabili. La cellulasi espressa in *E. coli* ha mostrato, così come rilevato per quella espressa in *S. solfataricus*, una preferenza per basso pH (2.0) ed alta temperatura (80°C). Tale proteina è risultata però instabile e contraddistinta da una bassa attività specifica.
Il lievito *Kluyveromyces lactis* è stato quindi scelto come ospite per effettuare ulteriori tentativi di espressione eterologa del gene *sso1354*. Tale microrganismo è attualmente impiegato nel campo delle biotecnologie per la produzione di diverse proteine eterologhe in scala industriale, e possiede infatti numerose caratteristiche che lo rendono particolarmente interessante per essere utilizzato a tale scopo: la manipolazione genetica è facile, possiede un apparato secretore versatile ed efficiente ed inoltre le proteine prodotte da *K. lactis* posseggono la certificazione GRAS (Generally Regarded As Safe) dell'FDA (Food and Drug Administration). Il gene *sso1354* è stato quindi adeguatamente clonato in un vettore integrativo (pKLAC1), e la sequenza è stata inserita in modo che venisse espressa come proteina di fusione con il peptide leader di secrezione del fattore α-MF (α-Mating Factor) di *K. lactis*. Tale peptide è necessario affinché il prodotto proteico venga correttamente secreto e venga rimosso dalle proteasi dell'ospite, garantendo la secrezione di una proteina nativa. Con il vettore così costruito sono state trasformate cellule di *K. lactis* GG799, un ceppo che non ha autotrofie, cresce rapidamente e raggiunge alte densità cellulari.
I risultati dell'espressione mostrano la produzione di proteina in forma attiva, ma che presenta una migrazione su gel SDS anomala, probabilmente attribuibile ad una iper-glicosilazione del prodotto proteico; tale dato sperimentale è coerente con dati riportati in letteratura circa i pattern di glicosilazione di proteine espresse in maniera eterologa in lievito. Inoltre, mentre la temperatura ottimale di attività (80°C) risulta essere in linea con i dati precedentemente acquisiti, il pH ottimale (pH 5) si discosta da quello atteso; inoltre, la proteina mostra una termostabilità di molto incrementata (emivita di 3h a 90°C) rispetto sia alla stessa SSO1354 prodotta in *E. coli* che a SSO1949. Entrambi gli effetti, sul pH e sulla termostabilità, potrebbero essere dovuti alla iperglicosilazione della proteina; in particolare, in letteratura è documentato l'effetto della glicosilazione nell'incrementare la resistenza delle proteine alle condizioni estreme; ciò potrebbe rappresentare infatti un modo che gli organismi (iper)termofili hanno evoluto per fornire al proprio corredo di proteine extracellulari un grado di extra stabilità.
La cellulasi espressa in lievito si è dimostrata capace di mantenere l'attività catalitica persino quando è stato utilizzato come mezzo di reazione un composto facente parte di una classe di solventi di nuova generazione, denominati *ionic liquids*. Tali sostanze, non acquose ma polari, sono completamente composte di ioni, e si trovano allo stato liquido a temperatura ambiente; posseggono inoltre determinate caratteristiche che hanno attirato l'interesse sempre maggiore del mondo scientifico

e produttivo: hanno una tensione di vapore prossima allo zero, sono termostabili ed hanno proprietà chimico-fisiche grandemente modificabili attraverso la sostituzione appropriata del catione e/o dell' anione, quali la polarità, l'idrofobicità e la miscibilità in altri solventi. Tali caratteristiche, abbinate ad un basso impatto ambientale, fanno degli *ionic liquids* una promettente classe di sostanze, che risultano estremamente vantaggiose da utilizzare in particolar modo per reazioni biocatalitiche. Determinati *ionic liquids* si sono dimostrati capaci di solubilizzare efficacemente la cellulosa, difficile da rendere solubile in solventi acquosi; in questo contesto, si comprende l'importanza nell'ottenere degli enzimi ad attività cellulasica che siano attivi in tali solventi. Gli *ionic liquids* potrebbero infatti essere utilizzati con successo in approcci volti ad impiegare materiali di scarto ad alto contenuto in cellulosa come materia prima nella produzione di biocarburanti in modo ecosostenibile. Ad oggi, la cellulasi SSO1354 è l'unico esempio di glicosil idrolasi in grado di esplicare la sua attività catalitica in questo tipo di solventi.

I risultati circa la caratterizzazione di SSO1354 confermano l'ipotesi del grande potenziale che tale enzima possiede dal punto di vista biotecnologico: risulta essere un candidato ideale da impiegare nei processi di degradazione della cellulosa, in particolar modo allo scopo di produrre biocarburanti. Inoltre, dal confronto delle proprietà della cellulasi espressa nei vari ospiti è possibile trarre interessanti conclusioni circa l'impatto della glicosilazione sulle proprietà chimico-fisiche e di attività dell'enzima; in particolare, risulta interessante dal punto di vista biotecnologico notare come la forma iperglicosilata della proteina mostri un forte incremento nella termoresistenza. Inoltre, il presente studio ha permesso di acquisire nuovi elementi circa la relazione tra struttura e funzione di SSO1354: sulla base dei dati derivanti dal confronto di sequenze e di quelli presenti in letteratura circa la caratterizzazione dell'attività cellulasica di *S. solfataricus* in vivo, e a differenza di quanto precedentemente ipotizzato per la cellulasi SSO1949, è stato possibile affermare che SSO1354 è un enzima legato alla membrana, e non rilasciato nel mezzo extracellulare. Tale ipotesi apre nuovi ed interessanti scenari di studio circa la strategia secondo la quale un ipertermofilo come *S. solfataricus* coordina ed organizza le attività coinvolte nell'utilizzo dei nutrienti.

Summary

The present PhD work was focused on the expression and characterization of the *Sulfolobus solfataricus sso1354* gene encoding a putative cellulase, which has not been characterized yet.
In this study, the results about the characterization of the *S. solfataricus* expression vector pMSSV are also reported. This plasmid was also able to propagate in *Escherichia coli* cells and was used to express the cellulase in a xenologous fashion.
This element had been previously constructed on the base of the pSSVx genetic element from *Sulfolobus islandicus* REY15/4, which is a hybrid between a plasmid and a fusellovirus, able to be maintained in nonintegrative form and to spread when the helper SSV2 virus is present in the cells. An expression cassette carrying *lacS* gene encoding a β-glycosidase from *S. solfataricus* under the control of the *Sulfolobus* chaperonin (thermosome tf55) heat shock promoter was inserted into pMSSV vector. Using the *lacS* gene as a genetic marker the vector could be characterized, in order to elucidate the potentiality of the genetic element as a gene transfer/expression vehicle. The results obtained pointed out some interesting features of this genetic system, namely it showed to be a powerful genetic tool, with a wide potential both in basic research and in the applicative field of biotechnology. In fact, the vector was able to spread efficiently through infected *S. solfataricus* cells as a virus, overcoming the usual low transformation efficiency of *Sulfolobus* cells. Moreover, the vector stably transformed *S. solfataricus* and propagated at high copy number with no rearrangement, recombination or integration into the host chromosome. The pMSSV vector was also shown to be an efficient cloning vehicle for the expression of passenger genes in *S. solfataricus*; in fact the genetic element was also able to drive the expression of the *lacS* sequence, leading to the production of a functional β-glycosidase enzyme.
The pMSSV vector was used to attempt the xenologous production of the recombinant SSO1354 enzyme from *S. solfataricus*. The *sso1354* gene was cloned under the control of the arabinose inducible *araS* promoter, and expressed into *S. solfataricus* cells. Moreover, the cellulase was expressed also in the mesophilic hosts *E. coli* and *Kluyveromyces lactis*, with the aim to optimize the yield of the active recombinant enzyme.
The comparison of the SSO1354 protein expressed in the different hosts have pointed out some interesting differences: when expressed in *E. Coli*, the cellulase showed to be unstable, whereas the recombinant counterpart produced by *K. lactis* was hyperglycosilated and was endowed with high thermostability (half life of 3h at 90°). These data confirm the theory about the glycosylation as an important stabilizing effector for proteins. Moreover, the SSO1354 from yeast showed a pH optimum (5.0) higher than SSO1354 from *E. coli* and the closely related *S. solfataricus* SSO1949 cellulase, indicating that the hyperglycosilation also affects the pH dependence of the enzyme.
The characterization of the SSO1354 cellulase revealed also its significant potential in biotechnological applications. Particularly, this is the first cellulase to date able to work in *ionic liquids*. These “molten salts” represent a new class of solvents which have a great potential as media for several biocatalytical reactions and are involved in promising strategies for the eco-sustainable production of bioethanol.
Moreover, new structure-function relationship was assigned to cellulases from *S. solfataricus* using SSO1354 sequence as a model. The alignment of this cellulase

sequence with those of *S. solfataricus* sugar binding proteins (SBP) and the characterization *in vivo* of the overall cellulase activity in *S. solfataricus* allowed to determine the cell localization of the SSO1354 protein as an extracellular and cell-surface associated cellulase. In fact, it was shown to be not free released in the culture media but anchored at the membrane most probably by its hydrophobic N-terminal region. This finding opens new and intriguing perspectives on the strategies adopted by a hyperthermophilic Archaeon, such as *S. solfataricus*, to coordinate and organize the activities involved in the sugar uptake and metabolism.

Introduction

Archaea

The Archaea domain has been distinguished in the phylogenetic tree only on late 1970s. Until then, all living organisms were divided between the prokaryotic bacteria and the four eukaryotic kingdoms. A revolutionary finding came from the studies of Dr Carl Woese on the relationship among prokaryotes; Woese and his colleagues identified this new group of prokaryotic microorganisms, by identifying their unusual small ribosomal RNA sequences. They differ significantly from the typical bacterial 16S rRNA sequences and hence required to be clustered together into an independent group, fairly distant also from eukaryotes. Therefore Woese proposed that life can be divided into three domains: Eukaryota, Eubacteria, and the new Archaebacteria. He later decided that the term Archaebacteria was a misnomer, and shortened it to Archaea to remark the separation with Eubacteria (Fig 1).

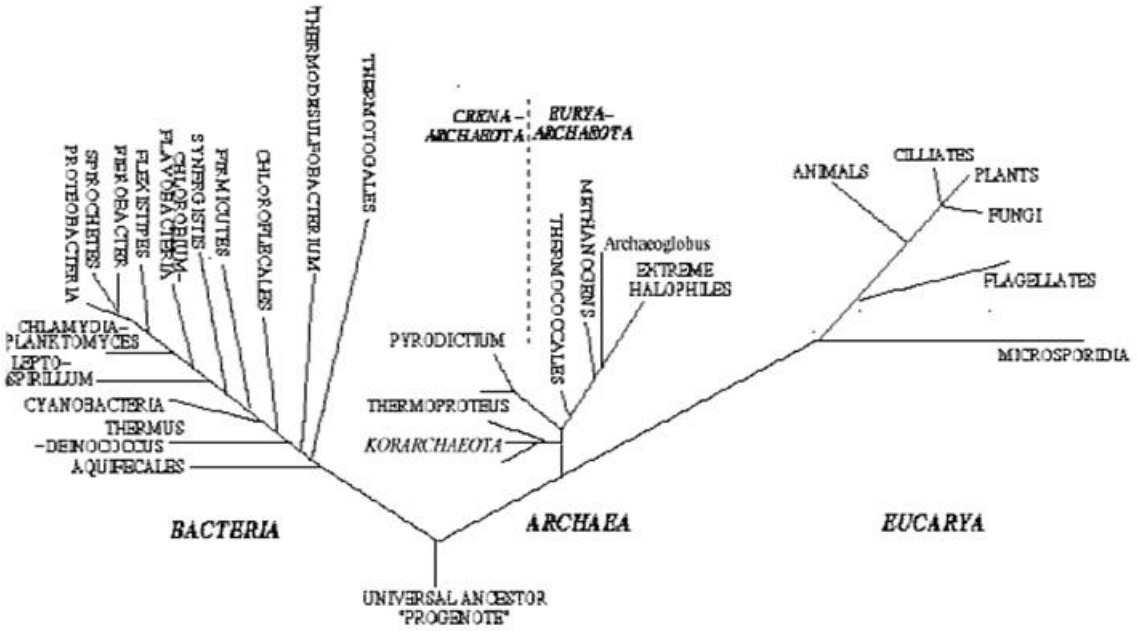

Fig 1. **The universal philogenetic tree of life.** It can be noted the subdivision in three different kingdoms: Bacteria, Eukarya and Archaea.

Further works about Archaea have revealed that these microorganisms are nearly ubiquitous on earth, and that some species live in extreme environments, with high temperature, pression and saline concentration, extreme pH values and strict anaerobiosis, such as hot springs and hydrothermal submarine vents, in alkaline or acidic waters. This peculiarity have captured the attention of scientific world, and the number of studies on archaeal organisms is constantly increasing. An important goal of scientist is to shed light on the molecular mechanisms of adaptation and resistance to extreme conditions. Moreover, enzymes from Archaea have several biotechnological applications, due to the ability of working under otherwise prohibitive physical and chemical conditions.

Several studies have clarified that Archaea are biochemically and genetically closer to Eukarya than to Bacteria; the genomic organization is similar to bacterial one, whereas some components of replication, transcription and translation complexes

resemble the eukariotyc homologous proteins, as indicated also by the presence of intron sequences into some genes and the shared insensibility to antibiotics inhibiting the ribosomal complex. Moreover, the close relation between Archaea and Eukarya has been demonstrated studying the sequence of duplicated genes. Archaeal microorganisms also show several unique features, with no similarity to the other domains of life, the structure and composition of cellular membrane and wall being the most distinctive.
The archaeal domain is now divided into three kingdoms: Crenarchaeota (hyperthermophylic and thermoacidophylic organisms); Euryarchaeota (hyperthermophylic, methanogenic and alophylic organisms); Korarchaeota (organisms biochemically and physiologically not completely defined yet)

Sulfolobus

The members of the genus *Sulfolobus* are aerobic thermoacidophilic microorganisms belonging to the crenarchaeal branch of the archaeal domain, with optimal growth occurring at pH 1-5 and temperatures of 75-85°C. The habitats of these organisms comprise globally spread geothermally active areas, like the aerobic zones of acidic pools and hot springs. Sulfolobales were earlier isolated from hot springs of the Yellowstone park (USA); since then many species have been found throughout the world in different spots, in Japan , Italy, Iceland, Russia and New Zealand.

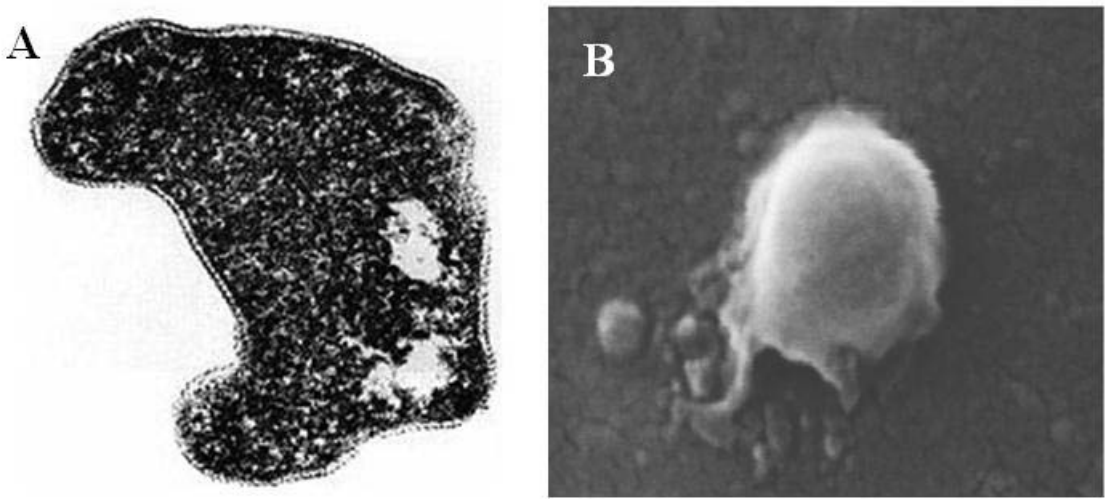

Fig. 2 Images of *Sulfolobus* cells. In the box **A** is shown the electron micrography of a thin section from a *S. acidocaldarius* cell; under an electron microscope, the cells look like irregular spheres often endowed with lobes. In the box **B** in displayed a differential scansion microscopy image of a *S. solfataricus* cell.

The organisms belonging to this genus show a lithotrophical or organotrophical grow, oxidizing elementary sulfur and complex organic substrates, respectively. Cells have a diameter ranging from 0.8 to 2µm and length up to 100µm, the shape is irregular and some cells can have one or more pili attached to their cell surface. (Fig.2). *S. solfataricus* and *S. acidocaldarius* strains can grow to high cell densities on organic substrates and a variety of sugars with doubling times in the range of 3–6 h during exponential growth and are relatively easy to cultivate in the laboratory.

Moreover, the complete genome sequence is available for both strains (*S. solfataricus* P2 genome was one of first crenarchaeal genome to be completely sequenced). Consequently, *S. solfataricus* and *S. acidocaldarius* have developed into important model organisms for the study of Crenarchaeota.

Genetic elements

The crenarchaeote *Sulfolobus* spp. is a host for a large spectrum of genetic elements. Some viruses, plasmids and insertion sequences have been found, and the elements often show a large extent of novelty with respect to morphology and genome content. These genetic elements has been extensively studied over the last few years, with the aim to develop shuttle vectors for *Sulfolobus* strains. Genetic methods for studying these organisms are not as advanced as for the halophilic and mathanogenic Euryarchaeota, and shuttle vectors are also important tools to (over)express homologous or heterologous proteins of interest. The proteins can be expressed and characterized in native host, be it for reasons of post-translational modifications or the determination of the localization of the protein or because the respective protein is difficult to express in functional form in mesophilic expression systems. Shuttle vectors are not only crucial for basic research applications but have also potential for exploitation in biotechnological applications, e.g., for the (over)production of heat stable enzymes.

Transposable elements

Transposable elements (TEs) are DNA sequences capable of moving into genome, and cause dramatic changes in genes and genomes via molecular events that do not depend on the proximity or similarity of the DNA sequences affected. These events include inactivation of functional genes by insertion, activation of cryptic genes by positioning of a promoter 5' to the coding region, deletion or inversion of DNA adjacent to the TE, and stable incorporation of DNA transferred from outside the lineage (Kleckner N. *et al.*, 1975; Ciampi M.S. *et al.*, 1982; Prentki P. *et al.*, 1986). In addition to these TE-promoted changes, homologous recombination between dispersed copies of a TE can rearrange large genomic segments (Haack K.R. and Roth J.R., 1995). These properties make TEs, which occur in nearly all organisms, a major source of genomic plasticity.

The mobility of TEs is due to the presence of repeats at the ends of sequence, and the elements need a transposase, normally encoded by the element itself, to effect their mobility. Insertion sequences (ISs) are the smallest TEs capable of independent transposition. A typical IS consists of a transposase-encoding gene flanked by inverted repeats (IRs), which provide the recognition and cleavage sites for the transposase (Mahillon J. and Chandler M., 1998). In situ, an IS is usually bounded by short direct repeats (DRs) that represent target-site duplications (TSDs) resulting from transposition. TEs were found in several bacterial, eukaryotic and archaeal genomes (Jurka J. 1998; Mahillon J. and Chandler M. 1998). In bacteria, TEs often inactivates genes by insertion, otherwise a TE can confer to the microorganism new useful properties, e.g. carrying a sequence for an antibiotic resistance or a degradative methabolic pathway.

The presence of TEs is also relevant in hyperthermophilic archaeal genomes. Each of the sequenced *S. solfataricus* and *S. tokodaii* genomes contains large numbers of putatively mobile elements, both IS elements (insertion sequence elements) and MITEs (miniature inverted-repeat transposable elements). There are 344 IS elements

in the 3.0 Mb genome of *S. solfataricus* P2 and 95 in the 2.7 Mb genome of *S.tokodaii* (Brügger K. *et al.* 2004). In the former they constitute more than 10% of the genome and the elements are mainly clustered in two broad areas (replicore 1 and replicore 2) separated by the putative oriC and terC (replication origin and termination) regions (Brügger K. *et al.* 2002); these sequences, fundamental for replication, should act as barriers for the mobility of transposable elements. Furthermore, experimental data suggest that transposition of IS elements occurs frequently. The ISC 1058, 1217, 1359 and 1439 have shown to be active and able to spread and interrupt a functional gene (Martusewitsch E. *et al.* 2000) by spontaneous excision-insertion or copying-insertion events; it has been proposed that they are probably mobilized by integrase and excisionase activities, often encoded by the same transposase open reading frame (ORF) (Brügger K. *et al.* 2002, Redder P. *et al.* 2001, Rezsöhazy R. *et al.* 1993) located on the IS elements themselves. Moreover the presence of repeats (identical copies) of some elements suggests a recent duplication. The large number of TEs gives a remarkable plasticity to genome, and this could be a key process in the adaptation strategy of hyperthermophlic organisms to extreme habitats.
In this regard, *lacS*- mutants of *S. solfataricus* defective in β-glycosidase activity were isolated (Bartolucci S. *et al.* 2003), in which mutations were caused by an insertion of transposable elements into *lacS* gene. One of the mutants, named GθW, showed a stable phenotype with no reversion; analysis of its chromosome revealed the total absence of the β-glycosidase gene (*lacS*), with an extended deletion of 13 kb including the whole *lacS* sequence, *xylS* (α-xylosidase) and *lacTr* (lactose transporter) genes. This chromosomal rearrangement was a nonconservative transposition event driven by the mobile insertion sequence element ISC1058, belonging to IS5 family (Mahillon, J. and Chandler M. 1998) and found in 14 complete copies in *S. solfataricus* P2 genome.
The isolation of stable complementable mutations from *S. solfataricus*, as *lacS* deletion found in GθW, opens new possibilities for the development of genetic tools. In this work, we have used this strain in complementation experiments in the aim of characterize the shuttle vector pMSSV (see above). Moreover, GθW mutant is faster growing in rich medium than the wild-type, with a doubling time of 4 versus 6 h; this is an interesting feature in the perspective of use the mutant in biotechnological application, e.g. expression of homologous/heterologous proteins.

Virus
Viruses are the largest reservoir of genetic material on the planet; in particular microbial viruses are extremely abundant on our planet; they have important and diverse roles in ecosystems and biogeochemical processes. Archaea and their viruses are poorly understood when compared with the Eukarya and Bacteria domains of life, so over the last few years efforts focused on the characterization and comparison of archaeal viruses are constantly increasing. All archaeal viruses isolated up to now have linear or circular, double-stranded (ds) DNA genomes. In common with their hosts, they show adaptation to extreme environments. Thus, viruses of extreme halophiles are stable only in solutions of high salt concentration (3–5 M) and are inactivated in solutions of low ionic strength (Witte A. et al. 1997). Similarly, viruses of hyperthermophiles are stable at extremely high temperatures (Prangishvili D. *et al.* 1999 ; Schleper C. *et al.* 1992).
From a morphological approach, DNA viruses of the Archaea have highly diverse and often exceptionally complex morphotypes. Many have been isolated from

geothermally heated hot environments, contradicting the widespread notion of limited biodiversity in extreme environments. Archaeal viruses exhibit a range of virion morphotypes, most of which have not been observed before for any dsDNA virus (Prangishvili D. *et al.* 2006a,); these viruses, particulary from hyperthermophilic Crenarchaeota, show exceptional forms, including fusiform, droplet and bottle shapes, as well as linear and spherical virions, and also more complex virions combining features of these different forms (Haring M. *et al.* 2005; Rice G. *et al.* 2001) (Fig.3).

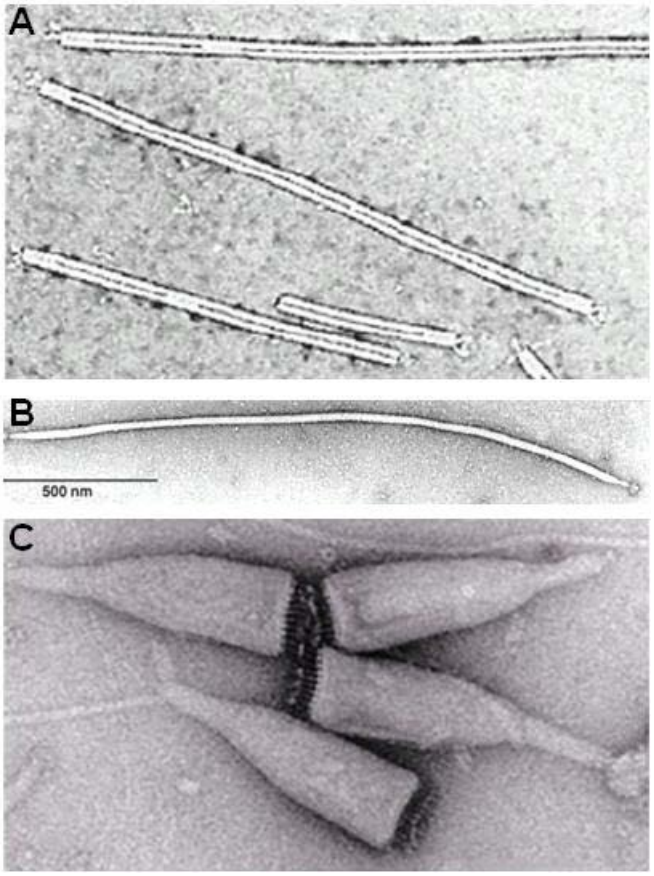

Fig. 3 Scansion electron microscopy images of *Sulfolobus* virus Viral particles of SIRV2, belonging to *Rudiviridae* family, are shown in the box **A;** in **B** is shown the SIFV virus (*Lipothrixviridae family)*; both viruses were isolated from *S. islandicus.* Box **C**: Acidianus bottle-shaped virus (ABV).

Moreover, genome-sequence analyses have demonstrated that most of the archaeal viruses are unrelated to other known viruses and suggest that they might have different, and possibly multiple, evolutionary origins (Prangishvili D. *et al.* 2006b). Most of the viruses isolated thus far infect hyperthermophilic hosts belonging to the *Sulfolobales* family. The viruses found in *Sulfolobus* are arranged to four families: *Fuselloviridae* (SSV1, SSV2, SSV RH, SSV K1 and the recently isolated SSV4), *Rudiviridae* (SIRV1 and SIRV2), *Lipothrixviridae* (SIFV) and *Guttaviridae* (SNDV); furthermore an icosahedral virus from the Yellowstone National Park has been recently discovered, which could not be assigned to a virus family (Rice G. *et al.* 2004; Lipps G. 2006).
However, the relatively largest amount of data provides information available about the *Fuselloviridae* family; the viruses belonging to this group show an inusual lemon

shape, which is solely found in archaeal domain. They have an approximately 60 to 90-nm size with sticky tails fibers extending from one end. These tail fibers are presumed to be involved in viral attachment to the host and association with membrane vesicles and are likely the cause of virus clustering into rosette formations as seen in culture. SSV viruses have an envelope made of coating hydrophobic proteins and probably host lipids. A recent comparative genomic analysis of four hyperthermophilic archaeal fusellovirus (Wiedenheft B. *et al.* 2004) allow us to make some interesting considerations (Fig.4).

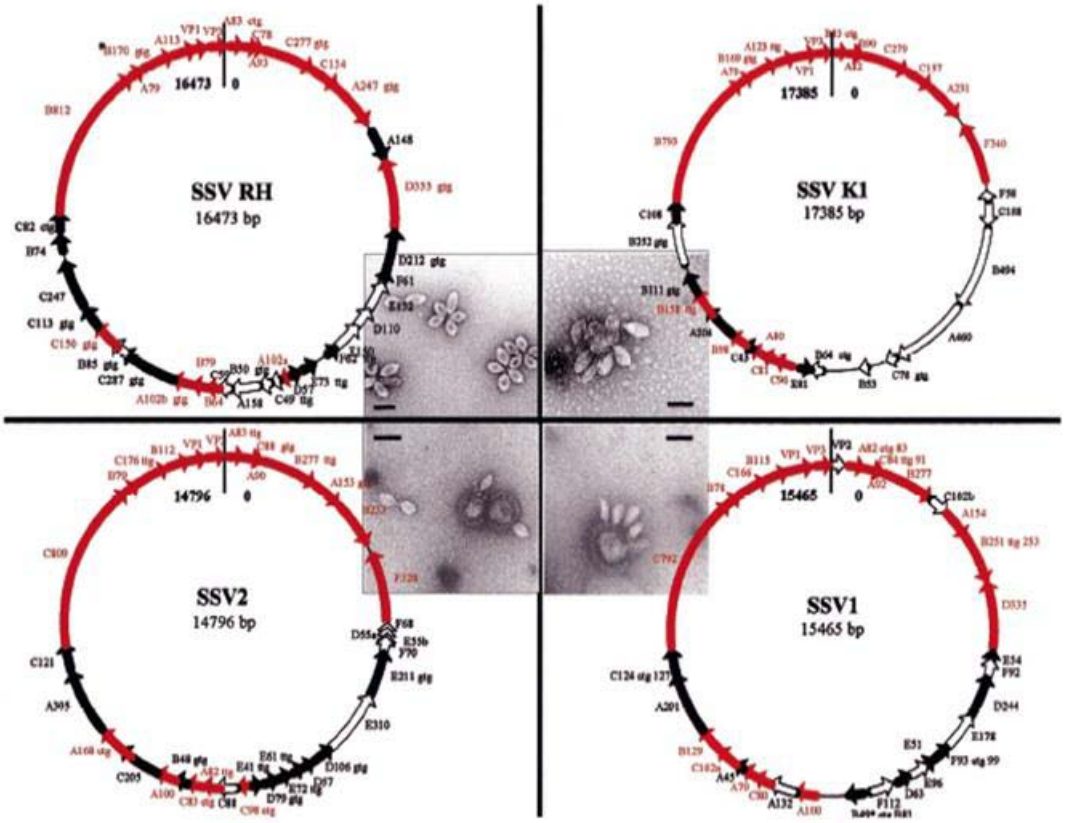

Fig. 4 Genomic maps of four SSV. Conserved ORFs shared by all genomes are shown as red arrows. ORFs shared between two or three of the SSV genomes are shown as solid black arrows, and ORFs unique to each isolate are shown as open arrows. Electron microscopy images of each virus are positioned next to maps of their respective genome (bars, 100 nm)

A first interesting feature of Fusellovirus*es* is their geographical distribution, as they are ubiquitous viruses. In fact, SSV1 was isolated in Beppu, Japan, SSV2 in Iceland, SSVRH in the United States and SSVK1 in Russia. Moreover, ORFs of these viruses show little or no similarity to genes in the public databases, as found for other archaeal viruses. However, among the four genomes, many ORFs share sequence similarity and are collinearly organized. In general, ORFs on one half of the genome are more highly conserved between the four isolates and are arranged in the same orientation, while genes on the other half of the genome are poorly conserved.
The genome comparison of all four viruses also reveals that there is a set of 18 ORFs common to all four SSV isolates. The products encoded by this set of 18 ORFs may represent viral common functions and clearly reflect a common evolutionary history, despite their geographic isolation. This set of genes may also represent the minimal replicon of the Fuselloviridae. The presence of unique ORFs in SSV

genomes are likely the consequence of their individual evolutionary history, geographic isolation, factors required for replication in their specific hosts, or adaptation to unique features of their respective thermal environments.
SSV1 is the first high-temperature virus to be characterized in detail and is the best-studied virus of the genus *Sulfolobus*. This Fusellovirus was isolated from the strain B12 of *Sulfolobus shibatae* in Beppu, Japan, (Grogan D. *et al.* 1990; Martin A. *et al.* 1984; Palm P. *et al.* 1991; Yeats S. 1982) but, like SSV viruses of different sources, it is also capable to infect *S. solfataricus* cells (Schleper, C. *et al.* 1992). In both hosts, virus production is UV inducible (Martin A. *et al.* 1984), and viral genome is packaged into viral particles in a positively supercoiled form, while the virus episomal DNA can exist in positively supercoiled, negatively supercoiled, or relaxed double-stranded DNA form inside the host cells (Nadal, M. *et al.* 1986). Infection and production of virions causes only a significant growth retardation of the host cells which can be visualized as turbid plaques around propagation foci on plated lawns of indicator host cells (Schleper, C. *et al.* 1992). SSV1 can integrate into the host arginyl tRNA gene using a tyrosine recombinase family integrase, in a similar way of other SSVs (SSV2 integrates into the glicyl tRNA gene, SSV RH and K1 can integrate into different tRNA genes; in addition SSVK1 can integrate into a non-tRNA spot of *S. solfataricus* genome, and this is the first example in SSV viruses). A 7.4-kbp segment inserted into an *S. solfataricus* arginyl tRNA gene shares extensive sequence similarity with a portion of the SSV1 genome and is likely a remnant of viral integration. The SSV1 genome (15465 bp) has 34 open reading frames (ORFs), which are tightly arranged in the genome (Palm P. *et al.* 1991); nine transcripts cover all 34 SSV1 ORFs (Reiter, W.D. *et al.* 1987), and this suggests that the viral genes are translated via a polycistronic strategy. However, only four genes of SSV1 could be assigned a function up to now; one of this ORFs encodes a type I tyrosine recombinase, which catalyses the site-specific integration (Muskhelishvili, G.P. *et al.* 1993), while the other three ORFs encode structural proteins, VP1, VP2, and VP3, which were assigned by sequencing of proteins from purified virus particles (Reiter W.D. *et al.* 1987). VP2 have no homologous proteins in other SSV viruses, unlike VP1 and VP3, even if the protein is known to play a key role in the virus infection mechanism, being involved in the packaging of genome into the viral particles.
The fusellovirus SSV2, isolated from *S. islandicus* REY 15/4, shares with SSV1 similar morphology, replication and DNA size (Stedman K.M. *et al.* 2003). The overall genome architecture is conserved but the low similarity in some regions of the sequences should be responsible for the higher copy number and the lack of a strong ultraviolet induction of episomal SSV2 DNA and particle production, as well as the different integration of the SSV2 genome which occurs into the host chromosome at the site of a glycyl tRNA (Wiedenheft B. *et al.* 2004). Recently a new fusellovirus, named SSV4, has been isolated (Peng X. 2008); this virus shows a similar genomic composition and organization of others SSVs, and integrates into a Glu t-RNA gene.

Plasmids

Two types of plasmids have been isolated for the genus *Sulfolobus*: cryptic plasmids, with genome sizes of 5–14 kb and conjugative plasmids, with genomes larger than 25 kb (Lipps G. 2006). The first conjugative plasmid discovered was the plasmid pNOB8 which was isolated from a Japanese strain of *Sulfolobus* (Schleper et al. 1995). The plasmid can propagate in liquid cultures through cells, that form aggregates of 2 to 30 individuals with intercellular cytoplasmic bridges connecting two or more cells. Conjugation is quite efficient and the plasmid is able to spread

through the whole culture. The conjugative plasmids are actually classified in three subfamilies: pNOB8, pING (Stedman K.M. *et al.* 2000) and pSOG2/4 (Erauso G. *et al.* 2006). These plasmids are smaller than PNOB8 (they are from 25kb to 35kb in length, while pNOB8 is 42kb); the genomes show each other a partial sequence homology, but also regions which are completely different. The plasmids have similar characteristics as the plasmid pNOB8 including high copy number and growth retardation and, excepting pSOG2/4, have shown to be stable upon conjugative transfer, undergoing high genetic variation only upon prolonged growth of the host cells. Comparisons of the genome sequences of the CPs sequenced allow to delineate the essential genes of this plasmid family (Greve B. *et al.* 2004). The analyses recognized three conserved and functionally distinct domains. The first one is a cluster of genes covering up to half of each genome (~13.5 kb) and contains six conserved ORFs implicated in conjugation; the second is a putative replication origin; finally there is a region including an operon with six to nine short genes some of which are involved in the initiation of plasmid replication, an integrase and the DNA binding protein, *PlrA*.

The other family of *Sulfolobus* plasmids comprises plasmids known as cryptic, because these elements give no distinguishable features to the hosts. The first cryptic plasmid completely sequenced was the plasmid pRN1, isolated from the *S. islandicus* strain REN1H1 (Keeling P.J. *et al.* 1996); this strain also harbours the plasmid pRN2. Detailed studies about replication of these plasmids have pointed out that both pRN1 and pRN2 can replicate and propagate independently, being not essential for the host cells (Purschke W.G. and Shäfer G. 2001). pRN plasmids give the name to a plasmid family composed of elements which share regions of high sequence similarity (Peng X. *et al.* 2000); pHEN7, a 7.8-Kb plasmid from *S. islandicus* HEN7H2 (Zillig W. *et al.*,1998), pDL10 (7.6 Kb) from the chemolithoautotrophic crenarchaeon *Acidianus ambivalens*, and pSSVx (5.7 kb) from the *S. islandicus* strain REY15/4 (Arnold H.P. *et al.*, 1999), pIT3 (5.0 kb) from *S. solfataricus* (Prato S. *et al.* 2006) belong to this family (Fig.5).

In the conserved regions, these plasmids share three ORFs. One of these conserved genes (named orf56 in pRN1) encodes for a sequence-specific double-stranded DNA-binding protein (CopG homologue), which could be involved in the plasmid copy number control. The orf80 genes of the plasmids are the most conserved genes of the plasmid family pRN, and encode for a DNA binding protein which have an unknown physiological function; the orf80 homologues, which are highly conserved also within the conjugative plasmids of *Sulfolobus*, are partly annotated as plrA, meaning plasmid regulatory gene A (Greve B. *et al.* 2004). The other conserved sequence is a large orf (2676 bp in pSSVx to 2970 bp in pRN2) which occupy about a third to a half of the plasmids. The proteins encoded by the Orf (904) from pRN1 (Lipps G. *et al.* 2003) and the Rep245 from pIT3 (Prato S. *et al.* 2008) have been expressed in *E.coli* and characterized, and showed to be multifunctional proteins with helicase, primase and DNA polymerase activity. This protein could play a crucial role in plasmid replication; has been suggested that Orf904 alone or in concert with another protein, probably Orf80, recognizes the origin of replication. Subsequently, the helicase domain of Orf904 will melt the replication origin using chemical energy from the hydrolysis of ATP and the prim/pol domain will synthesize a primer (Lipps G. 2006). In this contest Orf 80 could have a structural role, e.g. marking the replication origin, while the CopG homologous protein could limit the intracellular concentration of the replication protein and so down-regulate plasmidal replication initiation.

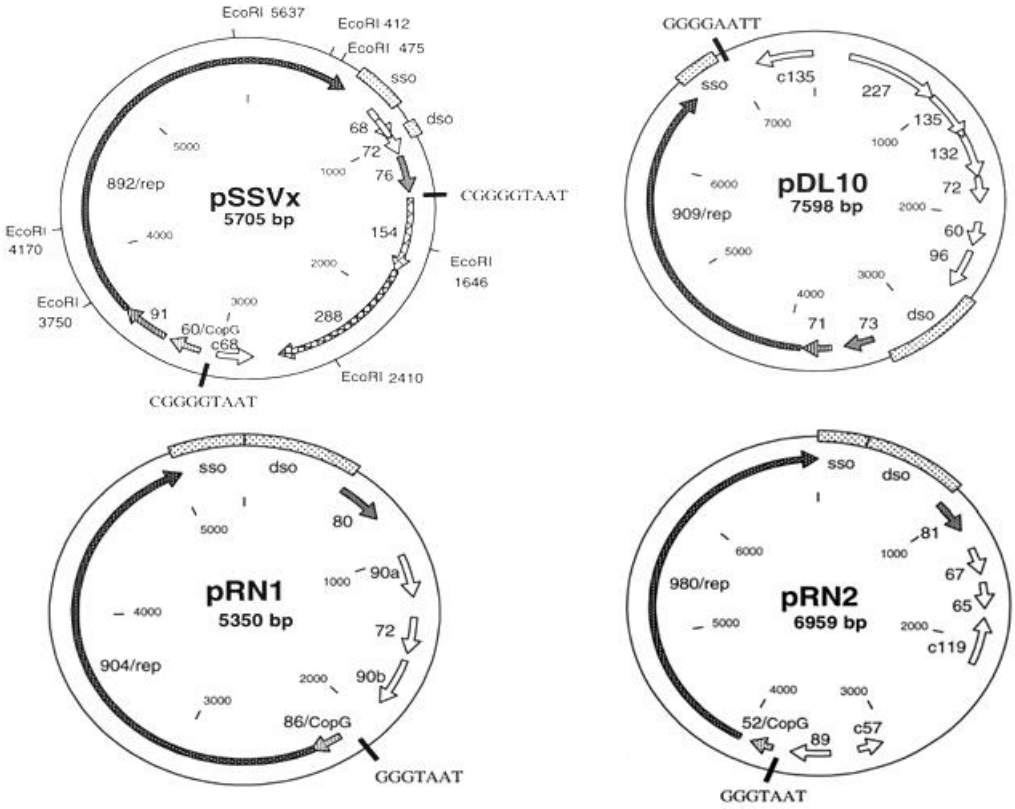

Fig. 6 Maps of the genomes of the plasmids pSSVx, pRN2, pRN1 and pDL10 forming the pRN family. ORFs were named by the number of codons. 'c' (complementary) indicates opposite orientation. ORFs in the `minimal replicon' (conserved region) are shown as filled arrows; sequence elements in this region as filled boxes. Unassigned ORFs in the variable region are shown as empty arrows, and SSV homologues are cross-hatched. Homologous ORFs are indicated by the same fill pattern. Consensus motifs located between conserved and variable regions are indicated.

Plasmid-fusellovirus genetic systems

pSSVx plasmid was isolated together with SSV2 from *S. islandicus* strain REY15/4 type (Arnold H.P. *et al.,* 1999). This element has been discovered in a screening procedure to detect extrachromosomal genetic elements of *Sulfolobus*, especially viruses of the SSV1 type. Arnold and collegues discovered the presence of spindle-shaped, short-tailed virus particles of two distinct sizes analyzing the supernatant of *S. islandicus* REY15/4 culture by electron microscopy (Fig.6). Both particles resembled those of the *Sulfolobus* virus SSV1 in shape (Martin A. *et al*., 1984); the larger particles had the same size as SSV1, 80nm x 55 nm, whereas the smaller particles measured only 60 x 40 nm. Further analysis allow Arnold to assert that the larger particles contained a fusellovirus genome, named SSV2, while the smaller one contained a 5.7kb plasmidic DNA, pSSVx. By sequence similarity, the pSSVx plasmid has been assigned to the pRN family of *Sulfolobales* plasmids (as previously described), but it shows an unique feature among pRNs, the ability to generate virus particles and spread, also in the cell cultures of *S. solfataricus*, using the packaging mechanism of SSV2. This genetic element lacks the genes that encode the major structural proteins of the virus, therefore is unable to package and spread without a virus helper. Therefore, pSSVx can be rewarded as a hybrid between a plasmid and a virus. Differently from the members of the pRN family that are incapable to spread

even in the presence of helper virus, pSSVx contains two open reading frames, named 154 and 288, showing high sequence similarity to a tandem of ORFs in both SSV1 and SSV2 genomes. It has been suggested that the proteins encoded by these ORFs can specifically recognize pSSVx DNA but associate with viral helper components necessary for capsid formation and packaging. The relationship of pSSVx to its helper fusellovirus SSV2 resembles that of the parasitic "defective" bacteriophage P4 to its compulsory helper, bacteriophage P2, especially in the packaging of these elements into helper-virus-like but smaller particles assembled by structural components of the helper. However, pSSVx appears to use the structural genes of SSV2 only for spreading, whereas bacteriophage P4 modifies the lifestyle of its helper. A trascriptional analysis of the pSSVx genetic element was developed by Contursi *et al.* (2007): a combination of Northern blot, primer extension and RT-PCR experiments, revealed the presence of nine major transcripts whose expression was differentially and temporally regulated over the growth cycle of *S. islandicus*.

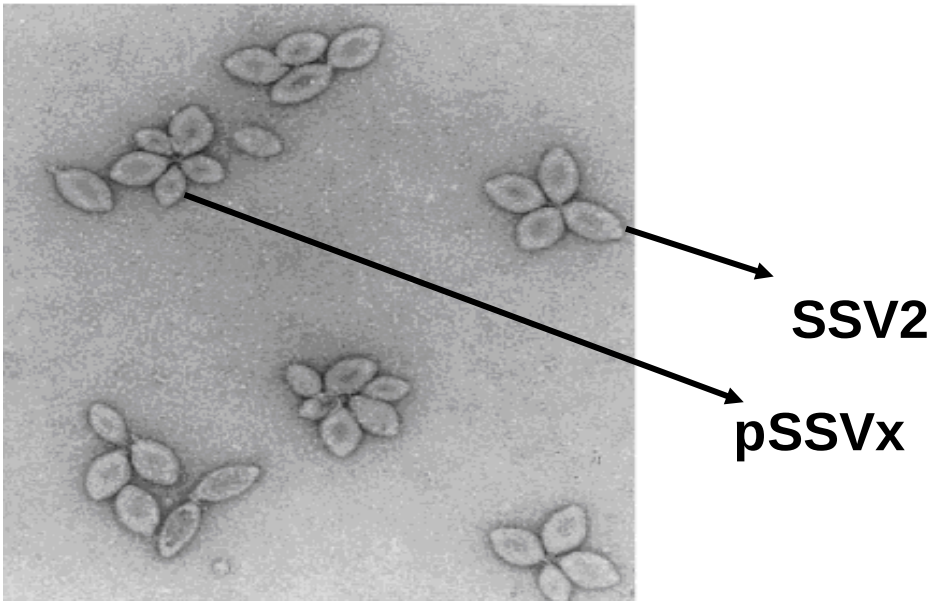

Fig. 6 Electron microscopy images of the *S. islandicus* REY 15/4 colture supernatants. The image shows mixtures of large SSV2 and small pSSVx particles, mostly in rosettes, from cultures of strain REY 15/4.

Recently two novel fusellovirus-plasmid systems have been discovered. An integrative non-conjugative extrachromosomal genetic element, denoted as pSSVi, has been isolated from a *S. solfataricus* P2 strain and characterized (Wang Y. *et al.* 2007). This genetic element is a double-stranded DNA of 5740 bp in size and contains eight ORFs. It resembles members of the pRN plasmid family in the genome organization but shows only weak similarity to the latter in the most conserved regions. pSSVi has a copG gene similar to that of a pRN plasmid, encodes a large replication protein which, unlike a typical pRN RepA, contains no polymerase/primase domain, and lacks the *plrA* gene. Interestingly, pSSVi encodes an SSV-type integrase which probably catalyzes the integration of its genome into a specific site (an Arg-tRNA gene) in the *S. solfataricus* P2 genome. Like pSSVx, pSSVi can be packaged into a spindle-like viral particle and spread with the help of SSV1 or SSV2. pSSVi genome is stably integrated into the host chromosome and it is excised in presence of SSV2 virus. In addition, both SSV1 and SSV2 appeared to

replicate more efficiently in the presence of pSSVi. Given the versatile genetic abilities, pSSVi appears to be well suited for a role in horizontal gene transfer.
Moreover, a new fusellovirus, SSV4, and a pRN-like plasmid, pXZ1, were co-isolated from the ARN3/6 strain of *S. islandicus* (Peng X. 2008). In contrast to the previously characterized virus–plasmid hybrids pSSVx and pSSVi, which can coexist intracellulary with a fusellovirus, pXZ1 is not packaged into viral particles and shows no viral infectivity. The virus and plasmid have a genome of respectively 15135 and 6970 bp, with 33 and 7 ORFs. Three ORFs of pXZ1 encode an atypical RepA, a PlrA and a CopG protein. A fourth ORF exhibits a high nucleotide sequence identity to the SSV4 integrase gene, which suggests that it has been transferred to the plasmid from SSV4. A single point mutation within an otherwise identical 500 bp region of the integrase gene occurs in the viral attachment site (attP), which corresponds to the anticodon region of the targeted tRNA gene in the host chromosome. This point mutation confers on pXZ1 the ability to integrate into the tRNAGlu[CUC] gene, which differs from the integration site of SSV4, tRNAGlu[UUC]. Comparing the features of pSSVi/pSSVx and pXZ1 systems, it could be assumed that packaging or integration are different strategies adopted by plasmid to survive in *Sulfolobus* cells growing under extreme conditions. When coexisting with a fusellovirus either they exploit the viral integrase, as in case of pXZ1, to integrate into a different chromosomal site or they exploit the viral packaging system and spread as a virus satellite, as found for pSSVx and pSSVi.

S. solfataricus-E. coli shuttle vectors

Over the last few years, great efforts have been made to develop shuttle vectors for hyperthermophilic Archaea; shuttle vectors are a powerful tool for genetic analysis and manipulation *in vivo*, and can also be successfully used in biotechnological applications, such as the overproduction of heat stable proteins which are difficult to produce in mesophilic hosts. Despite the importance of these systems, there are not many vectors constructed for these organisms, unlike the various systems set-up for the bacterial domain or for other Archaea (halophiles or methanogens) (Allers T. and Mevarech M. 2005), demonstrated to be hosts for very advanced vector development. This scarcity of available vectors can be a reflection of the difficulties in establishing efficient systems, due to low transformation efficiencies, inefficient selection and/or instability of the vectors in the host.
The common criterion to develop shuttle vectors has been based on the suitable modification of extrachromosomal elements of *Sulfolobus*. The highly efficient self-spreading capabilities of fusellovirus SSV1 has been exploited to construct several elements. The vector pEXSs is based on a part of the genome of the SSV1 cloned into pGEM5Zf(−) and contains a heterologous selectable marker gene coding for a thermostabilized version of the hygromycin phosphotransferase (Hph) from *E. coli* (Cannio R. *et al.* 2001). This vector was used to express a thermostable alcohol dehydrogenase from *Bacillus stearothermophilus* (Contursi P. *et al.* 2003), and to complement a *lacS* deletion mutant (*S. solfataricus* GθW) by expressing the genes *lacS* and *lacTr* coding for a β-glycosidase and a lactose transporter (Bartolucci S. *et al.* 2003). Stedman et al. (1999) has constructed a series of *Sulfolobus–E. coli* shuttle vectors based on the complete genome of the virus SSV1. The vector pBluescript has been inserted at different sites within the virus genome and constructs have been identified that have been not impaired in replication or infectivity. To one of

these constructs the *pyrEF* genes from *S. solfataricus* P2 coding for orotatphosphoribosyl transferase and orotidine-5′-monophosphate decarboxylase has been added as selectable marker. The expression of these marker genes allows for the complementation of uracil auxotroph recipients to uracil prototrophy. Additionally the *lacS* gene under control of the heat shock tf55α-promoter has been cloned into the vector as a phenotypic marker. The resulting shuttle vector pMJ03 (Jonuscheit M. *et al.* 2003) replicates to high copy numbers in the primary transformation mixture as an episome. After plating and isolation of single transformants this vector has been found to be integrated as a single copy into the chromosomal arginyl-tRNA gene of the recipient *S. solfataricus* PH1-16 (Martusewitsch E. *et al.* 2000) as previously observed for the wild type virus (Schleper C. *et al.* 1992). The pMJ03 vector has been improved by Albers et al. (2006) by the development of preassembly constructs and an inducible promoter. Recently a series of *Sulfolobus–Escherichia coli* shuttle vectors based on the small multicopy plasmid pRN1 from *S.islandicus* have been constructed (Berkner S. *et al.* 2007). The shuttle vectors do not integrate into the genome and do not rearrange; selection in suitable uracil auxotrophs is provided through inclusion of *pyrEF* genes in the plasmid. The plasmids allow functional overexpression of genes, as has been demonstrated for the β-glycosidase (*lacS*) gene of *S. solfataricus*.
In this PhD work a *S. solfataricus – E. coli* shuttle vector, named pMSSV, based on the hybrid genetic element pSSVx was used; it was obtained by a suitable site-specific insertion into pSSVx of the sequences from an *E. coli* plasmid essential for replication and selection of transformants (Fig. 7). This vector showed the ability to efficiently propagate in both hosts and to replicate at high copy number in a fashion similar to the wild type pSSVx.

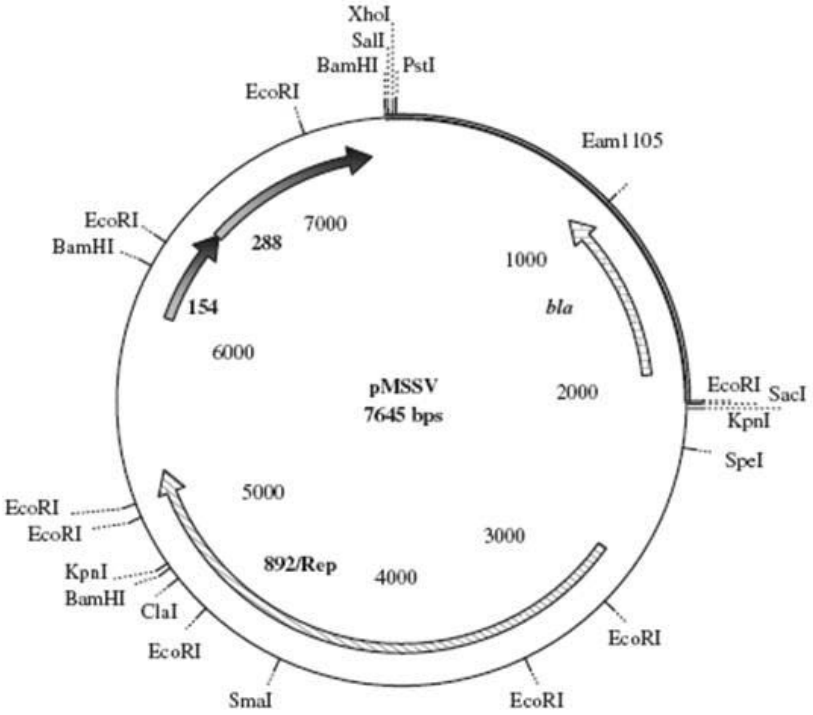

Fig. 7 Map of vector pMSSV. The sequences necessary for replication and ampicillin selection in *E. coli* (indicated as a solid bar) was inserted into pSSVx plasmid, to generate a new *E. coli- S. solfataricus* shuttle vector, named pMSSV.

Thermophilic enzymes as biocatalysts in industrial applications

The field of industrial enzyme applications has grown tremendously over the last years; enzymes are actually used in several industrial sectors, such as food, textile, detergent and pharmaceutical industry. Enzymes allow to improve the process execution, making the processes economically and environmentally more sustainable than the traditional chemical manufacturing approaches. Moreover, the utilization of enzymes provides the market with better quality products when compared to those obtained by traditional methodologies. Particularly, enzymes from (hyper)thermophilic organisms show a resistance to extreme physical-chemical conditions, prohibitive for their counterparts from mesophilic organisms.
These unique features make enzymes from hyperthermophiles very advantageous in most of the industrial processes that requires high temperature regimens, and for which the employment of the thermolabile mesophilic enzymes is limited or even completely unsuccessful. The main advantages of performing processes at high temperatures are the reduced risk of microbial contamination, lower viscosity, improved transfer rates, and improved solubility of substrates/products.
Many studies are still currently devoted to a defined understanding of the molecular basis conferring stability to thermozymes. There are no major conformational differences with mesophilic enzymes, and a small number of extra salt bridges, hydrophobic interactions, hydrogen bounds or a decreased length of surface loops seem to contribute to the extra degree of stabilization (Bruins M. *et al*. 2001). Recently, the increasing numbers of known three-dimensional structures and some studies on mutant proteins have clarified the importance of ion pairs, when organized in large networks. It is believed that these networks at the surface of proteins are a major stabilizing factor for thermostability (Sadeghi M *et al.* 2006; Ge M. *et al.* 2008). Stabilizing interactions between domains and subunits also contribute significantly to the intrinsic stability of proteins. Enzymes from (hyper)thermophilic organisms are known to exist in conformational states of higher-order association when compared to their mesophilic analogs, suggesting that the formation of oligomers is one way of increasing thermostability (Jaenicke R. 1998). Although most enzymes from (hyper)thermophilic organisms are intrinsically very stable, some intracellular enzymes obtain their thermostability from intracellular environmental factors. The presence of salts (hypersolutes), high-protein concentrations, coenzymes, substrates, activators, or general stabilizers such as thermamine, sorbitol, or cyclic polyphosphates can also further stabilize the enzyme (Bruins M. *et al*. 2001).

Cellulose

Structure and composition

Cellulose is the most widespread polysaccharide in nature and it forms the primary structural component of plants. Cellulose is a linear polymer made of repeated units of the D-glucopyranose linked by β-(1→4) bounds. Cellulose occurs naturally in pure form in cotton fibres, or it can be found in association with other biopolymers, like lignin and hemicelluloses (Fig.8).

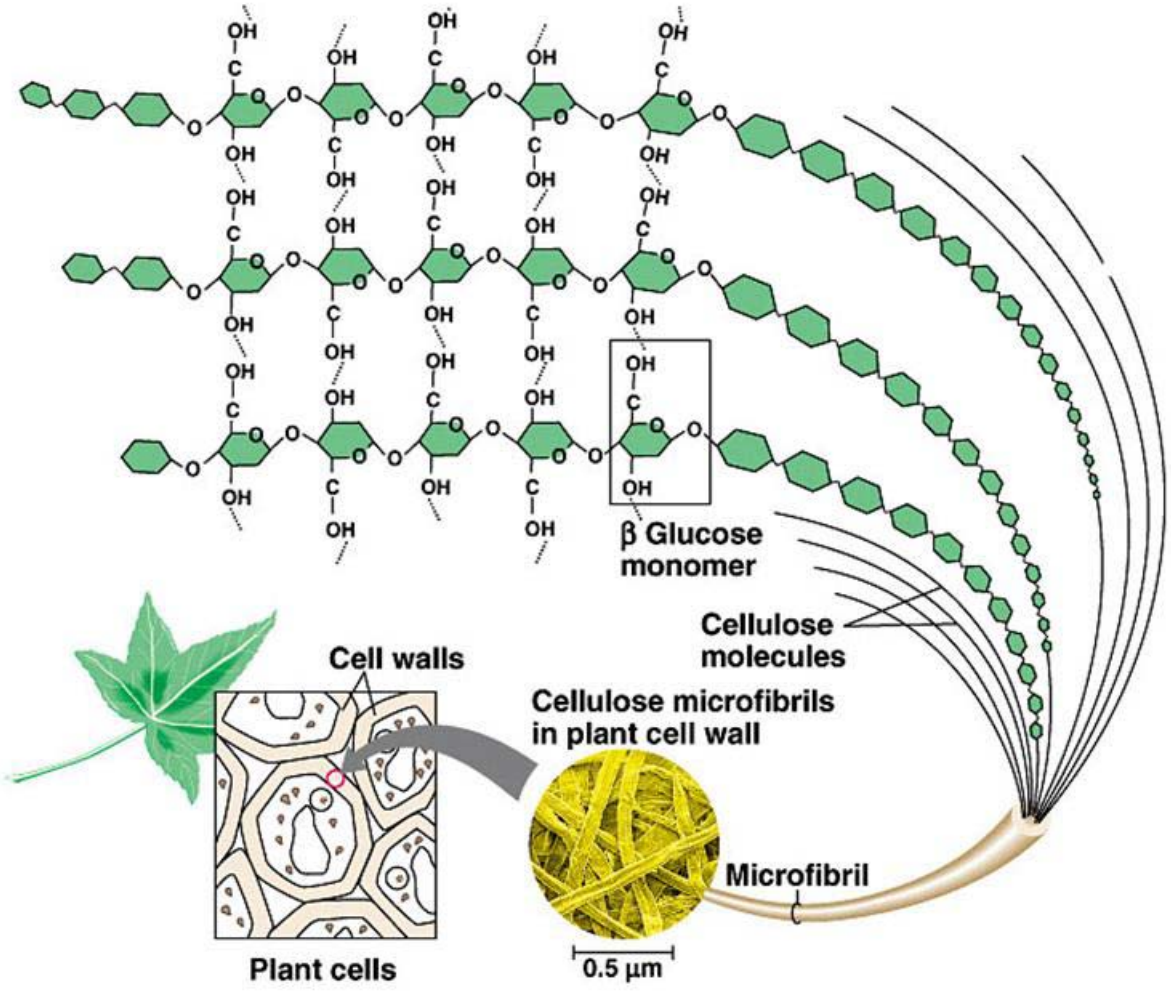

Fig. 8 Schematic representation of cellulose polymer in the plant cell wall. Glucose monomers are linked by β 1,4 glycosidic bonds; this causes the molecular chain of cellulose to extend in a more-or-less straight line, making it a good fiber-forming polymer. The single chains are organized in microfibrils, linked by interchain hydrogen bounds.

Cellulose was discovered and isolated as the common material of plant cell walls by Anselme Payen in 1838, and it was further characterized by Cross and Bevan in the early 1900s. They removed the related plant materials that occur in combination with cellulose by dissolving them in a concentrated sodium hydroxide solution. They designated the undissolved residue as α-cellulose. The soluble materials (designated as β-cellulose and γ-cellulose) were later shown not to be celluloses, but rather, relatively simple sugars and other carbohydrates. The α-cellulose of Cross and Bevan is what is usually meant when the term “cellulose” is used now. In the cellulose chain, glucopyranoside units are joined by acetal linkages between the C-1 of one pyranose ring and the C-4 of the next ring. The stereochemistry of these acetal linkages is very important for typical structural conformation of cellulose; the C-1 oxygen is in the opposite side of C-6 carbon, (β configuration). This β configuration, with all functional groups in equatorial positions, causes the molecular chain of cellulose to extend in a more-or-less straight line, making it a good fiber-forming polymer.

Because of the equatorial positions of the hydroxyls on the cellulose chain, they protrude laterally along the extended molecule. This positioning makes them readily available for hydrogen bonding. These hydrogen bonds cause the chains to group together in highly ordered (crystal-like) structures. Since the chains are usually longer than the crystalline regions, they are thought to pass through several different crystalline regions, with areas of disorder in between (the “fringed-micelle” model). The inter-chain hydrogen bonds in the crystalline regions are strong, giving the resultant fibers good strength and insolubility in most solvents. They also prevent cellulose from melting. In the less ordered regions, the chains are further apart and more available for hydrogen bonding to other molecules, such as water. Most cellulose structures can absorb large quantities of water (i.e., it is very hygroscopic). Thus, cellulose swells, but does not dissolve, in water.

Enzymatic degradation and biotechnological applications

There are three main enzymatic activities involved in cellulose degradation, which can be discerned as follows:

- Endoglucanases, or endo-1,4-β glucanases (EC 3.2.1.4), which randomly hydrolyze the 1,4-β bonds within cellulose chains acting on amorphous regions and producing oligosaccharides of various length.
- Exoglucanases, which include 1,4-β-D-glucan-4-glucohydrolases, also named cellodextrinases (EC 3.2.1.74), and 1,4- β -D-glucan cellobiohydrolases (EC 3.2.1.91); these enzymes act on reducing and nonreducing ends of cellulose chains, releasing respectively glucose (glucohydrolases) or cellobiose (cellobiohydrolases), and can also hydrolyze microcrystalline cellulose (Teeri T. T. 1998).
- β-glucosidases or β-D-glucoside glucohydrolases (EC 3.2.1.21), which catalyse the hydrolysis of soluble cellodextrins (oligosaccharides made of three to seven glucose molecules) and cellobiose to glucose.

These enzymes catalyze the hydrolysis of β -1,4-glycosidic bounds via general acid catalysis that requires two critical residues: a proton donor and a nucleophile/base. This hydrolysis occurs via two major mechanisms giving rise to either an overall retention, or an inversion, of anomeric configuration (Davies G. and Henrissat B.1995; Lynd L.R. *et al.* 2002) In both the retaining and the inverting mechanisms, the position of the proton donor is identical, in other words it is within hydrogen-bonding distance of the glycosidic oxygen. In retaining enzymes, the nucleophilic catalytic base is in close vicinity of the sugar anomeric carbon. This base, however, is more distant in inverting enzymes which must accommodate a water molecule between the base and the sugar.

The enzymes involved in cellulose hydrolysis have large biotechnological potentialities; in fact cellulases are widely utilized in industrial processes, e.g. in paper, textile and detergent industries. In the paper production, the step by step process used to separate cellulose from lignin and other wood components is known as pulping. It is a time and energy consuming process, involving the mechanical processing of wood or the treatment of wood with harsh chemicals. In biopulping, cellulase and xylanase enzymes are used to pre-treat wood and break down the lignin fibres. Removing lignin prior to further wood pulping saves time and energy, and decreases the quantities of chemicals used and water wasted. In textile industry, enzymes are used to treat and modify fibres, particularly during textile processing and in caring for textiles afterwards (desizing, biopolishing and stonewashing processes). Cellulases are also added to detergents to improve the cleaning and

softening power. In general, detergents containing enzymes are more environmental friendly than the traditional ones, allowing lower washing temperatures and shorter periods of agitation to be employed.
Cellulose is the earth's most widespread natural organic substance, making it an important biorenewable resource; in particular, cellulose can be the feedstock for the generation of renewable energy. In fact, the glucose obtained from cellulose hydrolysis can be used to produce ethanol by fermentation. The enzyme based approach is advantageous over chemical treatments also in this field. The main improvements are its higher conversion efficiency, the absence of substrate loss following chemical modifications and the use of more moderate and non-corrosive physical-chemical operating conditions, such as lower reaction temperatures, "neutral" pH and the use of biodegradable and non-toxic reagents. The production of bioethanol is particularly beneficial if it is obtained from cellulosic industrial or agricultural wastes (Sjöde A. *et al.* 2007; Chen Y. *et al.* 2007; Georgieva T.I. and Ahring B.K. 2007). This new promising approach could overcome the limits of traditional bioethanol production, which is performed using as feedstock agricultural biomasses like corn and sugarcane; these biomasses are specifically and intensively cultivated to produce bioethanol, diverting food resources and driving up food prices across the world. Moreover, intensive exploitation of cellulose as a biorenewable feedstock has to date been prevented by the lack of a suitable solvent that can be used in the chemical processes.
Rogers and co-worker have found that solutions of cellulose can now be produced for the first time at useful concentration using a new class of non-aqueous but polar solvents, named ionic liquids or "molten salts" (Swatloski R.P. *et al.* 2002); this new technology opens up substantial potential for cellulose processing and manufacturing of new cellulose-based materials. Ionic liquids are substances that are completely composed of ions and are liquid at or close to room temperature. Interest in these compounds, often heralded as the green, high-tech media of the future, is still increasing rapidly and stems from their near-zero vapor pressure, their thermal stability, and their widely tunable properties as regards polarity, hydrophobicity, and solvent miscibility behavior through appropriate modification of the cation and the anion (van Rantwijk F. and Sheldon R.A. 2007). Owing to their special properties and possible advantages, ILs may provide an ideal solvent for engineering media for biocatalytic reactions (Kragl U. *et al.* 2002); an increasingly number of studies are now focused to identify enzymes that are active in ionic liquids and to optimize these new biocatalysis reaction systems (Lee S.H. *et al.* 2008; Gorke J.T. *et al.* 2007)

endo-1,4-β glucanases

The endo-1,4-β glucanases cover 15 of the 113 families in which are actually divided the glycoside hydrolases (GH) on the base of amino acid sequence similarities; since the fold of proteins is better conserved than their sequences, some of the families can be further grouped in superfamilies or 'clans'. These enzymes, as well as the most of other glycoside hydrolases, show a modular structure; generally the proteins have a catalytic domain, which is reasonably large and represent more than 70% of the total protein, and a cellulose binding domain (CBD); these two domains are linked by a short amino acid sequence of varying length (6-59 residues), rich in proline and hydroxyamino acids which are often highly glycosylated (Bhat M.K. and Bhat S. 1997). Enzymes belonging to family 12 differ from this description, lacking the cellulose binding domain.

In the *S. solfataricus* P2 genome there are three ORFs coding for putative extracellular endoglucanases: SSO1354, SSO1949 and SSO2534. These enzymes belong to family 12 of glycoside hydrolases, clan C; SSO2534 (Limauro D. *et al.* 2001) and SSO1949 have been so far characterized and the *sso1949* gene expressed in *E. coli* (Huang Y. *et al.* 2005).

β-glucosidases

As previously described, β-glucosidases are involved in the hydrolysis of cellulose; consequently these are biotechnological relevant enzymes, finding application in several industrial processes. Recently, glycosidases are also used as base for the creation of new enzymatic activities, named glycosynthases, which are mutated versions of glycosidases (Perugino P. *et al.* 2004; Watts A.G. and Withers S.G. 2004). These enzymes efficiently synthesize oligosaccharides but do not hydrolyse them, and therefore can be used to create new biomolecules which have potential application as therapeutics. Considering their potentiality and application fields, in the last years β-glucosidases have been extensively studied. In particular, the *S. solfataricus* β-glucosidase (*lacS*) has undergone to extensive studies; a first crystallographic structure has been obtained (Aguilar C.F. et al. 1997), which has subsequently been integrated by further structural studies. This archeal protein, which is highly thermostable and thermophilic, has also been studied to elucidate the mechanisms allowing proteins to resist to high temperature. LacS belongs to GH family 1; it shows a tetrameric structure with a 240kD total molecular weight and the structure is a (beta/alpha) 8 barrel. The active site occurs at the centre of the top face of the barrel.

The lacS protein have a large ion-pair network located at the tetrameric interface of the molecule; circular dicroism, spectroscopy and mutagenic analysis have demonstrated that ion pair electrostatic interactions are fundamental for enzyme resistance and stability to high temperatures (D'Auria S. *et al.* 1998; Cobucci-Ponzano B. *et al.* 2002; Ausili A. *et al.* 2004), giving support to the current theories on the mechanism of ion pair stabilization in proteins from hyperthermophiles. Moreover, *lacS* has been expressed in *Saccharomyces cerevisiae* (Moracci M. *et al.* 1992), in *E. coli* (Moracci M. *et al.* 1995) and in mammal cells (Cannio R. *et al.* 1994); in these experiments the enzyme has been shown to retain the characteristics of thermostability and thermophilicity. The *lacS* gene is also widely used as marker in archaeal expression vectors, due to the simple method for the revelation of the protein gene expression product: a chromogenic artificial substrate, named X-gal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside) is used, which produce a blue coloration after β-glucosidase action. The activity is quantified by a spectrofotometric assay at 405nm using a p-nitrofenil-β-D-glucopyranoside. To efficiently use *lacS* as marker in vectors, defective cells in β-glucosidase activity can be employed in the transformation and expression experiments, in order to be sure that an eventually revealed activity is due to expression of the marker gene carried from the vector.

Concepts, methodology and objectives

The archaeal world strongly attracts the interest of scientists for several reasons; the attention is focused both to elucidate some physiological-molecular aspects that are still obscure and to give increasingly larger feasibility to biotechnological application of enzymes from (hyper)thermophilic organisms. The present PhD work is fitted in this dual context; on the one hand findings are reported about the characterization and tuning study of the *S. solfataricus* expression vector pMSSV, also able to propagate in *E. coli* cells (previously constructed in the same laboratory where this work has been performed); on the other hand a cellulase *S. solfataricus* gene was expressed both in xenologous and heterologous fashions, using in the former case the same pMSSV vector. The development of an effective expression vector for archaeal organisms is a key step to study their characteristics in depth. The genome sequencing progress provides a large quantity of DNA sequences and the function of the putative encoded proteins has been assumed by similarity analysis for some of them. In this contest an effective transformation-expression system can be very useful, allowing the gene product to be studied *in vivo* by homologus expression. Moreover, an expression vector could be a powerful tool to study the expression of heterologous proteins in archaeal hosts. The endolglucanases from hyperthermophilic organisms, as already explained, are biotechnogically relevant enzymes; in particular in *S. solfataricus* sequence among three cellulose sequences SSO1354 is the only one which remained still uncharacterized and unexpressed; a characterization of the product of *sso1354* gene could point out the applicative potentiality of this endoglucanase, allowing to make a comparison with the other characterized *S. solfataricus* endoglucanases. Therefore this work was developed through different steps:

- Cloning of *lacS* gene in the pSSVx-derived *E. coli-S. solfataricus* shuttle vector pMSSV, and characterization of this element.
- Cloning of *sso1354* gene in the pMSSV vector and expression of the endoglucanase in *S. solfataricus*
- Heterologous expression of *sso1354* gene in mesophilic hosts and characterization of the enzyme

Materials and Methods

Strains, enzymes and reagents used in this study

The reagents used for preparation of buffers and growth media of *Sulfolobus solfataricus* were supplied by Sigma Aldrich; the yeast extract and casaminoacids were supplied from Becton-Dickinson (BD). The reagents for polyacrylamide gels electrophoresis were supplied by Biorad. The restriction enzymes, the "modification enzymes" (alkaline phosphatase, T4 DNA ligase, T4 DNA polymerase) and the molecular weight markers for nucleic acids were supplied by New England Biolabs (NEB). The Phusion DNA polymerase was supplied by Finnzymes. The synthesis of the oligonucleotides was committed at PRIMM; the radioactive material was supplied by Perkin Elmer. tThe pET 28c vector was supplied by Novagen.
Escherichia coli mos-blue strain (*end*A1 *hsd*R17(r_{k12}-m_{k12}^{+}) *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac*[F′ pro$A^{+}B^{+}$ laqI^{q} ZΔM15: Tn*10* (Tet^{R})], used for cloning, were supplied by Amersham Pharmacia biotech; *E. coli* BL21-*Codon Plus* (DE3) RIL cells (F– *ompT hsdS*(rB–mB–) dcm+ *Tetr gal λ (DE3) endA The [argU ileY leuW Camr)*, and *E. coli arctic-express* (DE3) cells (*E. coli* B F^{-} *ompT hsdS*(r_B^{-} m_B^{-}) *dcm*$^{+}$ Tet' *gal* λ(DE3) *endA* Hte [*cpn10 cpn60* Gent'], used for overexpression, were supplied by Stratagene.
Kluyveromyces lactis GG799 competent cells and the pKLAC1 plasmid were supplied by NEB. Acetamide and YCB medium, used for selection of yeast transormants, were supplied by Becton-Dickinson (BD).
S. islandicus REY15/4 and the mutant derivative strains reported below were kindly provided by Dr. P. Contursi and Dr. Q. She at the University of Copenhagen. *Pyr*$^{-}$ mutants of two strains of *S. islandicus*, cured for pSSVx (REY31A) or for both SSV2 and pSSVx (REY 15A), was used as hosts for transformations with the modified pMSSV containing the *pyrEF* genes. *S. islandicus* REY31A is a pSSVx-cured strain isolated from REY15/4. *S. solfataricus* P2 (DSM1617) was purchased from the German Collections of Microorganisms and Cell Cultures. The strain *S. solfataricus* GθW, a spontaneous derivative mutant of Gθ strain lacking the β-galactodidase activity, was previously isolated in our laboratory (Bartolucci S. *et al.*, 2003).

E. coli transformation techniques

Transformation of *E. coli mos-blue* cells by electroporation

1μl of ligase solutions was added to aliquots of 40 μl of electrocompetent cells (previously preparated using the procedure described as "Transformation of *E. coli* by electroporation" in Sambrook J. and Russel D.W. 2001). The cells were incubated with the DNA for 1minute in ice, and then transferred into precooled electroporation couvettes (0.2 cm). The electroporator Gene Pulser^{TM} (Bio-Rad) was used, setting the following parameters: 25 μF, 2,5 kV, 200 Ohm. The transformed cells were immediately diluted with 1ml of LB, then transferred into a new 1.5 ml tube and incubated for 1h at 37° degrees under agitation.

Transformation of *E. coli* BL 21-*Codon Plus* (DE3) and *arctic-express* (DE3) cells

Aliquots of 200 µl of competent cells (previously preparated using the procedure described as "Preparation and transformation of competent *E. coli* using calcium chloride" in Sambrook J. and Russel D.W. 2001) were mixed with 50 ng of plasmidic DNA and stored on ice for ~20 min. The cells mixtures were transferred at 42 °C for 90 sec, on ice for two min and then diluted to 1 ml with LB medium. An incubation for ~1 h at 37°C under shaking was performed before plating the cells onto selective solid medium supplemented with 50 µg/ml kanamicin and 33 µg/ml chloramphenicol.

Proteins analyses

Determination of the protein concentration

The concentration of the proteins in solution was determined according to the Bradford' s method (Bradford, 1976). The Coomassie Brilliant (Bio-Rad) reagent was added to the samples and the absorbance at 595 nm was monitored. A solution of 1 µg/µl of bovine serum albumin (BSA) was used as standard.

Electrophoretic analysis of proteins (SDS-PAGE)

The electrophoresis on polyacrylamide gel in the presence of SDS (SDS-PAGE) was performed according Laemmli's protocol (Laemmli, 1970). The samples were denaturized at 100°C for 20 min in 2% SDS, 5% β-mercaptoethanol, 0.001% bromophenol blue and 10 % glycerol. The samples were then loaded on a 10% polyacrylamide gel and electrophoresed in 0.025 M Tris-HCl, 0.2M glycine pH 8.3 and 0.1% SDS. The electrophoresis was performed at 25 mA for ~1.5 hours; the proteins were then revealed by Coomassie Briliant-Blue; the gel was submerged in the staining solution (0.25% Coomassie Brilliant-Blue G250, 50% methilic alcohol and 10% acetic acid) for ~2 min with gentle agitation. The gel was washed in a solution containing 5% methilic alcohol and 7.5% acetic acid to remove the excess of Coomassie and then stored in 10% acetic acid. Alternatively the proteins were transferred by elettro-blotting from gel to a polivinildienefluoride (PVDF) membrane to sequence the N-terminal extremity of recombinant proteins (see below).

Protein Electroblotting

After electrophoresis on polyacrylamide gel in the presence of SDS, the proteins were transferred from gel to a polivinildienefluoride (PVDF) membrane by the following electro-blotting protocol. The PVDF membrane was submerged in methilic alcohol for 10 sec before using, to eliminate the hydrophobic properties, washed and equilibrated in transferring buffer (CAPS) for 5 min. The electro-blotting was performed for 2h (4°C) at 50 Volts using a Mini Trans Blot cell (BIORAD). Next, the membrane was washed with H_2O and submerged for few seconds in methilic alcohol. The proteins were then revealed by Coomassie Briliant-Blue: the membrane was submerged in the staining solution (0.25% Coomassie Brilliant-Blue G250, 50% methilic alcohol and 10% acetic acid) for 5 min. The membrane was washed in 50% methilic alcohol to remove the excess of Coomassie and then washed in H_2O. The

bands corresponding to the proteins of interest were cutted and stored in 1.5ml eppendorf tubes. The N-terminal sequences were performed by the protein sequencing core of IBP (Istituto di Biochimica delle Proteine, CNR, Napoli).

Cloning of the *lacS* gene into pMSSV and characterization of the vector

Growth of *Sulfolobus* strains

S. solfataricus strains P2 (DSM 1617), Gθ (Cannio R. *et al*., 2001) and the derivative mutant GθW [Δ*lacS,*] as well as *S. islandicus* REY 15/4 (Zillig W. *et al*., 1994) were grown at 75 or 80°C in TYS medium:

Ammonium sulfate	3g/l
Potassium sulfate	0,5g/l
Potassium chloride	0,1g/l
Glycine	0,7g/l
Yeast extract	0,5g/l

Brock's basal salts:

$MnCl_2 \times 4H_2O$	0,8 mg/l
$Na_2B_4O_7 \times 2H_2O$	2,1 mg/l
$ZnSO_4 \times 7H_2O$	0,11mg/l
$CuSO_4 \times 2H_2O$	25 μg/l
$Na_2MoO_4 \times 2H_2O$	15 μg/l
$VoSO_4 \times 5H_2O$	15 μg/l
$CoSO_4 \times 7H_2O$	5 μg/l
$NiSO_4 \times 6H_2O$	5 μg/l
$MgCl_2$	1 mM
$Ca(NO_3)_2$	0,3 mM
$FeSO_4 \times 7H_2O$	2 mg/l
HCl	0,1 mM

with 0.1% tryptone and 0.2% sucrose, at pH 3.2 adjusted using concentrated H_2SO_4. The shaking rate of the cultures was 150 rpm. For electroporation and plaque assays, cells were grown with phosphate buffered medium N.182 (M182) suggested by the DSMZ Catalogue of strains containing 0.1% glucose:

Ammonium sulfate	2,5 g/l
Magnesium sulfate	0,2 g/l
Potassium phosphate	3,1 g/l
Calcium chloride	0,25 g/l
Casaminoacids	1,0 g/l
Yeast extract	1,0 g/l

Basal salts:

$MnCl_2$ x $4H_2O$	1,8 mg/l
$Na_2B_4O_7$ x$10H_2O$	4,5 mg/l
$ZnSO_4$ x $7H_2O$	0,22 mg/l
$CuCl_2$ x $2H_2O$	0,05 mg/l
Na_2MoO_4 x $2H_2O$	0,03 mg/l
$VoSO_4$ x $2H_2O$	0,03 mg/l
$CoSO_4$ x $7H_2O$	0,01 mg/l

The pH of this medium was adjusted at 3,8 with H_2SO_4 10%. The optical density of liquid cultures was monitored at 600 nm. For solid media, gellan gum (Gelrite, Sigma) was added to a final concentration of 0.8% (0.35% for overlays), and $MgCl_2$ and $CaCl_2$ were added to 10.0 and 3.0 mM, respectively.

Viral DNA isolation and pMSSV*lacS* plasmid construction

Extrachromosomal SSV2 and pSSVx DNAs from both *S. islandicus* REY 15/4 and transfectants of *S. solfataricus* strains were performed with Qiaprep Spin Miniprep kit (Qiagen) following the standard procedure suggested by the manufacturer for *E. coli* cells.
An expression cassette of the *lacS* gene (Cubellis M.V. *et al.* 1990) was PCR amplified by the vector pMJ03 (Jonuscheit M. *et al.* 2003) and inserted between the XhoI and PstI sites of pMSSV generating the expression vector pMSSV*lacS*. Excision of the *E. coli* minimal plasmid was obtained by digestion of pMSSV*lacS* (1.0 mg) with SacI and purification of the pSSVx/*lacS* moiety from agarose gel. The DNA was re-circularized by ligation in a final volume of 5.0 μl, diluted in water to 20 ng/ml final concentration, and 2 μl were used for electroporation.

Transformation procedure and analyses of *Sulfolobus* transformants.

S. solfataricus cells were grown up to midlogarithmic phase (0.3–0.45 OD_{600}), harvested by centrifugation and repeatedly washed in 20 mM sucrose as described previously by Schleper *et al*. (1992). Aliquots of 1010 cells/ml (50 ml) were mixed with 1 ml DNA (10–100 ng/ml), incubated for 1 min on ice and transferred to chilled plastic cuvettes with an electrode gap of 0.1 cm (BioRad). High voltage electroporation (25 μF) was performed with a BioRad Gene Pulser Xcell™ at a field strength of 1.5 kV/cm and 400 Ω resistance; two successive shock pulses were applied to competent cells producing pulse length of ~10.0 and 9.8 ms, respectively. Immediately after electroporation cell mixtures were diluted with 1 ml of medium M182 containing 0.1% glucose, transferred to glass vials and incubated for 3 h at 75 or 80°C. After suitable scale-up, 5–15 ml aliquots of the cultures (the volumes varying in order to withdraw the same number of cells per aliquot) were harvested at

increasing cell density for DNA extractions. For monitoring propagation of the extrachromosomal elements, extrachromosomal DNA mini preparations and plaque assays were performed. Transformants were stored at ~80°C in 15% glycerol stocks.
For Southern blot analysis, ~2 µg of total cellular DNAs, extracted according to Arnold et al. (1999), and 5 ng of pMSSV*lacS* plasmid purified from *E. coli*, were cut with HindIII and BglII, and electrophoresed in a 0.8% agarose gel; DNA digests were blotted and hybridized according to standard procedures (Sambrook J. and Russel D.W. 2001). The probe was prepared by cutting out and purifying a HindIII restriction fragment from the pMSSV*lacS* vector encompassing the *lacS* gene and a portion (up to -302) of the tf55α 5' flanking region. This restriction DNA fragment was randomly labeled using the random prime DNA labeling kit (Boehringer Mannheim).

β-galactosidase complementation and isolation of mixed and single transformants

pMSSV*lacS* transformed cells were tested for β-galactosidase activity; 1 ml aliquots of cultures were centrifuged and cell pellets overlaid with a X-Gal solution (2 mg/ml in phosphate buffered medium) and incubated for 15 min at 75°C for blue color development. In situ assays were performed on the same cultures seeded on plates and grown as circular colonized areas, as previously described (Bartolucci S. *et al.* 2003). Supernatants of cultures grown up to 0.5, 1.0 and 1.3 OD_{600} were checked for plaque formation as already described and analyzed by X-Gal staining of plaques formed on continuous lawns of the *S. solfataricus* GθW strain.
For infection in liquid culture, 400 ml supernatants from pMSSV*lacS* transformants were added to a 20 ml culture of cells transfected only with SSV2 and grown up to 0.5 OD_{600}. After incubation under shaking for 48 h, cells were diluted 1:50, grown up to 0.36 and 1.0 OD_{600} and tested for β-galactosidase activity as already described.
Single transformants were selected either on rich (M182, glucose 0.1%) or on minimal (lactose 0.25% as the only nutrient) solid media, by plating or by streaking of electropored cells propagated until they showed positive staining with X-Gal and diluted after they had reached different cell densities (0.3–1.0 OD_{600}). Single clones formed after 2 weeks of incubation at 75°C were stained with X-Gal (incubation at 75°C for 3–4 h for color development on lactose plates), picked and resuspended in 100 ml lactose minimal medium and seeded as spots onto fresh rich (M182) medium plates.
After in situ X-Gal test on plates, cell spots were transferred to M182 medium containing glucose, propagated up to early stationary phase (1.0 OD_{600}) and analyzed for extrachromosomal DNA content as described above.

β-galactosidase enzyme activity

β-galactosidase activity of transformants was visualized and quantified as follows. Crude extracts were prepared by a freeze thaw technique, suspension of the cells in 10 mM Tris–HCl buffer, pH 8.0, placing at 80°C for 10 min, then at 50°C for 5 min. This procedure was repeated four times, and the extract was spun for 30 min at ~10000x g. The supernatant was either assayed immediately or stored at -80°C before testing.
Protein concentrations of the crude mcell extracts were determined by the Bradford assay (Bio-Rad). Detection of the β-galacosidase in protein extracts was performed on 10% acrylamide SDS–PAGE gels after electrophoretic separation and extensive

washing in 10 mM Tris–HCl buffer, pH 8.0. The specific enzyme band was mnvisualized incubating the gel 30 min at 75°C in the same buffer containing 2 mg/ml X-gal. The β-galactosidase assay procedure followed essentially the protocol of Pisani et al. (1990). A sample of 10 µl extract was transferred to a preheated (75°C) quartz cuvette containing 990 µl assay buffer (2.8 mM ONPG in 50 mM sodium phosphate buffer, pH 6.5). The ONPG hydrolysis reaction was followed spectrophotometrically at 75°C by measuring the increase in absorbance at 405 nm in a Beckman spectrophotometer with heatable cuvettes. One unit was defined as the amount of enzyme catalyzing the hydrolysis of 1 mmol of ONPG min^{-1} at 75°C with a molar absorption coefficient of 3100 M^{-1} cm^{-1} at 405 nm for ONP.

SSV plaque assay

The virus titre of a culture was determined by means of plaque assay using *S. solfataricus* P2 as an indicator strain according to Schleper *et al*. (1992). The plates were incubated for 2–3 days at 80°C and examined for the appearance of the turbid halos which the inhibition of host growth by SSV2 virus generates.

sso1354 gene expression in *S. solfataricus*

Plasmid constructions

mos-blue cells were used for all clonings in *E. coli*. The sequences of *sso1354* gene and *araS* promoter were amplified by PCR using the primers 1354-Bsph (TAAAGTAGGA**TCATGA**ATAAATTATATATTG), 1354-His (C**GTCGAC**TTAATGGTG ATGATGGTGATG**CTCGAG**GAGAGTTTCAGA); *araS*-Fw(TAAAACCCCAT**CTCGAG** AAATAATGAGGAG), *araS*-Rv (CTACGCCTAG**CCATGG**TCTCGGGTACTTTTATG). The primers were designed to insert a BsphI restriction site (bold letters) at the start of SSO1354 coding sequence and a sequence coding for a 6x(His) tag at 3' end; upstream and downstream of the His tag sequence a XhoI and SalI restriction sites (bold letters) were inserted, respectively. *araS* primers were designed to insert a XhoI and NcoI sites (bold letters) at the 5' and 3' ends. The *S. solfataricus* P2 genomic DNA was used as the template for amplifications. After amplifications, the fragments *sso1354* and *araS* were purified (with the Stratagene DNA purification kit) and digested with the enzyme mixtures BsphI/SalI and NcoI/XhoI, respectively. The restricted fragments were ligated with the pGEM T-Easy vector, previously digested with Xho I and Sal I. After isolation and partial sequence of the correct construct, the expression cassette *araS*1354 was excised from the pGEM-*araS*1354 with ApaI and NsiI restriction enzymes, and ligated with the 5' dephosphorylated pMSSV shuttle vector previously made compatible by digestion with ApaI and PstI.
A modified version of pMSSV, containing the *pyrEF* genes was also constructed. As a first step, an expression cassette containing the *glcS* promoter and the SSO1354 coding sequence was constructed. The *glcS* sequence was amplified using the primers *glcS*-Fw (CCCAATAACTA**CTCGAG**TTACTGACAACTC) and *glcS*-Rv (CTTCCTTTT CATGACAATTTATGGTAACC). The *sso1354* gene was amplified using the pGEM *araS*-1354 vector as the template in order to give a 5' blunt extremity; the primers were 1354-Fw (AATAAATTATATATTGTGCTTCCGG) and sp6 (ATTTAGGTGACACTATAG). The *glcS* amplification primers were designed to insert a XhoI site (bold letters) at the 5' end and to amplify the promoter sequence plus the

first three codons with the 3' extremity blunt. *S. solfataricus* P2 genome was used as the template for amplification of the *glcS* sequence. After amplifications, the fragments were purified (with the Stratagene DNA purification kit) and restricted with the following enzymes: *sso1354*, NsiI; *glcS*, XhoI; The vector pGEM T-Easy was digested with XhoI and NsiI. These three fragments were ligated and the correct vector pGEM T-Easy *glcS*-1354 obtained. The expression cassette was excised from this plasmid with SalI enzyme and ligated with *pyrEF* amplified sequence also digested with SalI. The *pyrEF* sequence had been amplified from the *S. solfataricus* P2 genome with the primers pyr-Fw TTTTGCTATCGAA **GTCGAC**TCCTCGGATGCAAAT and pyr-Rv CACACTTCTACCCTT**GTCGAC**TAAA TCGTTTTCG so that the SalI restriction site (bold letters) was inserted at both ends. The desired pGEM T-Easy vector containing the *glcS*-1354 cassette and the *pyrEF* genes was selected. The ApaI/NsiI fragment comprising the expression cassette and the *pyrEF* genes was ligated with the pMSSV shuttle vector made compatible with ApaI and PstI and dephosphorylated.

Expression and purification of *sso1354* gene

S. solfataricus GθW cells were transformed by electroporation with 50ng of pMSSV*araS*1354 vector (following the procedure described above). Transformed cells were scaled up in 182 medium up to 50ml. The cells were pelleted and resuspended in an equal volume of Brock's basal salts medium supplemented with 0.25% arabinose. After a scale-up to 500ml, the cells were grown up to 1.2 OD_{600} and harvested by centrifugation. The proteins of the supernatant were precipitated with ammonium sulphate (90% saturation); the pellet was resuspended in Tris-HCl buffer (25mM, pH 7.5), and the protein sample extensively dialyzed against the same buffer. After dialysis, the sample was loaded onto a resource Q anion exchange column (Amersham). After eluition (the SSO1354 protein was eluted at 500 mM NaCl), the recombinant enzyme carrying the 6x(His) tag was purified by nickel affinity chromatography (HIS-Select Spin Columns, Sigma).The active fractions (around 250 mM imidazole concentration) were pooled and dialyzed against Tris-HCl buffer pH 7.5.

CsCl gradient analysis

The supernatant from 500-ml coltures of *S. solfataricus* GθW cells transformed with pMSSV*araS*1354 vector was filtered through a 0.22-µm Seritop filter (Millipore). The filtrate was concentrated by ultrafiltration through a membrane (Amicon, Millipore) with a cut-off of 30 kDa, until the retained volume was ~15 ml. Further purification on a CsCl gradient (0.45 g/ml) by centrifugation for 48 h (SW60-Ti rotor, 250,000x g, 4°C, Beckman Optima LE-80K ultracentrifuge) was performed, and the opaque bands obtained were isolated and dialyzed extensively against 5 mM Tris-HCl buffer (pH 7) at 4°C.

Membrane preparation

Wet cells from 500 ml Brock's basal medium supplemented with 0.25% arabinose, harvested in the stationary phase, were suspended in 10 ml 50 mM Tris-HCl pH 7.0, and ground in a mortar with sand for 1 h. After centrifugation at 2,000x g for 10 min in order to remove sand and unbroken cells, the supernatant was ultracentrifuged at

55,000x g for 30 min. The clear crude extract was stored at 4°C, while the pellet, containing membrane fragments, was suspended in 25 ml 50 mM Tris-HCl pH 7.0 and ultra-centrifuged again. The clear pellet was resuspended in 10 ml of the same buffer containing 0.5% Triton X-100 and incubated overnight at 70°C.
After incubation, the suspension was ultracentrifuged as described above. The pellet was discarded and the supernatant was extensively dialyzed against 25 mM Tris-HCl pH 7.0.

sso1354 gene expression in *E. coli*

Plasmid constructions

The whole SSO1354 coding sequence plus a sequence coding for a C-terminal 6x(His) tag was amplified by PCR with the primers 1354-Bsph and 1354-His (see above in "Expression of *sso1354* gene in *S. solfataricus*" pharagraph). *S. solfataricus* P2 genome was used as the template for amplification. After amplification, the fragment was purified (with the Stratagene DNA purification kit), digested with the BsphI and SalI enzymes and cloned into pET-28c (previously made compatible with the NcoI and SalI enzymes).
The DNA sequence to obtain a N-terminally truncated SSO1354 version, namely a protein lacking the hydrophobic region (amino acids 1-27) was cloned as follows: a PCR fragment obtained with the forward primer 1354-NheI GAT**GCTAGC**CAGTCTCTCAGCGTTAAACCCGTAACTA and the reverse primer 1354-HindIII GGTCTTAG**AAGCTT**ATATTGTTTAGAGGAGAG was cut with NheI and HindIII and cloned into the vector pET-28c (the restriction sites are indicated as bold letters and the three nucleotides corresponding to the codon 28 of SSO1354 coding sequence are underlined).

Expression and purification of *sso1354* gene

The expression plasmids were used to transform *E. coli* BL21 (*codon plus*) cells and *E. coli arctic-express* cells. For expression, cells were grown overnight in 10 ml of LB medium with 50 µg/ml kanamycin and 50 µg/ml chloramphenicol at 37°C. After scale up to 1-litre colture, the growth was continued up to an A_{600} of about 0.6-0.8 and IPTG was added to a final concentration of 0.4 mM. The BL21 cells were incubated for further 16 h at 22°C, while the arctic-express cells were grown for 40h at 10°C. Cells were harvested by centrifugation, resuspended in 50 ml of buffer [100 mM Tris/HCl, pH 7.5, 300 mM NaCl, 2 mM 2-mercaptoethanol, 0.7 mM PMSF, 5% (v/v) glycerol], and disrupted by sonication for 5 cycles of 1min each (B. Braun sonicator). The recombinant cellulase versions carrying the 6x(His) tag were then purified by Nickel affinity chromatography. Both intact and N-terminus truncated SSO1354 enzymes were eluted at 250 mM imidazole concentration from a HIS-Select spin column (Sigma) and dialysed against Tris-HCl buffer pH 7.5.

sso1354 gene expression in *K. lactis*

Plasmid constructions

The SSO1354 coding sequence starting from the 28° codon was amplified by PCR with the primers 1354-Fw-Yeast (GAT**CTCGAG***AAAAGA*CAGTCTCTCAGCGTT AAACCCGTAACTACA) and 1354-Rv-Yeast (GGTCTT**AGATCT**TATATTGT TTAGAGGAGAG); these primers were designed to amplify the coding sequence from the 28° codon (underlined), and to insert XhoI and BglII restriction sites (bold letters) respectively upstream and downstream of the SSO1354 sequence. The coding sequence for the recognition site of the yeast Kex endoprotease (in italics letters, needed for the correct secretion of the recombinant proteins) was also added, immediately upstream to the SSO1354 coding sequence. After amplification, the fragment was purified (with the Stratagene DNA purification kit) and digested with the XhoI and BglII enzymes. The adapted sso1354 fragment was ligated with the pKLAC vector, previously digested with the same enzymes.

Expression and purification of *sso1354* gene

Cells of *K. lactis* GG799 strain were transformed with pKLAC-1354 vector following the protocol provided by the supplier (New England Biolabs) and also available at the http://www.neb.com/nebecomm/products/productE1000.asp web site. One out of 12 transformed clones was selected after 4-day growth in 10ml YPG medium and cellulase assay of the supernatants. A 10-ml colure of this clone was scaled-up to 1 liter and cells grown up to about 45 OD_{600}. The cells were centrifuged and the supernatant was 20-fold concentrated times using a stirred ultrafiltration system (Amicon, Millipore) with a YM 10 membrane. The proteins in the concentrated supernatant were precipitated with ammonium sulphate (90% saturation); the recombinant SSO1354 enzyme do not precipitate in ammonium sulphate, so the supernatant of the precipitation was further processed, namely extensively dialyzed against pH 7.5 Tris-HCl buffer (25mM), aliquoted and stored at -20°C.

Enzyme assays and analyses

Activity gel

The protein preparations was separated by SDS/PAGE in gels containing 0.1% (w/v) CMC (carboxymethylcellulose; Sigma). The protein was then renatured by several washes, once with a mixture (1:4, v/v) of propan-2-ol and 20 mM potassium phosphate buffer (in the pH range 1.8-2.5) or, alternatively, 20 mM phosphate citrate buffer (in the a pH range 3.0-5.0) for 30 min and, subsequently, with the same buffers without propanol for 30 min. Finally, the gel was incubated in 10 mM potassium phosphate or phosphate citrate buffer at 75◦C for 1 h with one buffer exchange. All gels were equilibrated with 50 mM potassium phosphate (pH 7) for 30 min, stained with 0.1% (w/v) Congo Red (Sigma,) for 30 min and destained with 1 M NaCl.

Enzyme assays

For enzyme activity, two assay methodologies were used. In the first assay, the soluble chromogenic substrates (AZO-polymers, polysaccharides dyed with RemazolBrilliant Blue R, Megazime). This assay was used to follow the enzyme activity during purification, using as substrate the AZO-CM-Cellulose and to determine the enzyme specificity towards pachyman and oat-spelt xylan polymers. The activity was measured by adding 170 µl of the specific 1% AZO-polysaccharide solution in 50mM phosphate-citrate buffer to 170µl of enzyme solutions and incubating the reaction mix at 80° C for 5–30 min at pH 5.0; pH was measured and adjusted in the final buffer–substrate mix. The reaction was stopped by addition of 850 µl 96% ethanol to the mixture, followed by incubation at room temperature for 10 min and centrifugation at 1,000xg for 10 min. The absorbance of the supernatant was measured at 590 nm. One unit of enzymatic activity (RBB-unit) was defined as the amount of enzyme required to increase the absorbance at 590 nm of 1 Δ/min under standard conditions.
Alternatively, enzyme activity was measured by determining the amount of reducing sugars released from the polysaccharide substrates. This assay was used also to evaluate the enzyme specificity toward debranched arabinan, curdlan and lichenan and to determine the pH and temperature optima. The standard reaction mixture consisted of 50 µl 1% polysaccharidic substrate in 50 mM phosphate citrate buffer, pH 5.0 and 50 µl of enzyme solution (the pH was always measured and adjusted in each final buffer–substrate mix); for the pH dependence, two buffers were used: potassium phosphate buffer in the the pH range 1.8-2.5 and phosphate citrate buffer in the pH range 3.0-7.5. After 10-60 min incubation at 80°C, the reactions were stopped on ice and the amount of reducing sugars released was measured at 520 nm by the Somogyi-Nelson method (Nelson 1944). One unit of enzymatic activity (SNunit) was defined as the micromoles of sugars released per minute per milliliter.

Reactions in ionic liquids and analysis of degradation products by TLC

Enzymatic reactions were performed in 2 hydrophilic ionic liquids, 1-ethyl-3-methyl-imidazolium-acetate and 1-ethyl-3-methyl-imidazolium-hydrogen sulfate. Solutions of 1 % α-cellulose were used for the assays. The assays were performed at 80°C for various incubation times adding a premixed solution of 30 µl of pure ionic liquid and 20 µl of enzymatic solution to 50 µl of ionic liquid-cellulose. Both solution were prewarmed at the assay temperature before mixing in order to reduce viscosity.
The assay was blocked by chilling and cellulose precipitated by adding 100 µl of H_2O. After removal of unsoluble cellulose by centrifugation, the hydrolysis products were separated from ionic liquids and protein by three water saturated phenol/chloroform (1:1) and one chloroform extractions. After extraction, a solution of ~150 µl was obtained. Aliquots (60 µl) were spotted on a silica 60 TLC plate (Macherey-Nagel), which was developed in ethyl acetate, H_2O, acetic acid, isopropanol, formic acid, (25:15:10:5:1, by vol.) for ~2 h. Reducing sugars were stained with the α-naphtol reagent after a 5-min incubation at 150°C.

Results

1. Characterization of pMSSV shuttle vector: cloning and expression of the *lacS* gene

Cloning of the expression cassette *tf55αlacS* in the pMSSV vector

Beta-D-galactosidase activity of *S. solfataricus* is displayed by the *lacS* gene product (Pisani F.M. *et al.* 1990; Grogan D. *et al.* 1991) and is responsible for the typical blue stain of the cells when exposed to the chromogenic substrate, X-gal; the lack of gene function in defective mutant strains is hence responsible for colorless colonies (Schleper C. *et al.* 1994). An expression cassette carrying the *lacS* coding sequence fused to a 448 bp fragment containing the promoter region of tf55α plus upstream sequences and the first five codons of its ORF was inserted in the polycloning site of the pMSSV vector by a directional cloning (*tf55α* is an heat shock inducible promoter, in fact *tf55α* gene encode for the α subunit of the tf55 thermosome, a protein involved in the heat shock response) (Fig.9).

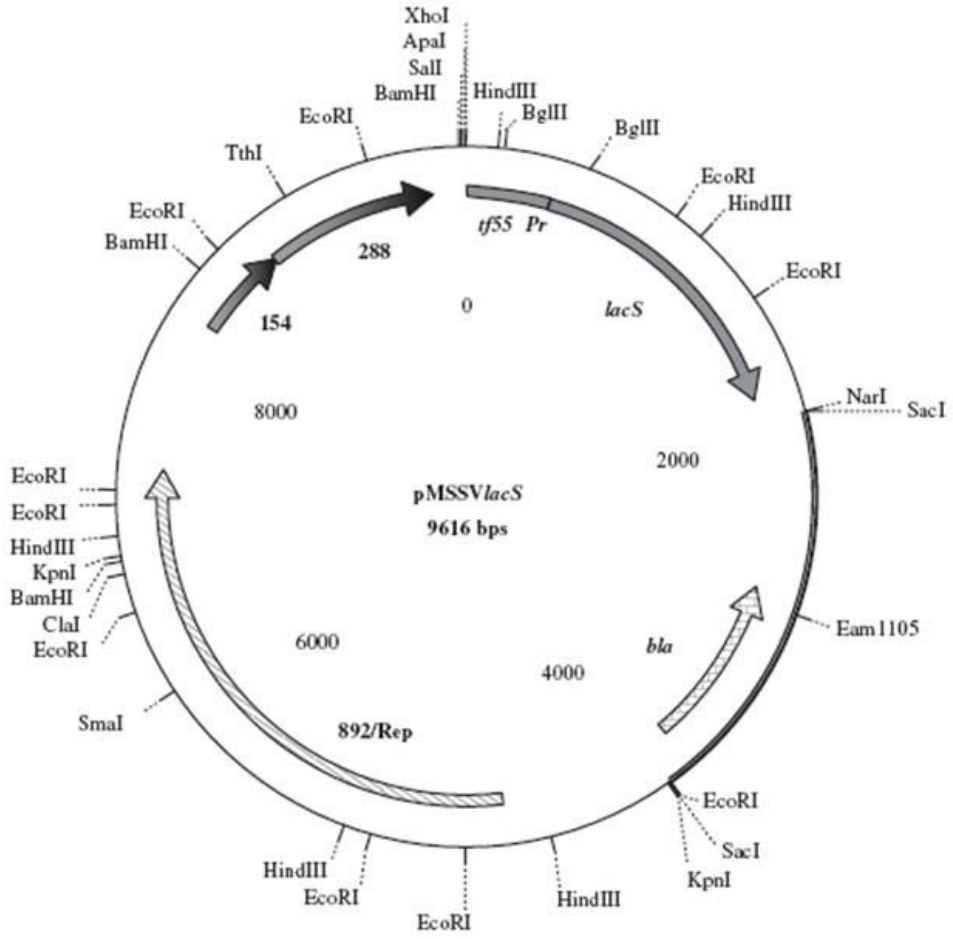

Fig. 9 Map of pMSSV*lacS*. An expression cassette containing the *lacS* gene fused to the thermosome tf55α subunit promoter was inserted into the shuttle vector pMSSV producing the expression vector pMSSV*lacS*.

Both the cassette and the vector were already available; the vector, as previously described, had been constructed on the base of pSSVx, at which had been added

the sequences essential for replication in *E. coli* from the plasmid pUC19, while the cassette had been provided by Dr C. Schleper and had been already demonstrated to efficiently complement a beta-galactosidase defect in *Sulfolobus* when carried by a viral SSV1- derived shuttle vector (Jonuscheit M. *et al.* 2003; Albers S.V. *et al.* 2006). The pMSSV*lacS* plasmid was introduced in *E. coli mos-blue* cells by electroporation; all the clones analyzed were recombinant, with no rearrangements, and the cells showed a regular growth. pMSSV*lacS* was subsequently transferred into *S. solfataricus* GθW cells, already transfected stably with the helper virus SSV2, by electroporation; the GθW strain was chosen because it is a stable beta-galactosidase mutant with an extended deletion in the *lacS* genetic locus (Bartolucci S. *et al.* 2003).

lacS expression in the β-galactosidase deficient mutant GθW

Transformed GθW/SSV2 cells were propagated, and then extrachromosomal DNAs were prepared and analyzed by agarose electrophoresis (Figure 10). The presence and growth-dependent accumulation of the vector in GθW/SSV2 revealed successful transformation and DNA replication; moreover the *lacS* gene is harmless and does not induce recombination of the vector. Identical results were obtained when SSV2 lysogens of *Sulfolobus* GθW were transformed with the plasmid pSSVx*lacS* lacking the *E. coli* sequences and obtained by cleavage with SacI and re-ligation (Fig. 10).

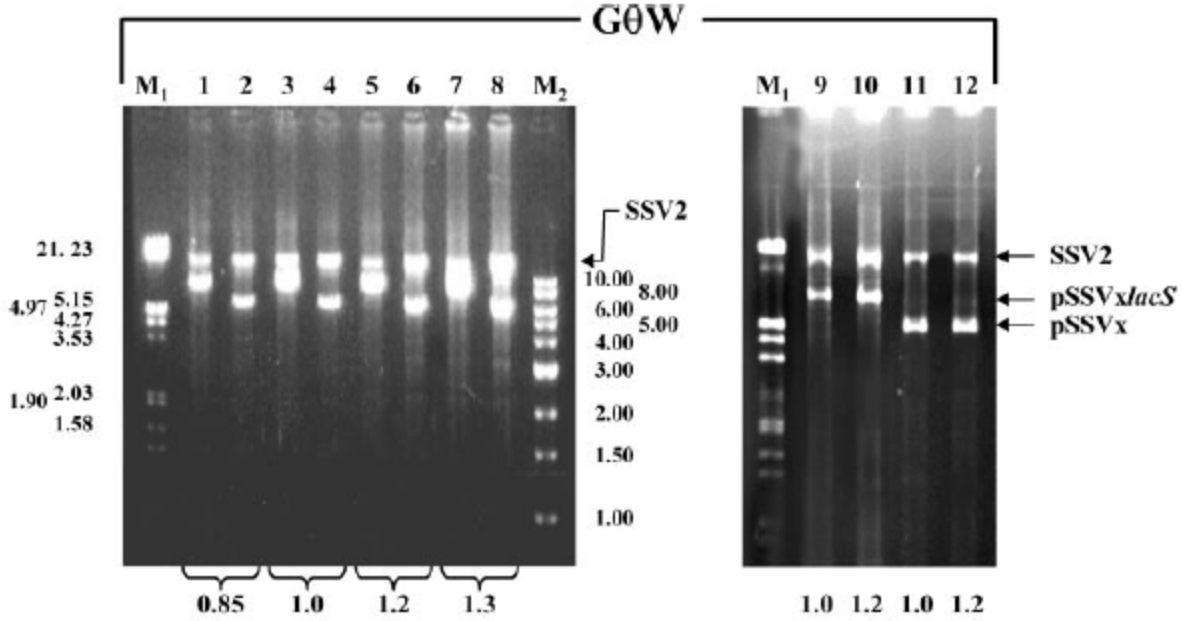

Fig. 10 Transformation of *S. solfataricus* with the vectors pMSSV and pMSSV*lacS*. Competent cells of *S.solfataricus* GθW infected with the SSV2 virus were electroporated with the vector carrying the β-glycosidase gene pMSSV*lacS* and also with pSSVx*lacS*, a pMSSV*lacS*-derived plasmid obtained by excision of the *E.coli*-specific sequences and religation. pMSSV and wild-type pSSVx were also transferred into the GθW/SSV2 lysogen for comparison of vector transfer and propagation efficiency in the presence and absence of the *lacS* gene. Extrachromosomal DNAs extracted from transformed cells after propagation of the cultures were checked by agarose electrophoresis for the presence of pMSSV*lacS* (lanes 1, 3, 5 and 7) and pMSSV (lanes 2, 4, 6 and 8) at the different growth stages indicated as cell densities (OD_{600}). Similarly, recovery of the pSSVx*lacS* plasmid (*SacI* cut, lanes 9 and 10) was monitored in comparison to the parental pSSVx (*Sma*I cut, lanes 11 and 12). Mobility of the virus SSV2 is highlighted. M1 and M2: molecular weight markers (Kb)

All pSSVx derived plasmids failed to transform *Sulfolobus* if linearized prior to transfer. In fact, X-Gal staining test was positive on *Sulfolobus* GθW/SSV2 transformed with pMSSV*lacS* on liquid-cultured (Figure 11A), plated (Figure 11B), and primary infected cells (Figure 11C). Plaques depicted in Figure 11C stained after incubation with the chromogenic substrate as a result of propagation of the engineered satellite virus. Since cell growth retardation in the plaques is directly proportional to the SSV2 virus titer which is maximal in the stationary phase cultures, turbidity and color faded progressively with increasing cell density.

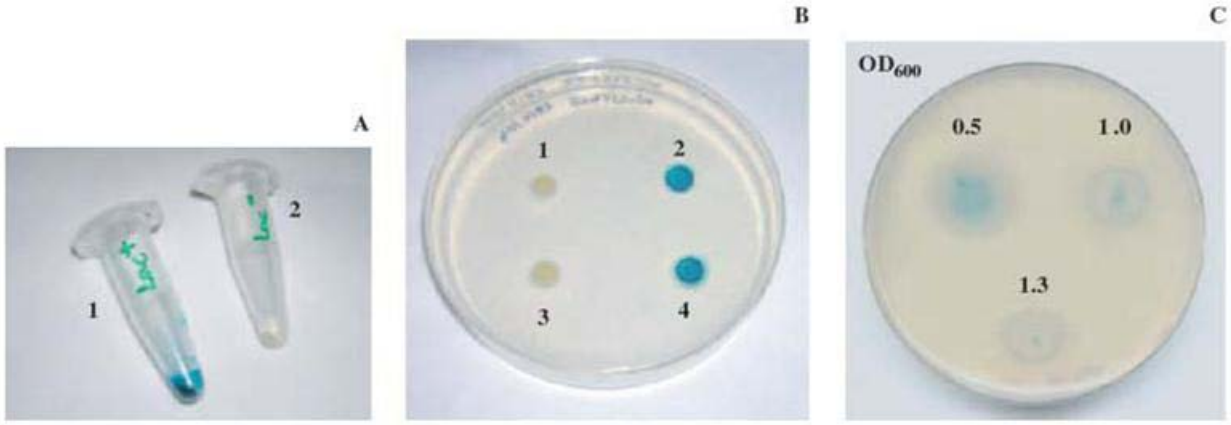

Fig. 11 Complementation of the β-glycosidase mutation in *S. solfataricus* GθW and spreading of the pMSSV*lacs* vector. The expression vector pMSSV*lacS* was transferred into *S. solfataricus* GθW cells lysogens for SSV2. Successful transformation was checked after propagation for several generation by direct exposure of cell pellets from liquid cultures to X-Gal and development of the blue color [**A** (1) Lac^+, pMSSV*lacS* transformed cells; (2) Lac^-, cells transfected with pSSVx]. Maintenance of the plasmid was also confirmed for colonized areas on plates and X-Gal test [**B** (1) Strain GθW; (2) wild-type strain Gθ; (3) SSV2/ pMSSV-infected GθW; (4) SSV2/ pMSSV*lacS* -infected GθW].
C: 4-ml aliquots of culture supernatant of SSV2/pMSSV*lacS*-infected GθW, withdrawn at the different cell densities indicated, were spotted onto a continuous lawn of uninfected GθW cells. The spreading of the recombinant satellite particles was revealed by the blue color developed on the plaques (primary infected cells) upon X-Gal exposure. Stain intensity depended on the extent of cell growth retardation in the plaques and was progressively less intense for increasingly higher virus titers.

Therefore, the β-galactosidase was expressed from the recombinant plasmid and the engineered satellite virus had spread efficiently throughout the culture. pMSSV*lacS* DNA prepared from *S. solfataricus* could be re-transferred into *E. coli* without suffering recombinational rearrangements.
Plasmid preparations and total DNA from *Sulfolobus* transformants were analysed by restriction analyses and Southern hybridization that confirmed the maintenance of the vector. No vector integration into the host chromosome occurred since the recombinant vector exhibited the same restriction pattern as the DNA prepared from *E. coli* and no signal relative to the plasmid could be detected on genomic DNAs in the Southern blots (Figure 12). Bands corresponding to chromosomal tf55α promoter and *lacS* gene (absent in GθW) could also be visualized as internal controls for detection of single copy chromosomal sequences. A similar Southern analysis confirmed that the pMSSV*lacS* copy number varied from 10–15 (mid-log phase) to 130 molecules per cell (stationary phase).

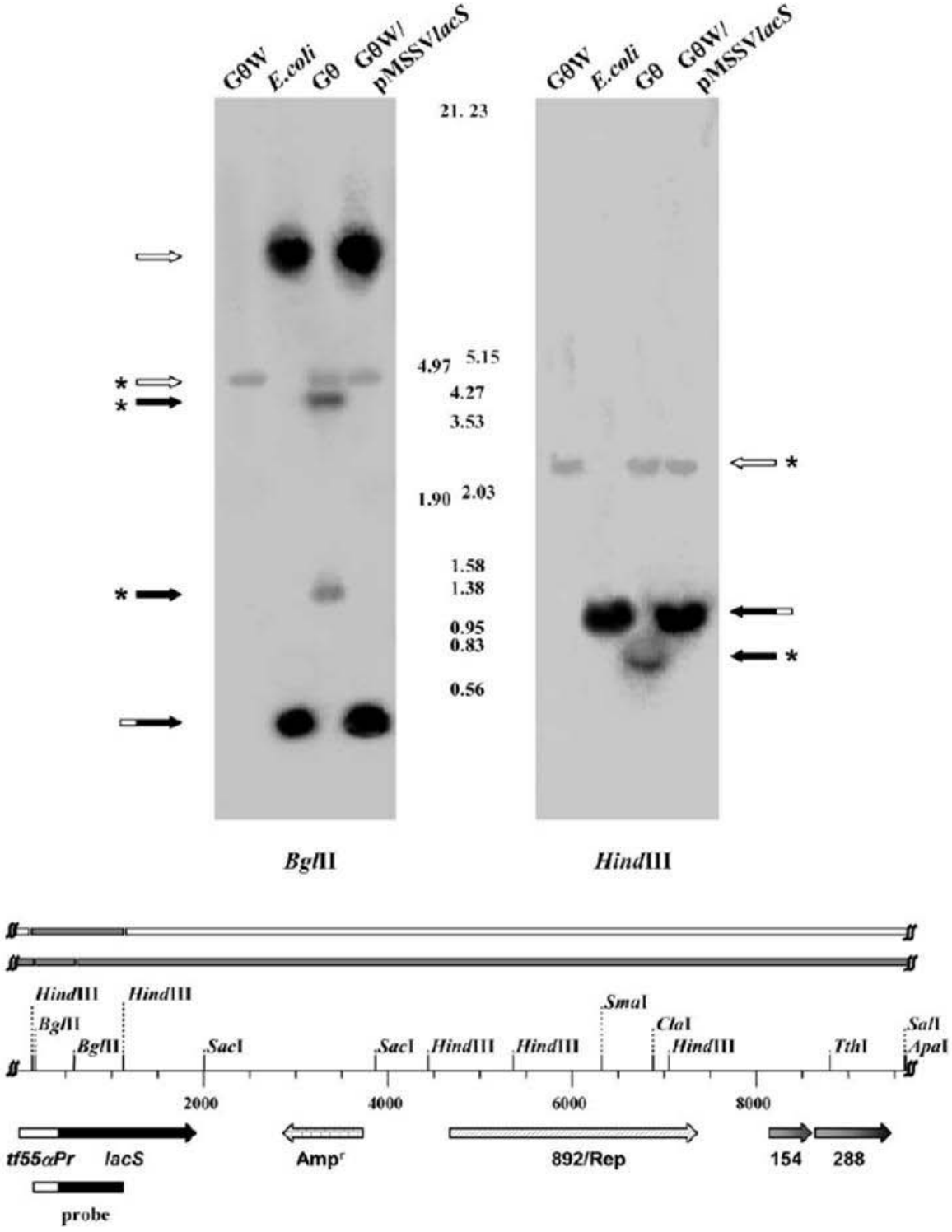

Fig. 12 Southern analysis of pMSSV*lacS* transformants. For hybridization, total DNAs from transformant cells (GθW/pMSSV*lacS*) and from the recipient mutant strain GθW as well as from wild-type cells Gθ were cut with *Bgl*II and *Hind*III as indicated. The recombinant pMSSV*lacS* vector prepared from *E.coli* before transformation was used as a reference for correct restriction patterns (molecular weight standards are indicated). White and black arrows indicate hybridization to tf55α and *lacS* gene sequences, respectively, the asterisks distinguishing signals of the chromosomal copies. Signals of the tf55α–*lacS* gene fusion on the vector are highlighted by white/black arrows. The scheme on the bottom represents the restriction patterns for the two enzymes on the map of the linearized pMSSV*lacS* and the fragments producing positive signals for hybridization (grey bars).

β-galactosidase assays on pMSSV*lacS* transfectants

Detection of β-galactosidase activity on denaturing gels by enzyme staining revealed that the *lacS* gene product was indistinguishable when expressed as heterologous in *E. coli* and xenologous in *S. solfataricus*, namely the *Sulfolobus* expression system did not interfere with the correct polypeptide syntesis (data not shown). β-galactosidase activity was measured using a spectrophotometric assay with the specific substrate ONPG in crude cell extracts of both transfected mutant GθW strain and wild type Gθ. The mutant recipient strain GθW has been shown to exhibit no detectable activity (Bartolucci S. *et al.* 2003), whereas the specific activities of pMSSV*lacS* transformants in primary transformation mixtures rose from undetectable levels to 1.2 U/mg protein; this value was ~2-fold higher than that found in wild-type cells expressing *lacS* under the control of its own promoter. The activity remained stable in diluted and propagated cells cultured as described above, when assays were performed at the same growth phase. Under identical growth conditions, the value of endogenous β-galactosidase activity in the P2 wild-type strain (i.e. the natural source of both the *lacS* gene and tf55a promoter sequences inserted in the pMSSV*lacS*), reached 0.1 U/mg as the highest value. Therefore, using this value as reference the expression level in GθW/pMSSV*lacS* is ~12-fold higher. For a single culture, activity increased up to late logarithmical growth phase and then maintained approximately the same value up to late stationary phase (Figure 13). This result confirmed that the recombinant satellite viral DNA was replicated inside the cells and that virus particle formed and spread throughout the culture; the increasing activity should therefore depend only on the copy number of the plasmid and hence on the number of the *lacS* gene copies per cell. Heat shock of the stably transfected cells, shifting the culture temperature from 75° to 88°C, induced an increase of the specific activity up to 2.5-fold (3.0 U/mg) after 3 h and remained constant in cells exposed to thermal stress for 24 h.

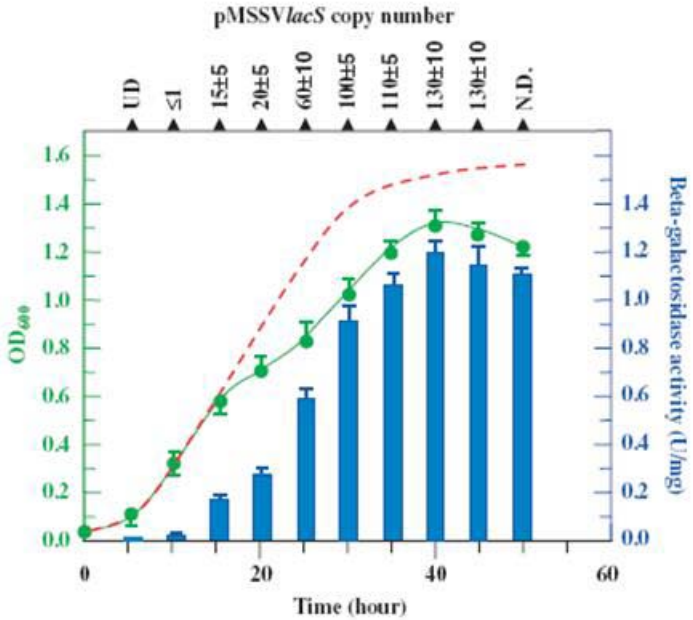

Figure 13. β-galactosidase activity and plasmid copy number in cultures of pMSSV*lacS* transformants. The β-galactosidase activity was measured in crude extracts of pMSSV*lacS* transformants harvested at different optical densities from a single culture. Specific activity (indicated by solid bars) is plotted together with the growth curve and the relative plasmid copy number. All values are the average of independent experiments performed on three different cultures and standard deviation is indicated for optical density and activity as well as for plasmid copy number. The dashed red line indicates a typical growth curve of the empty strain GθW. UD, undectectable; N.D., not determined.

2. Xenologous and heterologous expression of endoglucanse *sso1354* gene

Sequence analysis

Three putative extracellular endoglucanase sequences have been found in the complete genome sequence of *S. solfataricus* P2 (She Q., *et al.* 2001): *sso1354*, *sso1949* and *sso2534*. This work have focused on the still uncharacterized SSO1354 gene/protein. The first approach was to perform a deep sequence analysis using bioinformatic tools. The proteic sequence translated from *sso1354* gene was alligned with sequences avaiable in databanks using FASTA 3 program on EBI (European Bioinformatics Institute), website (http://www.ebi.ac.uk/services/index.html). The highest score was produced in the comparison with SSO1949, which shows 85% identity with SSO1354 (Fig.14) .

```
          1                                                          60
SSO1354 M--NKLYIVL -PVIVIIAIG VMGGIIYLHQ QSLSVKP--V TTTEFSTTTS TSTTTNAITT
SSO1949 MIMNKLYIII VPIIVIIVVG VIGGAIYLHH QSPNVKTSSI TVTTNETTTL MSITTNTVPT
SSO2534 M--NKLIPIF --VVVIIVLG II-------- ---------V SIEFGKFHQN ASLTRSTERF

          61                                                         120
SSO1354 TVTQTVTSIT SYNQLIYVTS SASSPTPVYL NNSTIPSFYL EVNMWNAKNY NGNYTMVFNP
SSO1949 TVTPTTSSIP ---QLIYVTS SASSPTPVYL NNSTVPSFYL EVNMWNAKTW NGNYTMVFNP
SSO2534 TLFPAHNRPF ---SVLGNYS SNSADALAIL NSSTNATLMV SPFLWNIGYA LGNVNMTIN-

          121                                                        180
SSO1354 LARTLSVSFN LTQVKPLEWT --NGYPEIYV GRKPW----- -DTAYAGNIF PMRIGNMTPF
SSO1949 LTRTLSVSFN LTQVNPLQWT --NGYPEIYV GRKPW----- -DTSYAGNIF PMRIGNMTPF
SSO2534 -INYLHVAIN LSQISKISSN VVDGYPGLMY GQELWWPFMY RTTQLQFLSL PMIVLRLPNF

          181                                                        240
SSO1354 MVSFYINLTK LDPSINFDIA SDAWIVRPQI AFSPGTAPGN GDIEIMVWLF -SQNLQ--PA
SSO1949 MVSFYINLTK LDPSINFDIA SDAWIVRPQI AFSPGTAPGN GDIEIMVWLF -SQNLQ--PA
SSO2534 YSILNYSVYL INGSID-DFS YDIWLSQNPN I----TSLQY GDFEIMIWMY WNENLSHTPY

          241                                                        300
SSO1354 GEQVGKVVVP IYINHTLVNA TFQVWEM-KS VPWGGWEYIA FRPD------ GWKVTNGYVS
SSO1949 GQQVGEVVIP IYINHTLVNA TFQVWKM-KN VPWGGWEYIA FRPD------ GWKVTNGYVA
SSO2534 FIYVGNMSIP TLINGKIENL SWEVYVLPRT GSANGWTGVY FLSPLKEPKA EFGVPIGYIL

          301                                                   356
SSO1354 YEPNLFIKAL SNFTSYNITN YYLTDWEFGT EWGTMTSNGT AYFSWTVSNF SETLL
SSO1949 YEPNLFIKAL NNFASYNITN YYLTDWEFGT EWGTMTSNGT AYFSWTISNF YETLL
SSO2534 KNMGSYIEK- AGVNIYNVNT YYLDAIQVGM EFSD--NQGT AIMGYYLYSW QIWLLS
```

Fig. 14 Alignment of SSO1354 protein sequence with the other two endo-1,4-β glucanase from *S. solfataricus* . Aligned are the protein sequences of the three putative extracellular endoglucanases found in the *S. solfataricus* P2 genome. High consensus and low consensus residues are indicated respectively in red and blue. SSO1354 shows 85% identity with SSO1949 and 26% identity with SSO2534 (*celS*).

Interestingly, these sequences show also some similarities in the flanking genomic regions; in particular, both sequences have an ORF immediately downstream (SSO1353 and SSO1948) encoding a putative glycosyl hydrolase. The corresponding proteic sequences of SSO1353 and SSO1948 show an high identity

percent value, similar to that determined for SSO1354 and SSO1949 (~86%). Moreover, significant aligments were produced also with other ORFs flanking SSO1354 and SSO1949, namely sequences coding for the putative transposases SSO1946, SSO1951 and SSO1367. Therefore, one of two sequences was presumably generated from a duplication/insertion event, which could have been mediated by the transposable elements (ISs) mentioned (Fig.15).

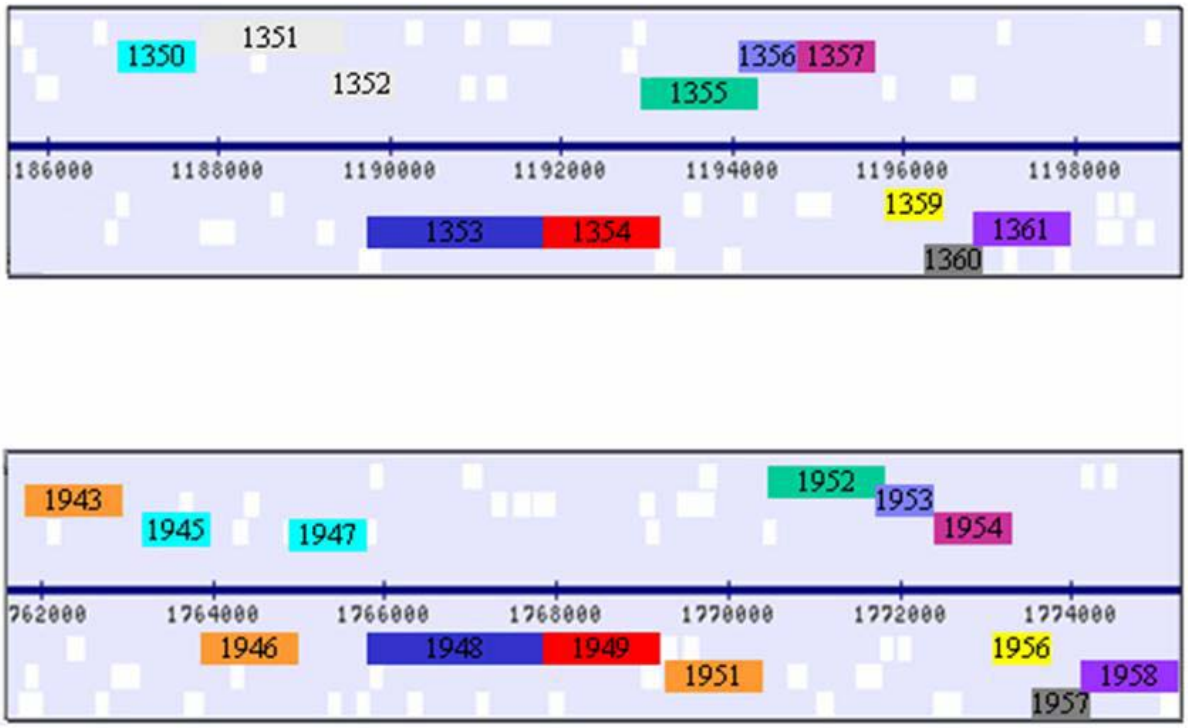

Fig 15 Comparison of the flanking regions of *sso1354* and *sso1949* genes on the *S. solfataricus* P2 genome. ORFs for which the corresponding protein sequences have an high percent of identity are marked with the same colour. *sso1354* and *sso1949* sequences are highlighted in red, the sequences coding for transposases in orange *(sso1943*, *sso1946 and sso 1951*). In light grey are two sequences (*sso1351* and *sso1352*) for which there are no corresponding sequences on *sso1949* region.

The inspection of the databanks revealed also only 26% identity between SSO1354 and the other putative cellulase, *celS*; furthermore, a good part of sequences for which the alignment scores was the most significant (escluding SSO1949) are from thermophilic Bacteria of the genus *Thermotoga* (particularly *maritima* e *neapolitana*). Moreover, the presence of the *glyco_hydro_12* conserved domain at the C-terminus was pointed out by the analysis of the SSO1354 sequence with the PFAM database program (available at website www.sanger.ac.uk); the protein can be assigned to family 12 of glycosyl hydrolase. Enzymes with four different catalytic activities belong to GH family 12 (clan C): endoglucanase (EC 3.2.1.4); xyloglucan hydrolase (EC 3.2.1.151); β-1,3-1,4-glucanase (EC 3.2.1.73); xyloglucan endotransglycosylase (EC 2.4.1.207); these enzymes have a retaining mechanism, in which two glutamic acid residues act as nucleophile and proton donor residues.
The structure of highly thermostable family 12 endoglucanase (Cel12A) from the thermophilic bacterium *Rhodothermus marinus* has been solved (Crennell S.J. *et al.* 2002); this enzyme has been shown to have one domain made of two β-sheets and one α-helix with a beta-jelly roll fold, resembling the structure of the xilanases family 11. The catalytic domain of SSO1354 enzyme has a significant amino acid

conservation compared with Cel12A (27.6% identity and 49.8% similarity), as previously seen for SSO1949 (Huang, Y. *et al.* 2004) (Fig. 16). The sequence similarity allowed a homology modelling of SSO1354 using the cellulase from *R. marinus* as template (fig. 17); only the modelling of the catalytic domain (aa 76-327) could be performed, since the N-terminal region does not show homology to Cel12A or to other known cellulose sequences (except SSO1949). The structural modelling combined with the sequence and structure based alignment clearly indicated the presence of a typical GH clan C fold in the SSO1354 protein (Bourne Y. and Henrissat B. 2001), with an active site which closely resembles that of other cellulases of family 12. By structural modelling, two putative catalytic glutamate residues, Glu-211 and Glu-310, were also identified as placed in the active site and highly conserved in other GH sequences. On the basis of known structures and well-characterized mechanisms of cellulases of the GH family 12, Glu-211 should act as a nucleophile and very likely Glu-310 displays the acid-base function.

```
            ---hydrophobic region------serine and threonine rich region-
SSO1354     1 MNKLYIVLPVIVIIAIGVMGGIIYLHQQSLSVKPVTTTEFSTTTSTSTTTNAITTTVTQT 60

            ------------------------
SSO1354    61 VTSITSYNQLIYVTSSASSPTPV 83

SSO1354    84 YLNNSTIPS--FYLEVNMWNAKNYNGNYTMVFNPLARTLSVSFNLTQVKPLEWTNGYPEI 141
  R.mar     8 RWDARDVAGGRYRVINNVWGAETAQCIEVGLETGNFTITRADHDNGNNVA--AYPAIYFG 65
  S.liv     7 PFGTTTIQG-RYVVQNNRWGSTAPQCV-TATDTG-FRVTQADGSAPTNGAPKSYPSVFNG 63
  T.res     6 QWATFTGNG--YTVSNNLWGASA------GSGFGCVTAVSLSGGASWHADWQWSGGQNNV 57

SSO1354  142 YVGRKPWDTAYAGNIFPMRIGNMTPFMVSFYINLTKLDPSINFDIASDAWIVRPQIAFSP 201
  R.mar   66 ----CHWGACTSNSGLPRRVQELSDVRTS--WTLTPITTG-RWNAAYDIWFS----PVTN 114
  S.liv   64 ----CHYTNCSPGTDLPVRLDTVSAAPSS--ISYGFVDGA-VYNASYDIWLD----PTAR 112
  T.res   58 ----KSYQNSQIAIPQKRTVNSISSMPTT--ASWSYSGSNIRANVAYDLFTA----ANPN 107

SSO1354  202 GTAPGNGDIEIMVWLFSQN-LQPAGEQVGKVVVPIYINHTLVNATFQVWEMKSVPWGGWE 260
  R.mar  115 SGNGYSGGAELMIWLNWNGGVMPGGSRVATV----------ELAGATWEVWYADWDWNY 163
  S.liv  113 TDG--VNQTEIMIWFNRVGPIQPIGSPVGTA----------SVGGRTWEVWSGGNGSND 159
  T.res  108 HVT-YSGDYELMIWLGKYGDIGPIGSSQGTV----------NVGGQSWTLYYGYNGAMQ 155

SSO1354  261 YIAFRPDGWKVTNGYVSYEPNLFIKALSNFTSYNITNYYLTDWEFGTE-WGTMTSNGTAY 319
  R.mar  164 IAYRRT---TPTTSVSELDLKAFIDDAVA-RGYIRPEWYLHAVETGFELWEGGAGLRSAD 219
  S.liv  160 VLSFVA---PSAISGWSFDVMDFVRATVA-RGLAENDWYLTSVQAGFEPWQNGAGLAVNS 215
  T.res  156 VYSFVA---QTNTTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVAS 212

SSO1354  320 FSWTVSNFSETLL  332
  R.mar  220 FSVTVQKL      227
  S.liv  216 FSSTVETG      223
  T.res  213 WTASIN        218
```

Fig. 16 Alignment of SSO1354 with cellulase sequences. Aligned are the catalytic modules of SSO1354 and the cellulases from *Rhodothermus marinus* (***R.mar***), *Streptomices lividans* (***S.liv***) and *Thricoderma reesei* (***T.res***). The complete sequence is indicated only for SSO1354. The N-terminal region (aa 1-83) has no sequence similarity to the other cellulases. High consensus and low consensus residues are indicated respectively in red and blue. The catalytic glutamate residues Glu-211 and Glu-308 are highly conserved and are highlighted in yellow.

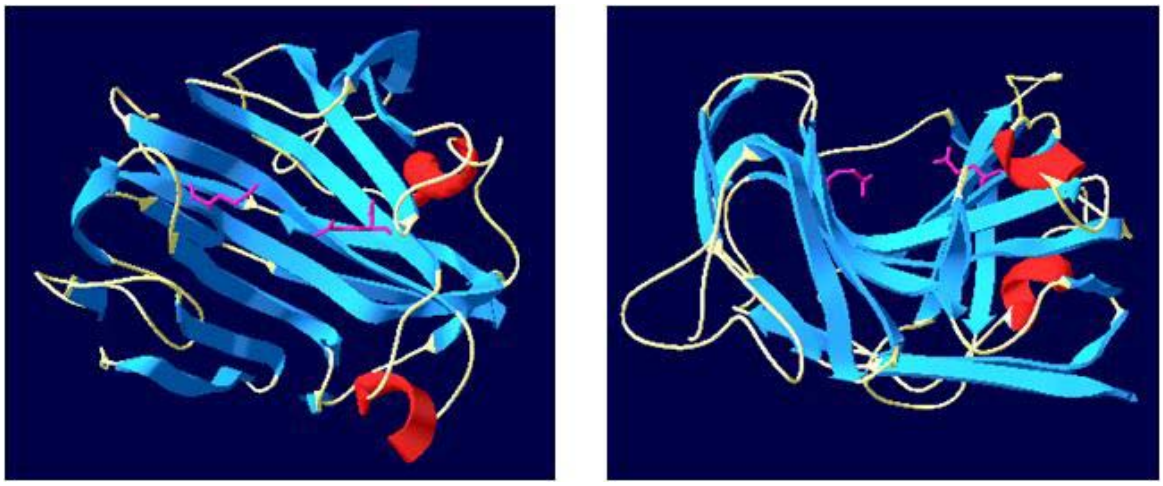

Fig. 17 Structural model of SSO1354 cellulase. The structure of catalityc domain of SSO1354 (aa 76-327) was modeled with the EsyPred3D program using the cellulase from *R. marinus* (PDB no. 1H0B) as template. The β-strands are coloured in light blue, α helices in red. The catalytic glutamate residues are depicted in purple. The active-site cleft is best viewed from the side (right picture)

Due to the striking difference with other cellulases in the corresponding region, the N-terminal sequence of SSO1354 was analyzed in depth, in order to complete a confident structural-functional prediction to the whole protein sequence. We found that SSO1354 can be successfully aligned to the sugar binding proteins (SBP) from *S. solfataricus* AraS, GlcS and TreS; these proteins, that respectively bind arabinose, glucose and trehalose, are the subunits of ATP binding cassette (ABC) transporters responsible for sugar uptake in *S. solfataricus* (Elferink M.G. *et al.* 2001) (Fig,18).

```
          1                                                          60
SSO1354   MNK-----------LYIVLPVIVIIAIGVMGGIIYLHQQSLSVKPVTTTEFST-------
   treS   MRR--------GLSTTTIIGIVVAIVIIVIGAVAAVTLLSHKPSQVVST-----------
   araS   MSRRRLY---KAISRTAIIIIVVVIIIAAIAGGLAAYYSSSKPPATSTSLTSTSSSLSVT
   glcS   MKRKYPYSLAKGLTSTQIAVIVAVIVIVIIIGVVAGFVLTKGPSTTAVTTTVTS------

          61                                                        120
SSO1354   TTSTSTTTNAITTTVTQTVTSITSYNQLIYVTSSASSPTPVYLNNSTIPSFYLEVNM---
   treS   --TSPSTSQSATSTSPSQVITITYFDDLSPSEANITQKIIIPQFEATHPNIKINYVD---
   araS   TSSTTSTLSSITTTTSTASSYVVDFINPWGAEDPVGLKWIGGNFSIYYPGYSVQFTSLPG
   glcS   TFTTTTTIPS--TTTSTPSNTVVFYTWWGGGDGGEALSQIIPAVK-QYAGLQMQTYSIPG
```

Fig.18 Aligment of N-terminal amino acid sequences of SSO1354 and SBPs. High consensus and low consensus residues are indicated respectively in red and blue; putative transmembrane segments are highlighted in yellow, and the long hydrophobic stretch rich in hydroxylated amino acids is indicated in light blue.

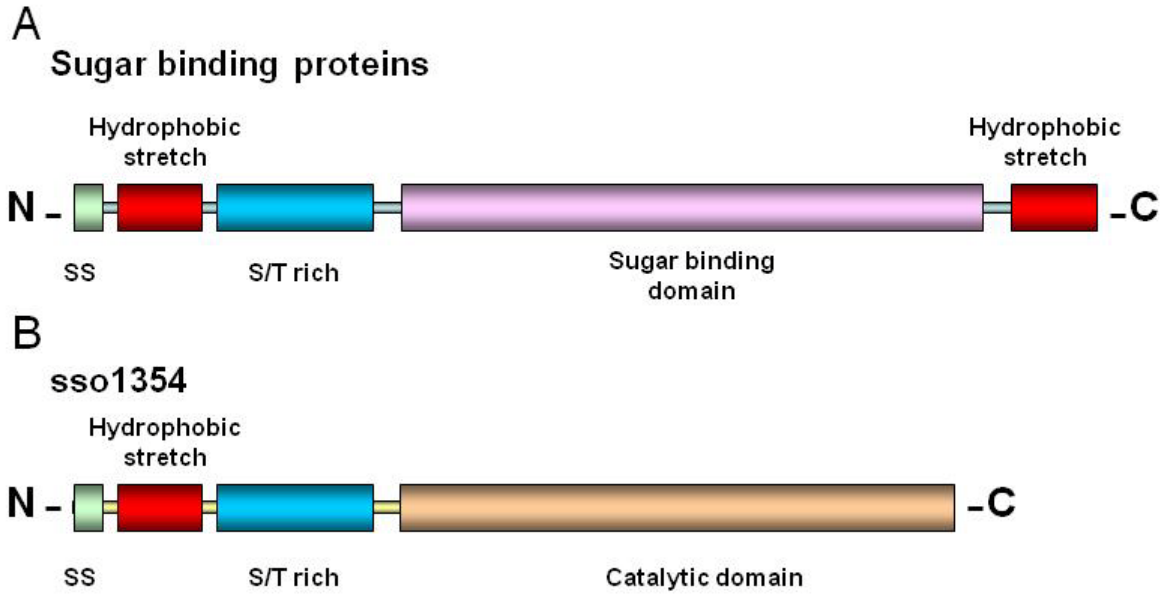

Fig.19 Domain organization of sugar binding proteins (SBP) A, and SSO1354 enzyme B. **SS**, signal peptide; **S/T rich**, linker region with high percentage of hydroxylated amino acids (mainly serine and threonine). SBPs and SSO1354 cellulase share a similar organization, which probably reflects a similar function of N-terminal domains.

The sequence alignment pointed out that SSO1354 shares an overall domain architecture with the sugar binding proteins (fig. 19): the hydrophobic region at the N-terminus of all proteins compared is linked to the catalytic (for the cellulase) or the sugar binding (for transporters) modules by a S-T rich stretch. This lipophilic domain serves as a transmembrane segment, namely an anchor to the membrane. Unlike SSO1354, the sugar binding proteins have an additional hydrophobic stretch at the C-terminus, with an identical function. The SSO1354 sequence was also analyzed for putative N-glycosilation sites, and 11 residues were found that match the consensus Asn-X-Ser/Thr in eukaryiotic extracellular proteins.

Cloning of the expression cassette *araS*1354 in the pMSSV vector and expression in *S. solfataricus*

The SSO1354 coding sequence plus a downstream sequence coding for a (His)x6 tag was fused with the promoter region of *araS*; this strong arabinose-inducible promoter have a low basal activity and was chosen to allow the induction of high-level expression without imposing stress on the host cells; moreover, *araS* promoter has been demonstrated to efficiently drive the expression of several recombinant proteins in *S. solfataricus* (Albers S.V. *et al.* 2006) .
The *araS*-sso1354 cassette was inserted into the polycloning site of the pMSSV vector by a directional cloning. The resultant pMSSV-*araS*1354 (Fig. 20A) vector showed to propagate without rearrangements in *E. coli*; therefore it was transferred into *S. solfataricus* GθW/SSV2 cells by electroporation. GθW mutant cells was used for the shorter doubling time than wild-type. After several propagation steps, the succes of transformation was confirmed by an electrophoresis of extrachromosomal DNA preparation (fig. 20B).

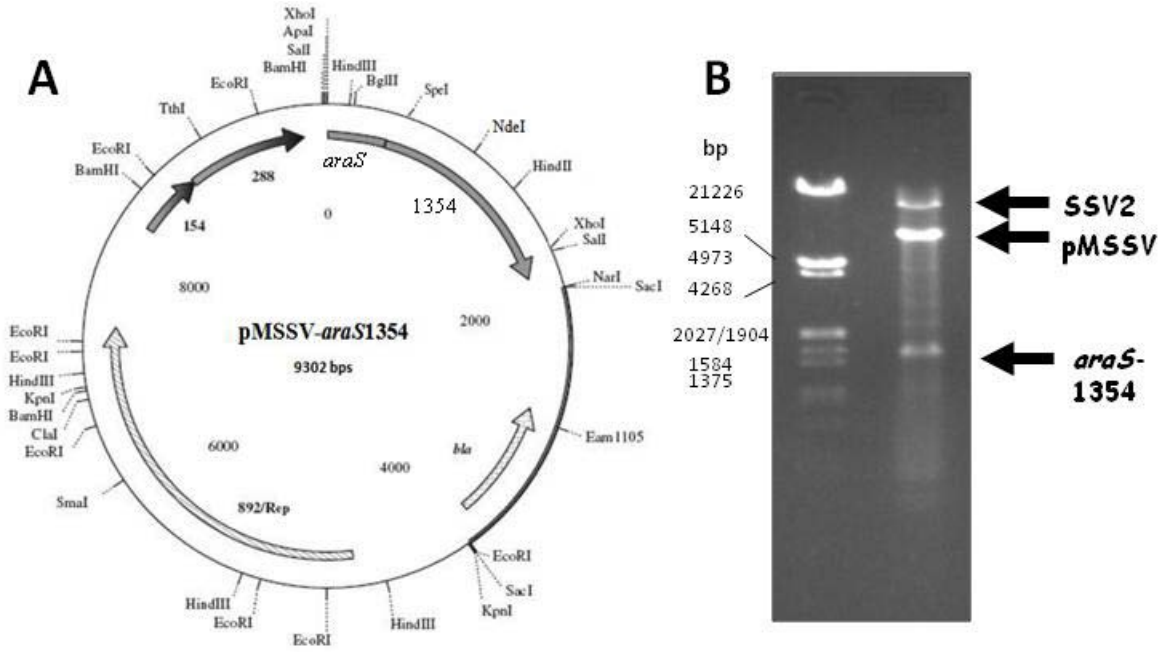

Fig.20 pMSSV-*araS*1354 map and vector propagation in *S. solfataricus* cells. The map of the pMSSV expression vector containing the *araS* promoter fused with the SSO1354 coding sequence is shown in panel **A**. Electrophoresis on 0.8% agarose gel of a Sal I digestion of an extrachromosomal DNA preparation from *S. solfataricus* GθW/SSV2 cells transformed with pMSSV-*araS*1354 vector is shown in panel **B**.

The cellulase activity was subsequently visualized by a zimographic assay, indicating that pMSSV*araS*1354 vector was succesfully capable to drive the expression of cellulase gene (Fig.21).

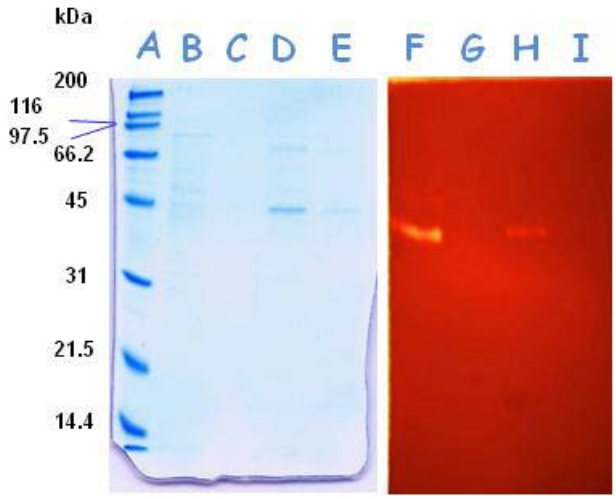

Fig.21 Expression of *sso1354* sequence in *S. solfataricus*. Proteins of the supernatant from transformed *S. solfataricus* cells were precipitated with ammonium sulphate (90% saturation); the pellet was resuspended and loaded onto an anionic exchange column. The fractions showing cellulase activity were pooled and loaded onto nickel affinity spin columns. Samples from the nickel affinity chromatography were analysed by SDS/PAGE followed by Comassie Blue and enzyme staining: marker(A), crude extract (B, F), wash(C, G), eluition with 250mM imidazole(D, H), eluition with 500mM imidazole(E, I)., After renaturation, cellulase activity was revealed in the zymographic gel (containing 0.1% CMC) by staining with Congo Red.

Unexpectedly, the most part of cellulasic activity was found in the supernatants of steady-state cultures, and not in the membrane fractions. We have hypotized that this free released activity can be ascribed to an amount of cell bound SSO1354 enzyme upon cell autolysis and membrane breakdown. In fact, isopycnic gradients of cesium chloride of concentrated conditioned media concentrated the SSO1354 activity in just one middle band. Moreover it was found only associated in complexes with proteins aggregated by membrane lipids.
SSO1354 was partially purified by ammonium sulfate precipitation and nickel affinity chromatography, and it was demonstrated to be active at low pH (2.0) and high temperature (80°C). The activity spot on zymography was corrispondent to a 45kD band on SDS-PAGE; the apparent molecular weight was higher than expected (37.2 kDa), indicating that the protein is glycosylated.
AraS promoter activity had been demonstrated to be inhibited from several compounds present in the standard *S. solfataricus* media (Lubelska J.M. *et al.* 2006). Therefore, different culture medium compositions were tested in order to optimize the expression of SSO1354 sequence; unfortunately, reproducibility of the results and high-level expression were unsuccessful, mainly because pMSSV-*araS*1354 vector showed to be unstable and to undergo rearrangements when the cells were grown in basal media with arabinose as the main carbon source.
Experiments aimed to increase the stabilty of the construct are now in progress. The genes *pyrEF* were chosen as a selectable marker for the complementation of uracil auxotrophic mutants and inserted into the vector pMSSV to impose a selective pressure for the maintenance of the plasmid; moreover, a new expression cassette, in which the *sso1354* gene was under the transcriptional control of the *glcS* promoter, was cloned into the vector (Fig. 22A). The *glcS* promoter is unsensitive to the inhibitors of the *araS* promoter activity, therefore it is more suitable to drive the expression of *sso1354* gene in culture media of virtually any composition. Furthermore, two mutants of the strain *S. islandicus* were used as hosts: one had been cured for both pSSVx and SSV2, while the other only for pSSVx. The construct was transferred into the cells by electroporation in the Quinxin She's laboratory at the Danish Archaea Center of the Univeristy of Copenhagen. This modified host/vector genetic system guarantees the stability and maintenance of the DNA construct through several generations as demonstrated by restriction maps of 5 different transformant clones (Fig. 22B).

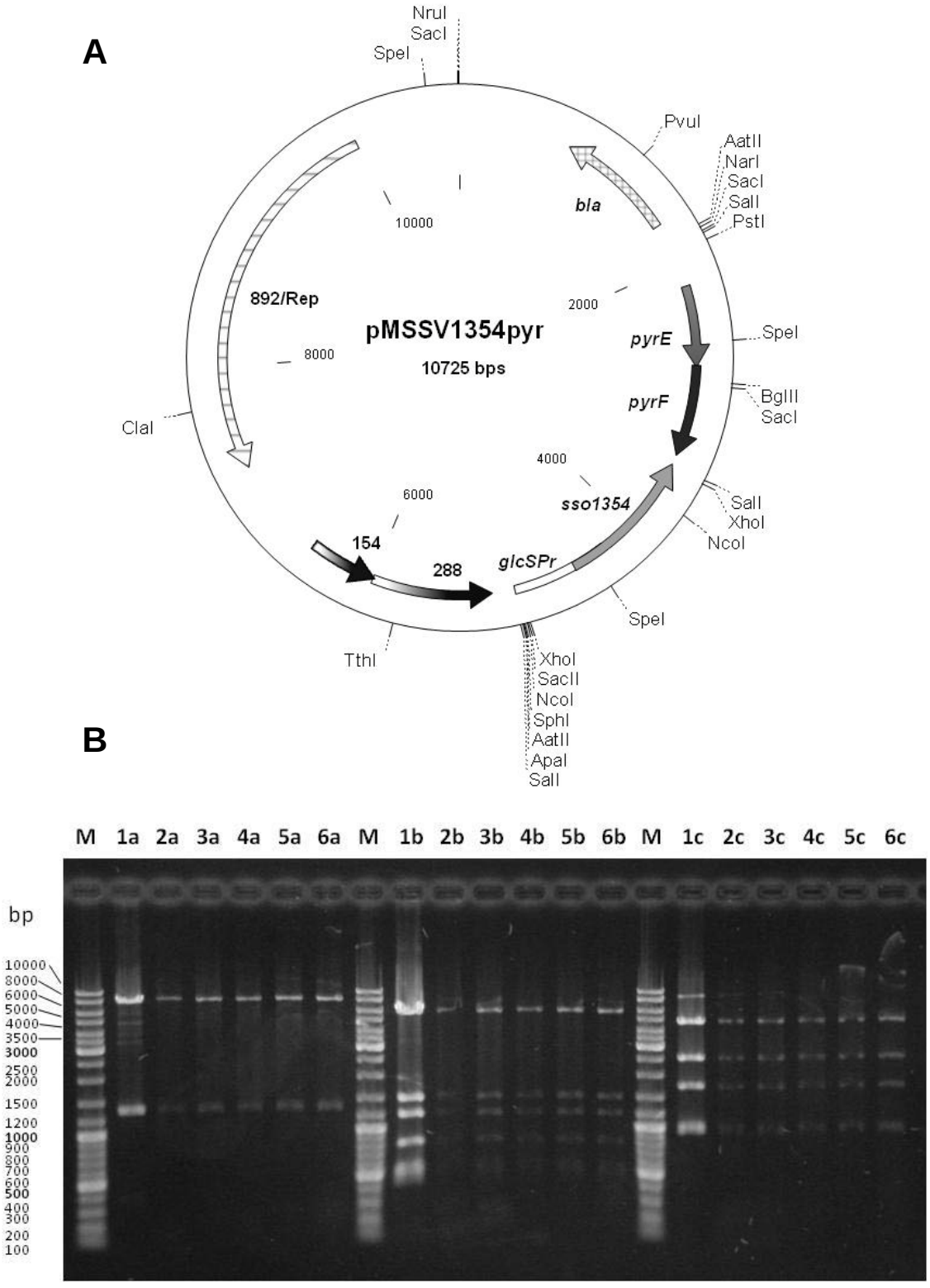

Fig. 22 Map of pMSSV-1354*pyr* and propagation of the construct. A: The expression cassette *glc*S1354 fused with the *pyrEF* genes was cloned into the vector pMSSV, producing the expression vector pMSSV1354*pyr* **B:** *S. islandicus pyr* - cells, cured for both pSSVx and SSV2, were electroporated with the pMSSV1354*pyr* vector. Five *S. islandicus* transformed clones were isolated. Extrachromosomal DNAs extracted from *E. coli mos blue* cells (lanes 1 a,b,c) and *S. islandicus* isolated clones (lanes 2-6 a,b,c), were checked by agarose electrophoresis for the presence of pMSSV1354*pyr* (lanes 1-6a: BamHI cut, lanes 1-6b: EcoRI cut, lanes 1-6c: HindIII cut). M: molecular weight markers (bp).

Expression of the *sso1354* gene in *E. coli*

The expression of *sso1354* sequence was undertaken in the conventional mesophilic host *E. coli* with the aim to obtain higher amounts of the enzyme and reproducibility than in the *S. solfataricus* expression system. Initially, the whole coding sequence of SSO1354 fused with a sequence coding for an (His)x6 tag at C-terminus was cloned into pET28c expression vector; this vector was used to transfom both *E. coli* BL21(DE3) and *arctic-express* (DE3) cells. Arctic strain was employed since allows the expression at low growth temperatures, minimizing the risks of protein loss due to aggregation, formation of inclusion bodies or degradation by heat shock proteases that are induced under overexpression conditions. The cellulase activity and the presence of a specific SSO1354 polypeptide were undetectable in the transformed cells of both strains. This complete lack of translation could be due to the presence of the putative membrane anchoring N-terminal peptide and/or to the addition of the C-terminal His tag, namely to the impairing of a correct folding of the SSO1354 polypeptide, which could favour its rapid degradation.
The SSO1354 DNA sequence was PCR re-amplified from the codon 28 and cloned into pET28c, so that the enzyme could be translated without the hyprophobic N-terminal peptide. Moreover, the cloning allowed the in-frame insertion of the sequence immediately downstream of a DNA stretch encoding a (His)x6 tag to obtain a recombinat enzyme with the His tag at the N-terminus. This strategy resulted in the successful of the recombinant enzyme production both from the *arctic-express* and BL21 *E. coli* cells, with comparable yield. The expression of *sso1354* gene was visualized on SDS-PAGE as 35,5 kDa band, which is in good agreement with the theoretical mass of 35154 Da calculated. The activity of the protein was qualitatively determined by a zimographic assays which revealed specific cellulase activity bands, absent in cell extract of untransformed *E. coli* (Fig. 23).

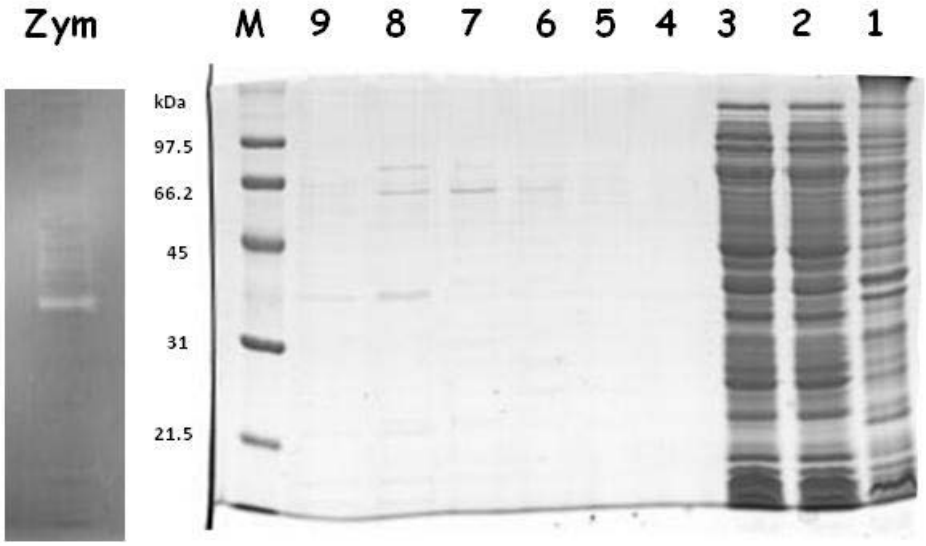

Fig.23 Expression of *sso1354* gene in *E. coli*. BL21 *E. coli* cells transformed with pET28c-1354 were grown up to 0.8 OD_{600} and gene expression induced by addition of 0.4 mM IPTG. After 16-h induction the cells were centrifugated, sonicated and the cell extract was loaded onto a Nickel affinity chromatography spin column. The samples were analysed by SDS-PAGE followed by Comassie staining: marker (M), membrane pellet (lane 1), crude cytosolic extract (2), flow through (3), wash I° (4), wash II° (5), eluition with 10mM imidazole (6), eluition with 50mM imidazole (7), eluition with 100mM imidazole (8), eluition with 250mM imidazole (9). A sample from the 100mM imidazole eluition was analyzed on a zymographic gel (containing 0.1% CMC) (lane Zym). After renaturation cellulase activity was revealed by staining with Congo Red.

As expected, SSO1354 producted in *E. coli* was optimally active at high temperature and low pH (80°C and pH 2), but unfortunatly it was also unstable and showed a very low specific activity when compared to the wild type enzyme.

Expression and characterization of the *sso1354* gene in *Kluyveromyces lactis*

As previously mentioned, SSO1354 sequence contain several glycosylation sites, and the native protein is likely glycosylated in *S. solfataricus*; the expression of *sso1354* gene in an eucaryotic host, like *K. lactis*, was attempted, with the aim to compare the recombinant enzyme expressed in yeast with the same protein prevoiusly expressed in the other hosts. Also for yeast expression, *sso1354* gene was cloned into the expression vector (pKLAC1) starting from codon 28; the sequence was cloned downstream of the coding sequence of the *K .lactis* α-mating factor domain (α-MF), hence resulting in the translation of chimeric protein. The α-MF domain is an efficient signal for the α-MF fusion proteins to be efficiently transported through the yeast secretory pathway. An α-MF fusion protein undergoes sequential processing by signal peptidase in the endoplasmic reticulum (ER) and the Kex protease in the Golgi, resulting in the efficient release of the protein in native form into the growth medium (Fig 24).

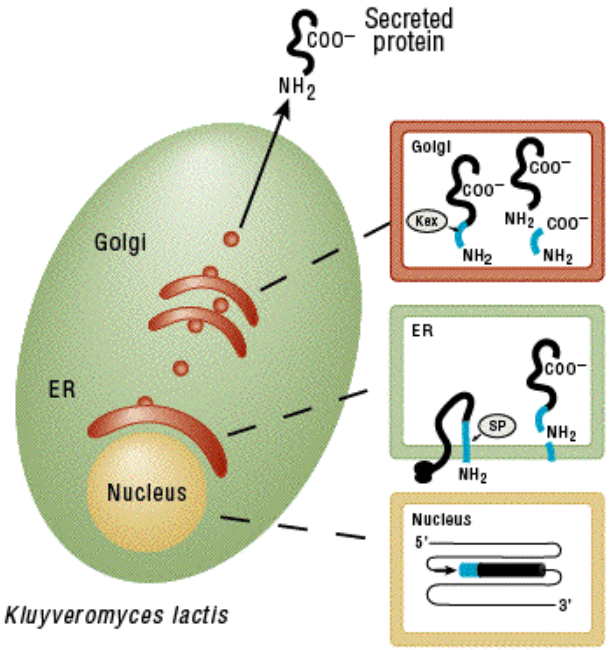

Fig.24 pKLAC1 based *K. lactis* expression sistem. The picture shows the secreted protein processing. In the nucleus, DNA encoding a protein fusion of the a-MF domain (blue) and a desired recombinant protein (black) is integrated into the yeast genome. Expression is driven by the $P_{LAC4\text{-}PBI}$ promoter (arrow). Upon expression of the fusion protein, a signal peptide in the a-MF domain directs translocation of the fusion protein into the lumen of the endoplasmic reticulum (ER) and is removed by signal peptidase (SP). Secretory vesicles (circles) transport the fusion protein to the Golgi where the Kex endoprotease cleaves the a-MF pro-domain, releasing a mature form of the desired protein. The protein of interest is then transported via vesicles to the plasma membrane (PM) where it is secreted from the cell.

Cells of the GG799 strain were used as a hosts for transformation with pKLAC1-*sso1354*; this strain has no auxotrophies, rapidly grows to high cell density, and efficiently secretes heterologous proteins.
With the yeast expression system, *sso1354* sequence was expressed and secreted, and the enzyme was catalytically active. Unexpectedly, the protein showed a marked retardation on SDS-PAGE, stopping at the separation line between stacking and separating compartments of a discontinuous gel; the most part of cellulase activity was found on the stacking gel of the zimograms (fig. 25). This electrophoretic slow mobility could be due to the hyperglycosylation of the polypeptide by yeast expression and/or to resistance of the protein to the denaturation before loading on the gel. The recombinant SSO1354 enzyme from concentrated supernatants of *K. lactis* cultures was further characterized. The protein was partially purified by a precipitation with ammonium sulphate; surprisingly, SSO1354 enzyme did not precipitate in ammonium sulphate (90% saturation) like most of other proteins, and the whole cellulasic activity was found in ammonium sulphate supernatants.

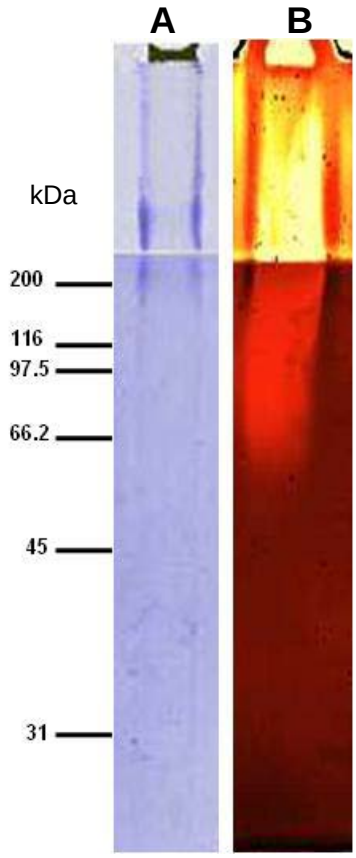

Fig. 25 Expression of *sso1354* in *K. lactis*. Proteins from a colture of *K. lactis* cells, transformed with the vector pKLAC1-*sso1354*, were concentrated and precipitated with ammonium sulphate (90% saturation). The figure shows the SDS-PAGE analysis followed by Comassie Blue staining (lane A) and the zymographic assay (lane B) of a sample of supernatants after precipitation.

The enzyme was active toward CMC-cellulose, as well as several other polymers: lichenan (a linear glucan of {1→3, 1→4} -β-glycosidic bonds in a ratio of 1:2) with a 120% activity with respect to CMC-cellulose; debranched arabinan (a polymeric chain of 1,5-a-linked L-arabinofuranosyl residues) with a 25% activity; curdlan and pachyman (both polymers of 1→3-β-linked D-glucosyl residues), with respectively a 26% and 22% activity, oat spelts xylan (a1→4-β-linked polymer contain predominantly D-xylose) with a 19% activity. Furthermore, SSO1354 cellulase did not hydrolyse *p*-nitrophenyl-β-D-cellobioside and *p*-nitrophenyl-β-D-cellotrioside, similarly to the behaviour described for the other *S. solfataricus* cellulase SSO1949, namely it

has no exo-cellulase activity, confirming that SSO1354 is mainly an endo-1,4-β glucanase.
pH and temperature dependance of recombinant SSO1354 activity, investigated in the range of 1.8-7.5 and 37-90°C, respectively, are shown in fig. 26. The temperature optimum was 80°C; at this temperature the highest activity was found at pH 5.0. Moreover, the recombinant cellulase showed a strong thermostability; when incubed at 90°C, the enzyme had an half life of 180 min.
SSO1354 was also tested to verify whether the enzyme retains the cellulasic activity in ionic liquids; first, the highest water concentration tolerated for solubility of α-cellulose in solutions of ionic liquids was determined as 20%. Qualitative activity detection by thin layer chromatography analysis of the idrolyzed products, revealed that the enzyme retained activity in 1-ethyl-3-methyl-imidazolium-acetate, producing cellobiose and cellotriose as main products of hydrolysis in a time dependent accumulation over 2.0-16.0 hour incubations (Fig. 27).

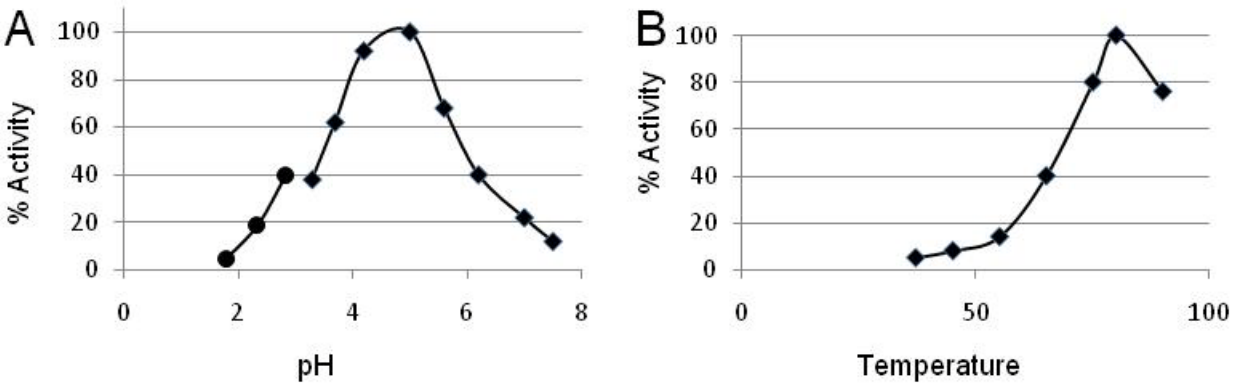

Fig.26 pH and temperature optimum of SSO1354 enzyme. **A**: The pH optimum was assayed at 80°C. The reactions were performed in potassium phosphate buffer (25mM) in the range of 1.8-2.8 (●); citrate phosphate buffer (25mM) was used in the range of 3.3-7.5 (♦). **B:** The cellulolytic activity of SSO1354 enzyme was assayed at various temperatures in 25mM citrate phosphate buffer (pH 5)

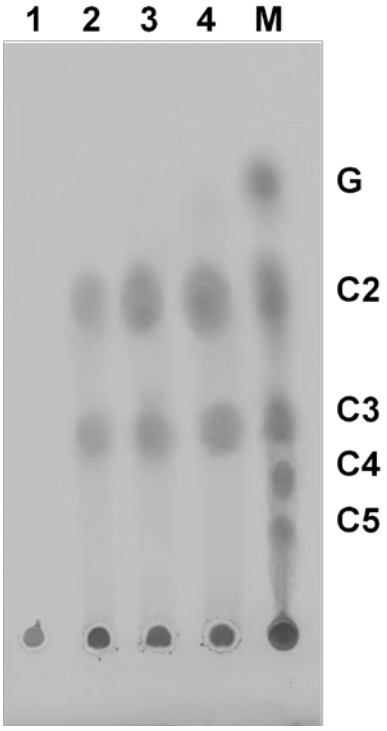

Fig.27 Degradation of cellulose in ionic liquids. A 1% α-cellulose solution in 1-ethyl-3-methyl-imidazolium-acetate was incubated with SSO1354 enzyme. Samples were withdrawn at several time points, and the ionic liquid was removed by three phenol/chloroforme and one chloroforme extractions. Reactions with (lanes 2-4) and without (lane 1) enzyme were spotted on silica plates and developed with a mixture of ethyl acetate, H_2O, acetic acid, isopropanol, formic acid, (25:15:10:5:1, by vol.) Sugars were visualized with α-naphtol after incubation of 5min at 120°C. Lane **M:** mix of cello-oligomers (3µl of a solution with a concentration of 5mM of each oligomer); **1:** reaction mix incubated for 16h without enzyme; **2:** 2h incubation; **3:** 4h incubation; **4:** 16h incubation. (60 µl of samples are loaded on lanes 1-4)

Discussion

The present PhD work was developed following two strongly interconnected experimental lines: in the one hand, a deep characterization of a shuttle vector *E.coli – S. solfataricus* has been performed, in order to elucidate the potentialities of this genetic element. In the other hand, the same vector has been used to express *sso1354* gene, which encodes for a thermostable enzyme with cellulasic activity from *S. solfataricus*, in a xenologous fashion; *sso1354* has been also expressed in two heterologous hosts, with the aim to optimize the expression yield and to compare the recombinant protein obtained from different sources.

Stable and efficient transformation/expression genetic systems for hyperthermophilic Archaea represent key tools both for basic and biotechnological researches. These systems can be used to efficiently characterize promoters and coding sequences; moreover, archaeal expression systems can be also employed to (over)express proteins with commercial potential, like the enzymes from (hyper)thermophiles, in the same organisms from which the sequences are derived, avoiding the common problems in expressing a protein from a thermophilic organism into a mesophilic hosts (i.e. incorrect folding and/or incompatible post-translational modifications).

A wide variety of systems have been developed for protein production in members of the Bacteria and Eukarya, using numerous combinations of vector and promoter systems. The development of genetic systems is not so advanced for archaeal microorganisms, particularly for that belonging to chrenarchaeal kingdom (in fact the most of the few existing transformation tools for the production of recombinant proteins have been developed for halophiles and methanogenes Euryarchaea). The crenarchaeote *S. solfataricus* has developed into an important model organism for molecular and biochemical studies of hyperthermophilic Archaea, and it is highly desirable to develop genetic tools for in vivo studies and for high-level production of proteins in this organism; *S. solfataricus* could be employed as “cell factory” in plants for the production of biotechnological relevant proteins.

In the present work, the expression shuttle vector pMSSV for *S. solfataricus* was tested, using an expression cassette containing the marker gene *lacS*, which encode for a *S. solfataricus* β-glycosidase. This system presents many advantages compared with others already mentioned. After DNA transfer, the vector propagates efficiently throughout the culture as a virus; the high propagation efficiency of the pMSSV vector, in SSV2 lysogens of the strain GθW, was found to be comparable to the wild-type pSSVx. Therefore, the pMSSV vector is proper to overcome the usual low transformation efficiency of *Sulfolobus* cells (10^{-3}-10^{-4} cells/μg of plasmidic DNA used). Moreover, this genetic system have shown another important feature: the vector is stably maintained at high-copy number in episomal form inside host cells (up to 130 copies per chromosome, the number depending on the cell growth phase), with no integration into the chromosome and hence no reduction in the number of episomal molecules. This feature is particularly advantageuos in the expression of passenger genes; the yield of expression would not be affected by the decrease of the copy number as consequence of the integration of the vector into the host chromosome. In fact, the analysis have shown that pMSSV vector is succesfully capable to drive the expression of the marker gene *lacS*.

Interestingly, β-galactosidase activity in complemented mutant cells followed an increasing trend which correlated to the plasmid accumulation up to stationary phase and was not dependent on the number of the generations. *lacS* gene expression was

inducible by thermal stress, namely it was confirmed to be transcriptionally regulated by the chaperonin gene promoter. However, heat shock did not produce the same β-galactosidase activity increase (~10-fold) described by Jonuscheit *et al.* (2003), although maximum gene expression resulted at comparable levels in the two host/vector systems (3 versus 5 U/mg). Therefore, the characterization of pMSSV have pointed out some interesting features of this system, which make it a powerful genetic tool, with a strong application potential in basic researches as well as in biotechnology.

Therefore, the vector/host system developed was used for the expression of the cellulase gene from *S. solfataricus*. As previously mentioned, the cellulases, particularly from (hyper)thermophilic organisms, have a wide range of industrial application, and consequently they are biotechnological relevant enzymes. The gene which encodes for the putative extracellular endoglucanase SSO1354 was expressed in xenologous and heterologous fashion; SSO1354 is the only one of three *S. solfataricus* putative cellulases for which a characterization study have been not performed yet.

Some interesting remarks can be made on the base of the sequence analysis of SSO1354; particularly interesting is the comparison with the sugar binding proteins (SBP) of *S. solfataricus* AraS, GlcS e TreS, that bind arabinose, glucose and trehalose, respectively. The sequence similarity is significant, particuarly in a "module" in the N-terminus, which instead does not produce any alignment with N-terminal regions in other known cellulase sequences (except SSO1949), namely the SSO1354 protein and the SBPs share a similar domain architecture. Therefore, SSO1354 N-terminal domains was assumed to have the same function as in SBPs. AraS, GlcS and TreS, such as various *S. solfataricus* solute binding proteins, have an unusual leader peptide that resembles bacterial type IV prepilin signal sequences and is cleaved by the *S. solfataricus* homolog of bacterial type IV prepilin peptidases, named *PibD* (Albers S.V. *et al.* 2003); SSO1354 putative signal sequence matchs only in part the relatively conserved consensus sequence which have been found to be recognized from *PibD*. However, this enzyme has been demonstrated to be equipped with a relatively broad substrate specificity, and to tolerate some mutations in the core of four amino acids around the cleavage site.

SSO1354 (and probably also the other two cellulases from *S. solfataricus* that also show a similar organization at N-terminus) could be a cell-bound cellulase, with the catalytic domain linked by the S-T rich stretch to the transmebrane region; this architercture could give the needed flexibility to the enzyme. In fact the S-T rich stretch has been previously demonstrated to be fundamental for catalytic activity in the closely related SSO1949 enzyme; the enzyme expressed without this domain has been shown to have no endoglucanasic activity. Our hypotesys of cell bound cellulases is confirmed by the data avaiable in literature: in fact the most part of cellulasic activity in *S. solfataricus* cultures has been found in membrane extracts (Cannio R. *et al.* 2004); however our hypotesys is in contrast with that previously assumed for SSO1949 (Huang Y. *et al.* 2005); Huang and collegues have been assigned to the N-terminal hydrophobic region the role of signal peptide by a bioinformatic prediction with SignalP, and consequently have assumed that SSO1949 is a medium-released cellulase. This hypothesis was indeed not supported by any data about characterization and localization *in vivo* of cellulase activity. The co-localization of the endoglucanases and the sugar binding proteins could be an advantageous strategy for the optimization of nutrient assimilation, particularly for

cells living in habitats where complex carbohydrate are a major, but scarce, energy source.
The SSO1354 cellulase was successfully producted both in homologous and heterologous hosts. The first strategy was to express the protein in *S. solfataricus*, considering the documented difficulty to express these cellulase genes from (hyper)thermophilic organisms in phylogenetical distant conventional hosts, like *E.coli*. In fact, the other two *S. solfataricus* cellulase genes *sso1949* (Huang Y. *et al.* 2005) and *sso2534* (data not shown) have been shown to be considerably difficult to handle in conventional mesophilic hosts. The cellulase gene was expressed under the control of a strong and inducible promoter, *araS*, using the vector pMSSV. SSO1354 was obtained in catalytical active form and, like the genome encoded natural product, it was glycosylated and active at extreme conditions. The membrane bound extracellular localization, predicted by the sequence analysis and suggested by literature data, was confirmed; SSO1354 activity was found mostly in culture surnatants, but the protein was demonstrated to be associated with lipids. Therefore, the cellulase was likely released in the culture medium as a consequence of cell lysis since the highest overall activity was obtained in cultures grown up to late stationary phase. The main problem encountered with the expression in *S. solfataricus* was the stability of the construct pMSSV-*araS*1354 when the cells were grown in specific media for the induction of the *araS* promoter activity. This bottleneck was overcome by imposing a selective pressure (a genetic selection marker was added to the vector) and using a different host strain; these experiments, which are now in progress, are giving very promising results. In fact, this strategy resulted in the improving of the genetic system based on the pMSSV vector, by providing it with increased genetic stability. Moreover, the vector with the selection genes was demonstrated to be able to replicate indipendently as a plasmid without the virus helper SSV2 under the selective pressure. This feature showed by the modified pMSSV is a fundamental requirement to next isolate single transformant clones and to use this vector for mutagenesis and isolation of single mutated gene variants.
SSO1354 was also producted in both a bacterial and eukaryotic mesophilic hosts, with the aim to diversify the strategies of expression and to improve the yield of recombinant protein. In *E. coli*, the exclusion of the N-terminal hydrophobic sequence was necessary to reach a succesful expression of the cellulase, suggesting that this sequence could interfere with translation or provoke a misfolding of the protein. SSO1354 producted in *E. coli* conserved the preference for low pH and high temperature of the wild type enzyme obtained from *S. solfataricus* but unfortunately showed also a marked instability and low specific activity.
Therefore, due to the more strict similarity in the secretion/post translational modification apparatus, *K. lactis* has been chosen as a more suitable host to express *sso1354* gene; *K. lactis* have a large potential for the biotechnological production of recombinant proteins, as demonstrated by a large number of sequence which have been successfully expressed in this host (van Ooyen A.J. *et al* 2006). Moreover, *K.lactis* has a number of advantages over other yeast expression systems; these include easy genetic manipulation, GRAS certification (generally regarded as safe) and an efficient and versatile secretory apparatus. The cellulase expressed in yeast was found to be hyperglycosylated, according with the data present in literature about expression patterns of heterologous proteins in yeast. The protein was characterized, and the comparison of the main features with those found for SSO1354 producted in the other hosts pointed out some interesting differences: the temperature optimum (80°C) was in agreement with the precedent findings, whilst the

pH optimum (5.0) diverged from that expected, even if the enzyme retained the activity also at lower pH; moreover, the enzyme showed a strong increase of thermostability (half-life of 3h at 90°C) when compared to SSO1354 producted in *E.coli* and SSO1949 (which has been characterized after the production in *E. coli*). The effect on pH optimum was probably due to the different hyperglycosilation of the protein in the yeast host when compared to the wild type enzyme produced by *S. solfataricus*. Hyperglycosylation was also responsible for the higher thermostability of the recombinant enzyme expressed in *Sulfolobus* and yeast cells with respect to the non glycosilated polypeptide produced in *E. coli*. In particular, the effect of glycosylation in increasing the resistance of proteins to extreme conditions is well documented (Koseki T. *et al.* 2006; Kim Y.O. *et al.* 2006). This could be a post-translational modification mechanism selected by evolution to provide an extra stability even to the proteins from hyperthermophiles.

The cellulase expressed in yeast was also demonstrated to retain the catalytic activity in ionic liquids. This result is particularly relevant; in fact these new class of solvents, as previously discussed (see “Introduction”), have a great potential in eco-sustainable production of bioethanol: it can be used as reaction media for suspension and subsequent increased susceptibily to hydrolytic attack of cellulosic materials. To our knowledge, this is the first example of a glycosyl hydrolase capable to work in ionic liquids and in reaction environments at very low water content.

The findings obtainend in this work shed light the great biotechnological relevance of the SSO1354 cellulase, which is an ideal candidate to be employed in several industrial processes, and particularly in the biofuel production. Furthermore, new elements on structure-function relationship was acquired: from sequence comparison and literature background about the characterization of *S. solfataricus* cellulasic activity *in vivo*, was assumed that SSO1354, as well as SSO2534 and SSO1949, is a cell-bound enzyme, and not free released in extracellular medium. This finding opens new and intriguing perspectives on the strategy adopted by a hyperthermophilic archaeon as *S. solfataricus* to coordinate and organize the activities involved in the nutrient metabolism.

Bibliography

Aguilar CF, Sanderson I, Moracci M, Ciaramella M, Nucci R, Rossi M, Pearl LH (1997) "**Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability"** .J Mol Biol. **271**, 789-802.

Albers SV, Szabó Z, Driessen AJ. (2003) "**Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity**". J Bacteriol. **185**, 3918-25.

Allers T, Mevarech M (2005) "**Archaeal genetics—the third way**". Nat Rev Genet **6**, 58–73

Arnold HP, Zillig W, Ziese U, Holz I, Crosby M, Utterback T, Weidmann JF, Kristjanson JK, Klenk HP, Nelson KE, Fraser CM (2000) "**A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus***". Virology **267**, 252–266

Arnold HP, She Q, Phan H, Stedman K, Prangishvili D, Holz I, Kristjansson JK, Garrett R, Zillig W. (1999) "**The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus**".Mol Microbiol. **34,** 217-26.

Ausili A, Di Lauro B, Cobucci-Ponzano B, Bertoli E, Scirè A, Rossi M, Tanfani F, Moracci M. (2004) "**Two-dimensional IR correlation spectroscopy of mutants of the beta-glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus* identifies the mechanism of quaternary structure stabilization and unravels the sequence of thermal unfolding events**". Biochem J. 2004 **384** 69-78.

Bartolucci S, Rossi M, Cannio R. (2003) "**Characterization and functional complementation of a nonlethal deletion in the chromosome of a beta-glycosidase mutant of *Sulfolobus solfataricus***".J Bacteriol. **185**, 3948-57.

Bettstetter M, Peng X, Garrett RA, Prangishvili D (2003) "**AFV1, a novel virus infecting hyperthermophilic archaea of the genus *acidianus***". Virology **315**, 68–79

Blake Wiedenheft, Kenneth Stedman, Francisco Roberto, Deborah Willits, Anne-Kathrin Gleske, Luisa Zoeller, Jamie Snyder, Trevor Douglas, and Mark Young (2004) "**Comparative Genomic Analysis of Hyperthermophilic Archaeal Fuselloviridae Viruses**" Journal of Virology, February **78**, 1954-1961

Bourne Y, Henrissat B. (2001) **"Glycoside hydrolases and glycosyltransferases: families and functional modules"**. Curr Opin Struct Biol. **11**, 593-600.

Cannio R, Di Prizito N, Rossi M, Morana A. (2004) "**A xylan-degrading strain of *Sulfolobus solfataricus*: isolation and characterization of the xylanase activity**". Extremophiles. Apr; **8**,117-24.

Cannio R, Contursi P, Rossi M, Bartolucci S. (2001) "**Thermoadaptation of a mesophilic hygromycin B phosphotransferase by directed evolution in hyperthermophilic Archaea: selection of a stable genetic marker for DNA transfer into *Sulfolobus solfataricus***".Extremophiles. **5**, 153-9.

Crennell SJ, Hreggvidsson GO, Nordberg Karlsson E. (2002) "**The structure of *Rhodothermus marinus* Cel12A, a highly thermostable family 12 endoglucanase, at 1.8 A resolution**". J Mol Biol. **320**, 883-97

Cubellis,M.V., Rozzo,C., Monteucchi,P. and Rossi,M. (1990) "**Isolation and sequencing of a new beta-glycosidase-encoding archaebacterial gene**" Gene, **94**, 89–94.

D'Auria S, Moracci M, Febbraio F, Tanfani F, Nucci R, Rossi M. (1998) "**Structure-function studies on beta-glycosidase from *Sulfolobus solfataricus*. Molecular bases of thermostability"**.Biochimie. **80**, 949-57.

Davies G, Henrissat B. (1995) "**Structures and mechanisms of glycosyl hydrolases"**. Structure **3**, 853-9.

Elferink MG, Albers SV, Konings WN, Driessen AJ. (2001) **"Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters".** Mol Microbiol. **39**, 1494-503.

Erauso G, Stedman KM, van de Werken HJ, Zillig W, van der Oost J. (2006) „**Two novel conjugative plasmids from a single strain of *Sulfolobus*"**. Microbiology. 152, 1951-68.

Ge M, Xia XY, Pan XM. (2008) "**Salt bridges in the hyperthermophilic protein Ssh10b are resilient to temperature increases"**. J Biol Chem. **8**

Georgieva TI, Ahring BK. (2007) "**Potential of agroindustrial waste from olive oil industry for fuel ethanol production"**.Biotechnol J. **2**, 1547-55.

Gorke JT, Okrasa K, Louwagie A, Kazlauskas RJ, Srienc F. (2007) "**Enzymatic synthesis of poly(hydroxyalkanoates) in ionic liquids"**.J Biotechnol. **132**, 306-13.

Greve B, Jensen S, Brügger K, Zillig W, Garrett RA. (2004) "**Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*"**. Archaea. **1**, 231-9.

Grogan, D., P. Palm, and W. Zillig. (1990). "**Isolate B12, which harbours a virus-like element, represents a new species of the archaebacterial genus *Sulfolobus*, *Sulfolobus shibatae***", Arch. Microbiol. **154**, 594-599.

Haack KR, Roth JR. (1995) "**Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*".** Genetics**.141**, 1245-52.

Haring, M., Rachel, R., Peng, X., Garrett, R.A., Prangishvili, D. (2005a). "**Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique**

archaeal virus, *Acidianus* bottle-shaped virus, from a new family, the Ampullaviridae" J. Virol. **79**, 9904-9911.

Häring M, Rachel R, Peng X, Garrett RA, Prangishvili D. (2005b) "**Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, *Acidianus* bottle-shaped virus, from a new family, the *Ampullaviridae***".J Virol. **79**, 9904-11.

Huang Y, Krauss G, Cottaz S, Driguez H, Lipps G. (2005) "**A highly acid-stable and thermostable endo-beta-glucanase from the thermoacidophilic archaeon *Sulfolobus solfataricus*."** Biochem J. **385**, 581-8

Jaenicke R. (1998) "**What ultrastable globular proteins teach us about protein stabilization"**.Biochemistry (Mosc). **63**, 312-21.

Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C. (2003) "**A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector"**.Mol Microbiol. **48**, 1241-52.

Keeling PJ, Klenk HP, Singh RK, Feeley O, Schleper C, Zillig W, Doolittle WF, Sensen CW. (1996) "**Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1".** Plasmid. **35**, 141-4.

Kim YO, Kim HW, Lee JH, Kim KK, Lee SJ. (2006) "**Molecular cloning of the phytase gene from *Citrobacter braakii* and its expression in *Saccharomyces cerevisiae***". Biotechnol Lett. **28**. 33-8.

Kleckner N, Chan RK, Tye BK, Botstein D. (1975) "**Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition"**.J Mol Biol. **97** 75

Koseki T, Takahashi K, Handa T, Yamane Y, Fushinobu S, Hashizume K. (2006) **"N-linked oligosaccharides of *Aspergillus awamori* feruloyl esterase are important for thermostability and catalysis"**. Biosci Biotechnol Biochem. **70**, 2476-80.

Kragl U, Eckstein M, Kaftzik N.(2002) "**Enzyme catalysis in ionic liquids"**.Curr Opin Biotechnol. **13**, 565-71.

Lee SH, Ha SH, Hiep NM, Chang WJ, Koo YM. (2008) "**Lipase-catalyzed synthesis of glucose fatty acid ester using ionic liquids mixtures".**J Biotechnol. **133**, 486-9.

Limauro D, Cannio R, Fiorentino G, Rossi M, Bartolucci S. (2001) "**Identification and molecular characterization of an endoglucanase gene, *celS*, from the extremely thermophilic archaeon *Sulfolobus solfataricus***." Extremophiles. **5**, 213-9.

Lipps G (2006) "**Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus***". Extremophiles.**10**,17-28.

Lipps G, Röther S, Hart C, Krauss G. (2003) ”**A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity**”.EMBO J. **22**, 2516-25.

Lubelska JM, Jonuscheit M, Schleper C, Albers SV, Driessen AJ. (2006) “**Regulation of expression of the arabinose and glucose transporter genes in the thermophilic archaeon *Sulfolobus solfataricus***”. Extremophiles. **10**, 383-91.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. (2002) “**Microbial cellulose utilization: fundamentals and biotechnology”**.Microbiol Mol Biol Rev. **66**, 506-77

Mahillon J, Chandler M. “**Insertion sequences**”. (1998) Microbiol Mol Biol Rev. **62**, 725-74.

Martin, A., S. Yeats, D. Janekovic, W.-D. Reiter, W. Aicher, and W. Zillig. (1984). “**SAV1, a temperate U. V. inducible DNA virus like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12**”. EMBO J. **3**, 2165-2168.

Martusewitsch E, Sensen CW, Schleper C. (2000) “ **High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements”**. J Bacteriol. **182**, 2574-81

Moracci M, Nucci R, Febbraio F, Vaccaro C, Vespa N, La Cara F, Rossi M. (1995) “**Expression and extensive characterization of a beta-glycosidase from the extreme thermoacidophilic archaeon *Sulfolobus solfataricus* in *Escherichia coli*: authenticity of the recombinant enzyme”**. Enzyme Microb Technol. **17**, 992-7.

Moracci M, La Volpe A, Pulitzer JF, Rossi M, Ciaramella M**.** (1992) **“Expression of the thermostable beta-galactosidase gene from the archaebacterium *Sulfolobus solfataricus* in *Saccharomyces cerevisiae* and characterization of a new inducible promoter for heterologous expression”**. J Bacteriol. **174**, 873-82.

Muskhelishvili, G., P. Palm, and W. Zillig. (1993). “**SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*”**. Mol. Gen. Genet. **237**, 334-342.

Nadal, M., G. Mirambeau, P. Forterre, W.-D. Reiter, and M. Duguet. (1986). “**Positively supercoiled DNA in a virus-like particle of an archaebacterium**”. Nature **321**, 256-258.

Nelson N. (1944) “**A photometric adaptation of the Somogyi method for the determination of glucose**”. J Biol Chem **153,** 375–380.

Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W.-D. Reiter, and W. Zillig. (1991). “**Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*”**. Virology **185**, 242-250.

Peng X. (2008) “**Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of *Sulfolobus* and the implications for plasmid survival**”.Microbiology. **154**, 383-91.

Peng X, Holz I, Zillig W, Garrett RA, She Q. (2000) "**Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*"** J Mol Biol. **303**, 449-54

Perugino G, Trincone A, Rossi M, Moracci M. (2004) "**Oligosaccharide synthesis by glycosynthases".**Trends Biotechnol. **22,** 31-7.

Pisani FM, Rella R, Raia CA, Rozzo C, Nucci R, Gambacorta A, De Rosa M, Rossi M (1990) "**Thermostable beta-galactosidase from the archaebacterium *Sulfolobus solfataricus*. Purification and properties**".Eur J Biochem. **187**, 321-8.

Prangishvili D, Forterre P and Garrett R A. (2006a) "**Viruses of the Archaea: a unifying view"** Nature Reviews Microbiology **4**, 837-848.

Prangishvili D, Garrett RA, Koonin EV. (2006b)"**Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life**".Virus Res.;**117**, 52-67.

Prangishvili D., Arnold H.P., Gotz D., Ziese U., Holz I., Kristjansson J.K. and Zillig W., (1999) "**A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of the *sulfolobus* viruses SIRV1 and SIRV2"**, Genetics **152**, 1387–1396.

Prangishvili D, Albers SV, Holz I, Arnold HP, Stedman K, Klein T, Singh H, Hiort J, Schweier A, Kristjansson JK, Zillig W. (1998) "**Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*"**. Plasmid. **40**, 190-202.

Prato S, Vitale RM, Contursi P, Lipps G, Saviano M, Rossi M, Bartolucci S. (2008) **"Molecular modeling and functional characterization of the monomeric primase-polymerase domain from the *Sulfolobus solfataricus* plasmid pIT3".** FEBS J. **275,** 4389-402.

Prato S, Cannio R, Klenk HP, Contursi P, Rossi M, Bartolucci S. (2006) "**pIT3, a cryptic plasmid isolated from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* IT3".** Plasmid. **56**, 35-45.

Prentki P, Teter B, Chandler M, Galas DJ. (1986) "**Functional promoters created by the insertion of transposable element IS1"**.J Mol Biol. **191**, 383-93.

Purschke WG, Schäfer G. (2001) "**Independent replication of the plasmids pRN1 and pRN2 in the archaeon *Sulfolobus islandicus***". FEMS Microbiol Lett. **200**, 97-102.

Redder P, She Q, Garrett RA (2001) "**Non-autonomous mobile elements in the crenarchaeon *Sulfolobus solfataricus***".J Mol Biol. **306**, 1-6.

Reiter, W.-D., P. Palm, S. Yeats, and W. Zillig. (1987). "**Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1"**. Mol. Gen. Genet. **209**, 270-275

Rezsöhazy R, Hallet B, Delcour J, Mahillon J. (1993) ”**The IS4 family of insertion sequences: evidence for a conserved transposase motif**”. Mol Microbiol. **9**, 1283-95.

Rice G, Tang L, Stedman K, Roberto F, Spuhler J, Gillitzer E, Johnson JE, Douglas T, Young M (2004) “**The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life”**. Proc Natl Acad Sci U S A. **101**, 7716-20.

Rice, G, Stedman, K, Snyder, J, Wiedenheft, B, Willits, D, Brumfield, S, McDermott, T and Young M. J. (2001) **“Viruses from extreme thermal environments”** Proc Natl Acad Sci USA **98**, 13341-13345.

Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjbar B. (2006) “**Effective factors in thermostability of thermophilic proteins”**. Biophys Chem. **119**, 256-70.

Sambrook J, Russell DW (2001) “**Molecular cloning, a laboratory manual**”. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schleper C, Holz I, Janekovic D, Murphy J, Zillig W (1995) „**A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating”**. J Bacteriol. **177**, 4417-26.

Schleper C, Röder R, Singer T, Zillig W. (1994) “**An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous beta-galactosidase gene**”. Mol Gen Genet.;**243**, 91-6.

Schleper C., Kubo K. and Zillig W. (1992) “**The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA”**, Proc. Natl. Acad. Sci. U.S.A. **89** 7645–7649.

Serre, M. C., Letzelter C., Garel J. R., and Duguet M. (2002) „**Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase”**. J. Biol. Chem. **277**, 16758-16767

Sjöde A, Alriksson B, Jönsson LJ, Nilvebrant NO. (2007) “**The potential in bioethanol production from waste fiber sludges in pulp mill-based biorefineries”**.Appl Biochem Biotechnol. Apr; **137-140**, 327-37.

Snyder JC, Stedman K, Rice G, Wiedenheft B, Spuhler J, Young MJ (2003) “**Viruses of hyperthermophilic Archaea”.** Res Microbiol **154**, 474–482.

Stedman, K. M., Q. She, H. Phan, H. P. Arnold, I. Holz, R. A. Garrett, and W. Zillig. (2003). “**Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2”**. Res. Microbiol. **154**, 295-302.

Stedman KM, She Q, Phan H, Holz I, Singh H, Prangishvili D, Garrett R, Zillig W. (2000) **“pING family of conjugative plasmids from the extremely thermophilic**

archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota". J Bacteriol. **182**, 7014-20.

Stedman KM, Schleper C, Rumpf E, Zillig W. (1999) "**Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors"**.Genetics. **152**, 1397-405.

Swatloski RP, Spear SK, Holbrey JD, Rogers RD. (2002) "**Dissolution of cellulose [correction of cellose] with ionic liquids**." J Am Chem Soc. **124**. 4974-5.

Teeri TT, Koivula A, Linder M, Wohlfahrt G, Divne C, Jones TA. (1998) "***Trichoderma reesei* cellobiohydrolases: why so efficient on crystalline cellulose?"**Biochem Soc Trans. **26**, 173-8.

Uccelletti D, Farina F, Rufini S, Magnelli P, Abeijon C, Palleschi C. (2006) **"The *Kluyveromyces lactis* alpha1,6-mannosyltransferase KlOch1p is required for cell-wall organization and proper functioning of the secretory pathway".** FEMS Yeast Res. **6**, 449-57.

van Ooyen AJ, Dekker P, Huang M, Olsthoorn MM, Jacobs DI, Colussi PA, Taron CH (2006). **"Heterologous protein production in the yeast *Kluyveromyces lactis***". FEMS Yeast Res. **6** 381-92.

van Rantwijk F, Sheldon RA. (2007) "**Biocatalysis in ionic liquids".** Chem Rev. **107**, 2757-85.

Wang Y, Duan Z, Zhu H, Guo X, Wang Z, Zhou J, She Q, Huang L. (2007) "**A novel *Sulfolobus* non-conjugative extrachromosomal genetic element capable of integration into the host genome and spreading in the presence of a fusellovirus**". Virology. **363**, 124-33.

Watts AG, Withers SG. (2004) "**Glycosynthase-catalysed formation of modified polysaccharide microstructures"** .Biochem J. **380**, 9-10.

Witte A., Baranyi U., Klein R., Sulzner M., Luo C., Wanner G., Kruger D.H. and Lubitz W., (1997) "**Characterization of *Natronobacterium magadii* phage phi Ch1, a unique archaeal phage containing DNA and RNA"** Mol. Microbiol*.* **23** 603–616

Yeats, S., P. McWilliam, and W. Zillig. (1982). **"A plasmid in the archaebacterium *Sulfolobus acidocaldarius***". EMBO J. **1**, 1035-1038.

Zillig W, Arnold HP, Holz I, Prangishvili D, Schweier A, Stedman K, She Q, Phan H, Garrett R, Kristjansson JK (1998) "**Genetic elements in the extremely thermophilic archaeon *Sulfolobus***". Extremophiles. **2**. 131-40.

.

Pubblications:

Aucelli T., Contursi P., Girfoglio M., Rossi M., Cannio R.
"**A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus***."
Nucleic Acids Res. (2006)

Comunications:

Girfoglio M., Rossi M., Cannio R.
"**Acetate kinase from moderately thermophilic and hyperthermophilic sources: structure-function relationship**"
International Conference on Extremophiles, 17-21 Sept. 2006 Brest, France

Published online 13 September 2006 *Nucleic Acids Research, 2006, Vol. 34, No. 17* ***e114***
doi:10.1093/nar/gkl615

A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*

Tiziana Aucelli[1], Patrizia Contursi[1], Michele Girfoglio, Mosè Rossi and Raffaele Cannio*

Istituto di Biochimica delle Proteine, Via P. Castellino 111, 80131 CNR, Napoli, Italy and [1]Dipartimento di Biologia Strutturale e Funzionale, Università degli Studi di Napoli Federico II, Via Cinthia, 80126 Napoli, Italy

Received July 5, 2006; Revised and Accepted August 5, 2006

ABSTRACT

The pSSVx genetic element from *Sulfolobus islandicus* REY15/4 is a hybrid between a plasmid and a fusellovirus, able to be maintained in non-integrative form and to spread when the helper SSV2 virus is present in the cells. In this work, the satellite virus was engineered to obtain an *Escherichia coli–Sulfolobus solfataricus* shuttle vector for gene transfer and expression in *S.solfataricus* by fusing site-specifically the pSSVx chromosome with an *E.coli* plasmid replicon and the ampicillin resistance gene. The pSSVx-based vector was proven functional like the parental virus, namely it was able to spread efficiently through infected *S.solfataricus* cells. Moreover, the hybrid plasmid stably transformed *S.solfataricus* and propagated with no rearrangement, recombination or integration into the host chromosome. The high copy number of the artificial genetic element was found comparable with that calculated for the wild-type pSSVx in the new host cells, with no need of genetic markers for vector maintenance in the cells and for transfomant enrichment.

The newly constructed vector was also shown to be an efficient cloning vehicle for the expression of passenger genes in *S.solfataricus*. In fact, a derivative plasmid carrying an expression cassette of the lacS gene encoding the β-glycosidase from *S.solfataricus* under the control of the *Sulfolobus* chaperonine (thermosome tf55) heat shock promoter was also able to drive the expression of a functional enzyme. Complementation of the β-galactosidase deficiency in a deletion mutant strain of *S.solfataricus* demonstrated that *lacS* gene was an efficient marker for selection of single transformants on solid minimal lactose medium.

INTRODUCTION

Host/virus interaction modes have provided windows to study microbial diversity (1) as well as genetic processes at the molecular level, in particular for prokaryotes, and hence have helped in clarifying the physiological mechanisms, the dependence on the specific biochemical environment and evolution of their host cells (2,3).

Very few viruses have been identified from Archaea (4) as compared with Bacteria and Eukarya and detailed description has been reported for those from hyperthermophilic archaea (5,6) with representatives that replicate in the genus *Sulfolobus* being the majority within the kingdom Crenarchaeota (6–8). To date, the *Fuselloviridae* are the most widespread on earth in the *Sulfolobus* genus with viruses sharing similar morphology as well as DNA genome size and organization (9–12).

Sulfolobus spindle-shaped virus 1 (SSV1) is the best studied member of this family and demonstrated to be temperate both in *Sulfolobus shibatae* and in non-natural but related *Sulfolobus* hosts, such as *Sulfolobus solfataricus* (13,14); infection, integration of DNA into the host chromosome and production of virions cause apparently no phenotype change but a significant growth retardation of the host cells which can be visualized as turbid plaques around propagation foci on plated lawns of indicator host cells (14–18).

More recently, another fusellovirus, SSV2 from *Sulfolobus islandicus* strain 15/4 was isolated, characterized and its complete genomic sequence determined. SSV2 shares with SSV1 similar morphology, replication and DNA size (19). The overall genome architecture is conserved but the low similarity in the sequences should be responsible for the higher copy number and the lack of a strong ultraviolet induction of episomal SSV2 DNA and particle production, as well as for the different integration of the SSV2 genome which occurs into the host chromosome at the site of a glycyl tRNA instead of arginyl-tRNA (12).

S.islandicus REY15/4 harbors also a small plasmid, pSSVx, assigned to the pRN family (20,21) of *Sulfolobales* plasmids; pSSVx is also capable of spreading in the cell cultures of

*To whom correspondence should be addressed at Istituto di Biochimica delle Proteine, Consiglio Nazionale delle Ricerche, Via Pietro Castellino 111, 80131, Naples, Italy. Tel: +39 081 613 2285; Fax: +39 081 613 2248; Email: r.cannio@ibp.cnr.it

S.solfataricus but only in the presence of either SSV2 or SSV1, necessary as helpers (9). In fact, pSSVx contains two open reading frames showing high-sequence similarity to a tandem of ORFs in both SSV1 and SSV2 genomes; the proteins encoded by these ORFs are probably necessary for specific recognition of the pSSVx DNA but need viral helper components for capsid formation and packaging (9,19).

In general, the choice of *S.solfataricus* as a model for fundamental understanding of the genetics of extremely thermophilic archaea is due to growth conditions operatively non-prohibitive (22) and capability of maintaining and propagating either natural or genetically modified extrachromosomal DNAs (23,24) from other sources. The complete genome of *S.solfataricus* has also been determined (25), the biochemical characterization of many gene products obtained (26) and the development of post-genomics tools such as proteomics and metabolic pathway reconstruction recently attempted (27,28). Some progress has been made to develop stable transformation (29–32), specific gene disruption methods (33) as well as overexpression of foreign and homologous genes (34); nevertheless none of the systems described so far has been proven efficient for reproducibility and stability of gene cloning and protein expression levels in *Sulfolobus*, probably due, in most cases, to low transformation efficiencies, inefficient selection and/or instability of the vectors in the host as well as changes in the ratio episomal versus integrative forms occurring in the cell during replication of virus-derived constructs (35).

In this study, a genetic system for *Sulfolobus* was developed that is based on the satellite virus pSSVx from *S. islandicus* 15/4. The different recombinant *Escherichia coli–Sulfolobus solfataricus* shuttle vectors constructed retained the wild-type capability to replicate at high copy-number and to spread in cell cultures in the presence of its helper virus SSV2. *Sulfolobus* transformants were demonstrated to be stable and propagate the pSSVx derived plasmids in a reproducible and constant fashion without any rearrangement, recombination or integration into the chromosome.

Moreover, stable complementation of a β-galactosidase mutant of *S.solfataricus* previously isolated and characterized in our laboratory (32) and reproducible gene expression levels were also obtained by introducing the β-galactosidase gene (*lacS*) as a reporter under the control of a strong and heat-inducible promoter into the shuttle vector.

MATERIALS AND METHODS

Growth of *Sulfolobus* strains and isolation of SSV2-infected *S.solfataricus* GθW and P2

S.solfataricus strains P2 (DSM 1617), Gθ (23) and the derivative mutant GθW [Δ*lacS*, (32)] as well as *S.islandicus* REY 15/4 (22) were grown at 75 or 80°C in glycine buffered Brock's medium (36) with 0.1% tryptone, 0.05% yeast extract and 0.2% sucrose at pH 3.2. For electroporation and plaque assays, cells were grown with phosphate buffered medium N.182 (M182) suggested by the DSMZ Catalogue of strains containing 0.1% glucose. For isolation of independent clones, medium M182 contained only 0.25% lactose and no other nutrient. The optical density of liquid cultures was monitored at 600 nm. For solid media, gellan gum (Gelrite, Sigma) was added to a final concentration of 0.8% (0.35% for overlays), and $MgCl_2$ and $CaCl_2$ were added to 10.0 and 3.0 mM, respectively.

The *S.solfataricus* strains GθW and P2 transfected with the SSV2 and pSSVx were extracted from zones of growth inhibition (plaques) formed on indicator lawns around spots of the *S.islandicus* REY 15/4 culture supernatants (2–6 μl for each spot), as described by Arnold *et al*. (9). Clones cured selectively for pSSVx were obtained by isolation of single colonies formed on plates by the cells extracted from plaques and revitalized in liquid cultures.

Viral DNA isolation and plasmid constructions

Extrachromosomal SSV2 and pSSVx DNAs from both *S.islandicus* REY 15/4 and transfectants of *S.solfataricus* strains were performed with Qiaprep Spin Miniprep kit (Qiagen) following the standard procedure suggested by the manufacturer for *E.coli* cells.

Plasmid pSSVrt was constructed by cloning the pSSVx DNA linearized at the position 2812 with AflIII and modified with Klenow DNA polymerase into the SmaI site of the pUC19 *E.coli* vector. Linearization was obtained by AflIII partial digestion in the presence of ethidium bromide using a protocol already described for the SSV1 virus DNA linearized with Sau3AI (37). Clones with insertion at the specific position were selected by restriction analysis of the resulting plasmid collection. A size reduced derivative of pSSVrt was also constructed; the polycloning sequence (between the AatII and EcoRI sites) of the pUC28 vector was inserted into a 1812 bp AatII/AflIII DNA fragment from pUC18 after suitable modification of incompatible ends. The pSSVx sequence was excised from the pSSVrt plasmid with SacI and PstI and inserted into the same sites of the minimal plasmid obtained to produce the pMSSV vector. An expression cassette of the *lacS* gene (38) was PCR amplified by the vector pMJ03 (35) and inserted between the XhoI and PstI sites of pMSSV generating the expression vector pMSSV*lacS*. Excision of the *E.coli* minimal plasmid was obtained by digestion of pMSSV*lacS* (1.0 μg) with SacI and purification of the pSSVx/*lacS* moiety from agarose gel. The DNA was re-circularized by ligation in a final volume of 5.0 μl, diluted in water to 20 ng/μl final concentration, and 2 μl were used for electroporation.

Transformation procedure and analyses of *Sulfolobus* transformants

S.solfataricus cells of SSV2 lysogenes were grown up to mid-logarithmic phase (0.3–0.45 OD_{600}), harvested by centrifugation and repeatedly washed in 20 mM sucrose as described previously by Schleper *et al*. (14). Aliquots of 10^{10} cells/ml (50 μl) were mixed with 1 μl DNA (10–100 ng/μl), incubated for 1 min on ice and transferred to chilled plastic cuvettes with an electrode gap of 0.1 cm (BioRad). High voltage electroporation (25 μF) was performed with a BioRad Gene Pulser Xcell™ at a field strength of 1.5 kV/cm and 400 Ω resistance; two successive shock pulses were applied to competent cells producing pulse length of ∼10.0 and 9.8 ms, respectively. Immediately after electroporation cell mixtures were diluted with 1 ml of medium M182 containing 0.1%

glucose, transferred to glass vials and incubated for 3 h at 75 or 80°C. After suitable scale-up, 5–15 ml aliquots of the cultures (the volumes varying in order to withdraw the same number of cells per aliquot) were harvested at increasing cell density for DNA extractions. For monitoring propagation of the recombinant satellite virus, extrachromosomal DNA mini preparations and plaque assays were performed. Transformants were stored at −80°C in 15% glycerol stocks.

For Southern blot analysis, ∼2 μg of total cellular DNAs, extracted according to Arnold *et al*. (9), and 5 ng of pMSSV*lacS* plasmid purified from *E.coli*, were cut with HindIII and BglII, and electrophoresed in a 0.8% agarose gel; DNA digests were blotted and hybridized according to standard procedures (39). The probe was prepared by cutting out and purifying a HindIII restriction fragment from the pMSSV*lacS* vector encompassing the *lacS* gene and a portion (up to −302) of the tf55α 5′ flanking region. This restriction DNA fragment was randomly labeled using the random prime DNA labeling kit (Boehringer Mannheim).

β-galactosidase complementation and isolation of mixed and single transformants

pMSSV*lacS* transformed cells were tested for β-galactosidase activity; 1 ml aliquots of cultures were centrifuged and cell pellets overlaid with a X-Gal solution (2 mg/ml in phosphate buffered medium) and incubated for 15 min at 75°C for blue color development. *In situ* assays were performed on the same cultures seeded on plates and grown as circular colonized areas, as previously described (32). Supernatants of cultures grown up to 0.5, 1.0 and 1.3 OD_{600} were checked for plaque formation as already described and analyzed by X-Gal staining of plaques formed on continuous lawns of the *S.solfataricus* GθW strain. For infection in liquid culture, 400 μl supernatants from pMSSV*lacS* transformants were added to a 20 ml culture of cells transfected only with SSV2 and grown up to 0.5 OD_{600}. After incubation under shaking for 48 h, cells were diluted 1:50, grown up to 0.36 and 1.0 OD_{600} and tested for β-galactosidase activity as already described.

Single transformants were selected either on rich (M182, glucose 0.1%) or on minimal (lactose 0.25% as the only nutrient) solid media, by plating or by streaking of electropored cells propagated until they showed positive staining with X-Gal and diluted after they had reached different cell densities (0.3–1.0 OD_{600}). Single clones formed after 2 weeks of incubation at 75°C were stained with X-Gal (incubation at 75°C for 3–4 h for color development on lactose plates), picked and resuspended in 100 μl lactose minimal medium and seeded as spots onto fresh rich (M182) medium plates. After *in situ* X-Gal test on plates, cell spots were transferred to M182 medium containing glucose, propagated up to early stationary phase (1.0 OD_{600}) and analyzed for extrachromosomal DNA content as described above.

β-galactosidase enzyme activity

β-galactosidase activity of transformants was visualized and quantified as follows. Crude extracts were prepared by a freeze thaw technique, suspension of the cells in 10 mM Tris–HCl buffer, pH 8.0, placing at 80°C for 10 min, then at 50°C for 5 min. This procedure was repeated four times, and the extract was spun for 30 min at ∼10 000*g*. The supernatant was either assayed immediately or stored at −80°C before testing. Protein concentrations of the crude cell extracts were determined by the Bradford assay (Bio-Rad). Detection of the β-galacosidase in protein extracts was performed on 10% acrylamide SDS–PAGE gels after electrophoretic separation and extensive washing in 10 mM Tris–HCl buffer, pH 8.0. The specific enzyme band was visualized incubating the gel 30 min at 75°C in the same buffer containing 2 mg/ml X-gal.

The β-galactosidase assay procedure followed essentially the protocol of Pisani *et al*. (40). A sample of 10 μl extract was transferred to a preheated (75°C) quartz cuvette containing 990 μl assay buffer (2.8 mM ONPG in 50 mM sodium phosphate buffer, pH 6.5). The ONPG hydrolysis reaction was followed spectrophotometrically at 75°C by measuring the increase in absorbance at 405 nm in a Beckman spectrophotometer with heatable cuvettes. One unit was defined as the amount of enzyme catalyzing the hydrolysis of 1 μmol of ONPG min^{-1} at 75°C with a molar absorption coefficient of 3100 M^{-1} cm^{-1} at 405 nm for ONP.

RESULTS

Transfection and isolation of *Sulfolobus solfataricus* SSV2 transfectants

The strains P2 and GθW, a β-glycosidase defective mutant isolated earlier (32), were tested for susceptibility to infection by SSV2 and pSSVx viruses produced by *S.islandicus* REY15/4. Supernatants of REY15/4 cultures generated growth inhibition halos on continuous lawns of both *S.solfataricus* strains (Figure 1A), similar to those observed for the strain P1 (9); like for SSV1 virus (37), plaques did not form on *Sulfolobus acidocaldarius* cells, namely no infection occurred. Cells extracted from areas of these plaques and revitalized in liquid cultures contained extrachromosomal DNA indistinguishable from that isolated from REY15/4, indicating the presence and active replication of both virus and satellite elements (Figure 1B). From this culture, clones infected only by the SSV2 virus, namely cured for the pSSVx, were isolated as single colonies on plates after suitable dilutions, as described for the P1 strain by Arnold *et al*. (9). The absence of the satellite virus was confirmed by restriction analysis of extrachromosomal DNA (Figure 1C) and Southern hybridization (data not shown).

Shuttle vector construction

A sequence analysis on the pSSVx DNA was performed in order to locate regions that could be manipulated without affecting DNA replication and particle proliferation/spreading, and thus representing candidate targets for site-specific insertion of foreign DNA sequences. A 200 bp segment was identified which contains the tail-to-tail intergenic region between the ORFs C68 and 288. The segment shows archaeal transcription termination signals (41) and tendency to form hairpin loops as for rho-independent termination mechanisms (42,43). Moreover, it has an AflIII site (cut at the position 2812 on the pSSVx genome), useful for cloning, since it is situated in the 3′ direction beyond the ORF288

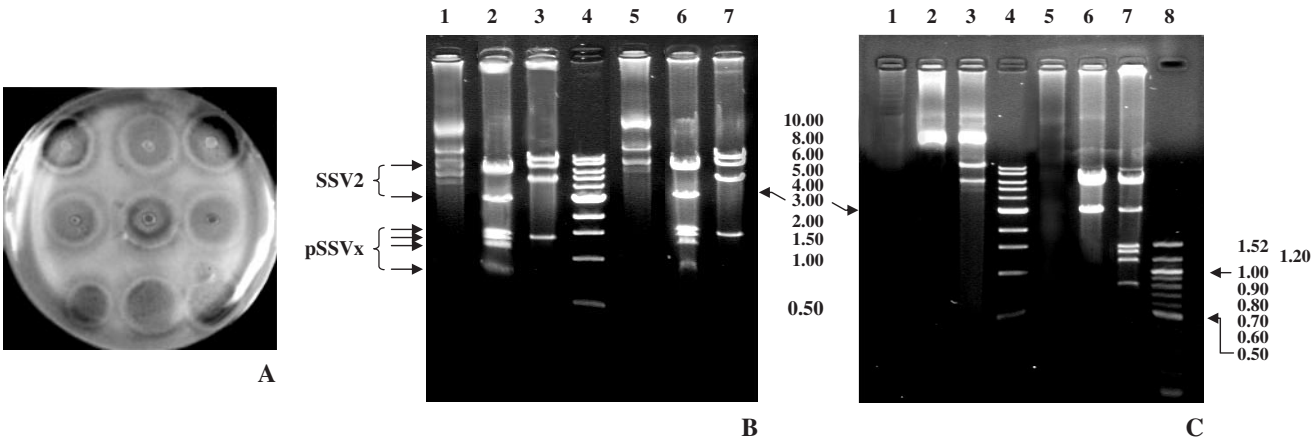

Figure 1. Transfection of *S.solfataricus* GθW with conditioned medium of *S.islandicus* REY15/4. The spontaneous mutant GθW lacking the β-glycosidase activity was first tested for infection with both SSV2 and pSSVx by formation of growth inhibition areas in cell lawns on plate. Virus plaques of SSV2 and pSSVx were obtained by spotting aliquots of REY 15/4 supernatants (2 µl for the three plaques in the first row, 4 µl for those in the second row and 6 µl for those in the third row) onto a continuous indicator lawn of the GθW cells (**A**). (**B**) Extrachromosomal DNAs (uncut in lanes 1 and 5), extracted from both the natural source *S.islandicus* (lanes 1–3) and the infected *S.solfataricus* (lanes 5–7), propagated after extraction from the plaques and revitalization, was cut with EcoRI (lanes 2 and 6) and BamHI (lanes 3 and 7). Episomal DNAs were confirmed to be identical and to contain both SSV2 and pSSVx DNAs, as demonstrated by the two distinct EcoRI restriction patterns. Lane 4: DNA molecular weight standards (kb). (**C**) The cells entrapped inside the plaques have been extracted from the plates, revitalized in liquid medium and then streaked onto solid medium to obtain single colonies. The comparative analysis with extrachromosomal DNAs from non-transformed recipient cells (lanes 1 and 5) and from the cells transfected with both viral and satellite DNAs (lanes 3 and 7) revealed the absence of the pSSVx plasmid in one colony isolated on plate and analyzed (lanes 2 and 6). Lanes 1–3: uncut DNAs; lanes 5–7: EcoRI digests; lanes 4 and 8: DNA molecular weight standards (kb).

transcription termination region, and far upstream of the ORF c68 stop codon. Since the AflIII site is present in five copies on the pSSVx sequence (positions 677, 814, 982, 2812 and 4994), singly cut pSSVx genomes were generated by digestion with the restriction enzyme under conditions that allowed single cleavage of the DNA molecules presumably at every specific site (37). After suitable modification of termini, these DNAs were inserted into *E.coli* pUC19 plasmid vector and specific insertion at the position 2812 produced the fusion plasmid pSSVrt (Figure 2).

pSSVrt shuttle vector transfection and spreading

The plasmid DNA pSSVrt selected from *E.coli* was transferred by electroporation into *Sulfolobus*, using different strains, namely SSV2 lysogens of the strains P2, MT4 and GθW, which is a stable β-galactosidase mutant with an extended deletion in the *lacS* genetic locus (32). To test transformation efficiency, *S.solfataricus* was transformed with varying amounts of the plasmid pSSVrt (10–100 ng) and then checked for the presence and amount of the vector at different cell densities and after several generations. Three hours after electroporation no extrachromosomal DNA could be detected, confirming the low transformation efficiency already determined both by Schleper *et al*. (14) and Cannio *et al*. (23). Nevertheless the plasmid pSSVrt spread efficiently throughout a culture after transformation; in fact after the first scale-up from 3 to 50 ml of the culture, the amount of the plasmid and of the helper virus SSV2 increased and could be detected by ethidium bromide fluorescence on agarose gels (Figure 3). Identical results were obtained with all the different DNA concentrations used for electroporation but a lower amount (~5-fold) of the hybrid plasmid could be detected in the P2 and MT4 (data not shown) strains when compared with the GθW strain (Figure 3).

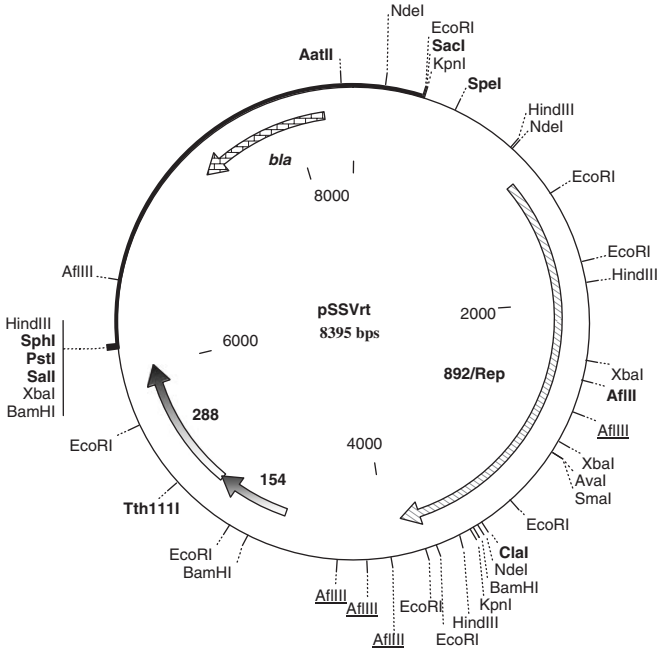

Figure 2. Plasmid map of the pSSVx/pUC19 shuttle vector. The pSSVrt shuttle vector has been obtained by cloning the pUC19 bacterial sequence (heavy bar) into an AflIII restriction site (all AflIII sites are underlined) downstream of the ORF288 of the pSSVx plasmid and by propagation in *E.coli*. Main ORFs allow easy location of archaeal (replicase ORF892/Rep as well as ORF154 and ORF288 putative DNA packaging components of pSSVx) and bacterial (*bla*, ampicillin resistance gene) sequences; bold face letters indicate unique restriction sites.

As expected, the same experiments performed on wild-type strains not infected with SSV2 produced no transformation by the shuttle vector, confirming the need of the helper virus for the propagation of the engineered plasmid.

A long-term growth experiment was also carried out with transformed *S.solfataricus* GθW. A culture containing pSSVrt was grown under standard conditions until it reached an optical density of 0.8, then diluted 1:10 in the same medium and grown and diluted again twice in the same manner. At this point samples were withdrawn at different cell densities and the DNA from the cells was analyzed; no change in the DNA replication and accumulation was observed (Figure 4) even after storing the propagated culture as a frozen −80°C glycerol stock and repeating the dilution/growth cycles. The relative fluorescence intensities of the pSSVrt and pSSVx were almost identical and the comparison with DNA fragments loaded at known concentrations on the gels allowed the estimation of ∼130 and 150 copies per cell (density of 1.0–1.2 OD_{600}) for the engineered and the unmodified viral DNAs.

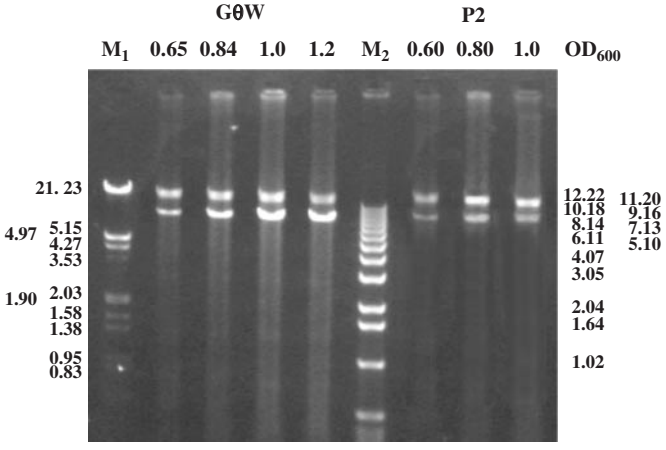

Figure 3. Transformation of *S.solfataricus* with the shuttle vector pSSVrt. Transfer into *S.solfataricus* GθW and P2 strains was performed by electroporation and after propagation of the cultures, the cells were harvested at different growth phases (indicated as optical density values at 600 nm on the top) for extrachromosomal DNA preparation. The upper and lower bands in each preparation correspond to SSV2 virus and recombinant satellite pSSVrt, respectively, both linearized with SalI. M_1 and M_2: DNA molecular weight markers (kb).

pSSVx-derived shuttle vectors for the *lac*S gene transfer and expression in the β-galactosidase deficient mutant GθW

β-D-galactosidase activity of *S.solfataricus* is displayed by the *lacS* gene product (40,44) and is responsible for the typical blue stain of the cells when exposed to the chromogenic substrate, X-gal; the lack of gene function in defective mutant strains is hence responsible for colorless colonies (45). The *lacS* coding sequence with a 648 bp 3′-untranslated region (3′-UTR) was fused to a 448 bp fragment containing the promoter region of tf55α and the first five codons of its ORF; this 2578 bp expression cassette was inserted into the polycloning site of the pSSVrt vector, increasing its size to ∼11 kb. A similar gene fusion was already demonstrated to efficiently complement a β-galactosidase defect in *Sulfolobus* when carried by a viral SSV1-derived shuttle vector (35). The plasmid was used to transform the strain GθW (Δ*lacS*) and after electroporation, cell culture was regenerated to allow spreading and subsequently seeded as spots onto plates as previously described (32). After 2–3 days of incubation, the colonized areas were overlaid with X-gal and reincubated at 75°C. No color was developed even after prolonged incubation, namely no expression of β-galactosidase activity could be detected, suggesting the failure of the recombinant plasmid to transport and/or to express the *lacS* gene. Extrachromosomal DNA preparation did not contain the recombinant pSSVx confirming that interference could have occurred either at the level of DNA propagation or of the recombinant particle proliferation/spreading. Southern analysis of extrachromosomal DNAs from this primary transformants demonstrated that the DNA transfer and maintenance were unaffected whereas spreading was impeded, as indicated by specific bands that could be detected at constant but very

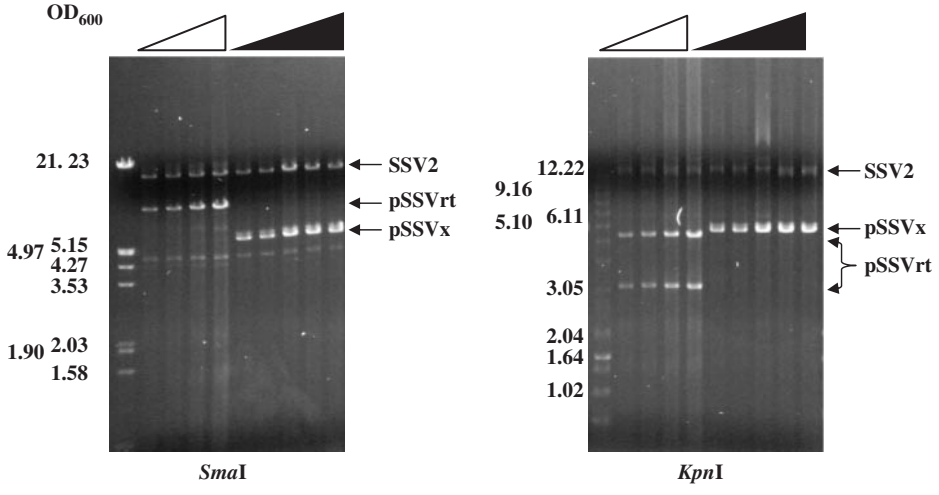

Figure 4. Maintenance and propagation of pSSVrt in *S.solfataricus*. After storage, SSV2 lysogens of *S.solfataricus* GθW transformed with pSSVrt and wild-type pSSVx were grown for several generations to check transformation stability. The propagated cultures were tested by restriction analysis of extrachromosomal DNAs (extracted at increasing cell densities from ∼0.5 to 1.2 OD_{600}); the SmaI enzyme produced single cuts in both satellite and SSV2 elements, whereas KpnI had 1 and 2 recognition sites in pSSVx and pSSVrt, respectively, and no site on SSV2. Standards of DNA molecular weight are indicated (kb).

weak intensity for cells withdrawn at different generation stages (data not shown). To confirm this preliminary results a size reduction of the pSSVrt vector was obtained by eliminating a redundant sequence in the *E.coli* plasmid moiety not necessary for replication and ampicillin selection and producing the vector pMSSV. Moreover a smaller (2025 bp) *lacS* expression cassette provided by Dr C. Schleper (35) and containing a shorter 3′-UTR replaced the one used in the first attempt. This *lacS* expression cassette was inserted into the polycloning site of pMSSV, generating the expression vector pMSSV*lacS* (Figure 5). Both pMSSV and pMSSV*lacS* were transferred into *S.solfataricus* GθW/SSV2

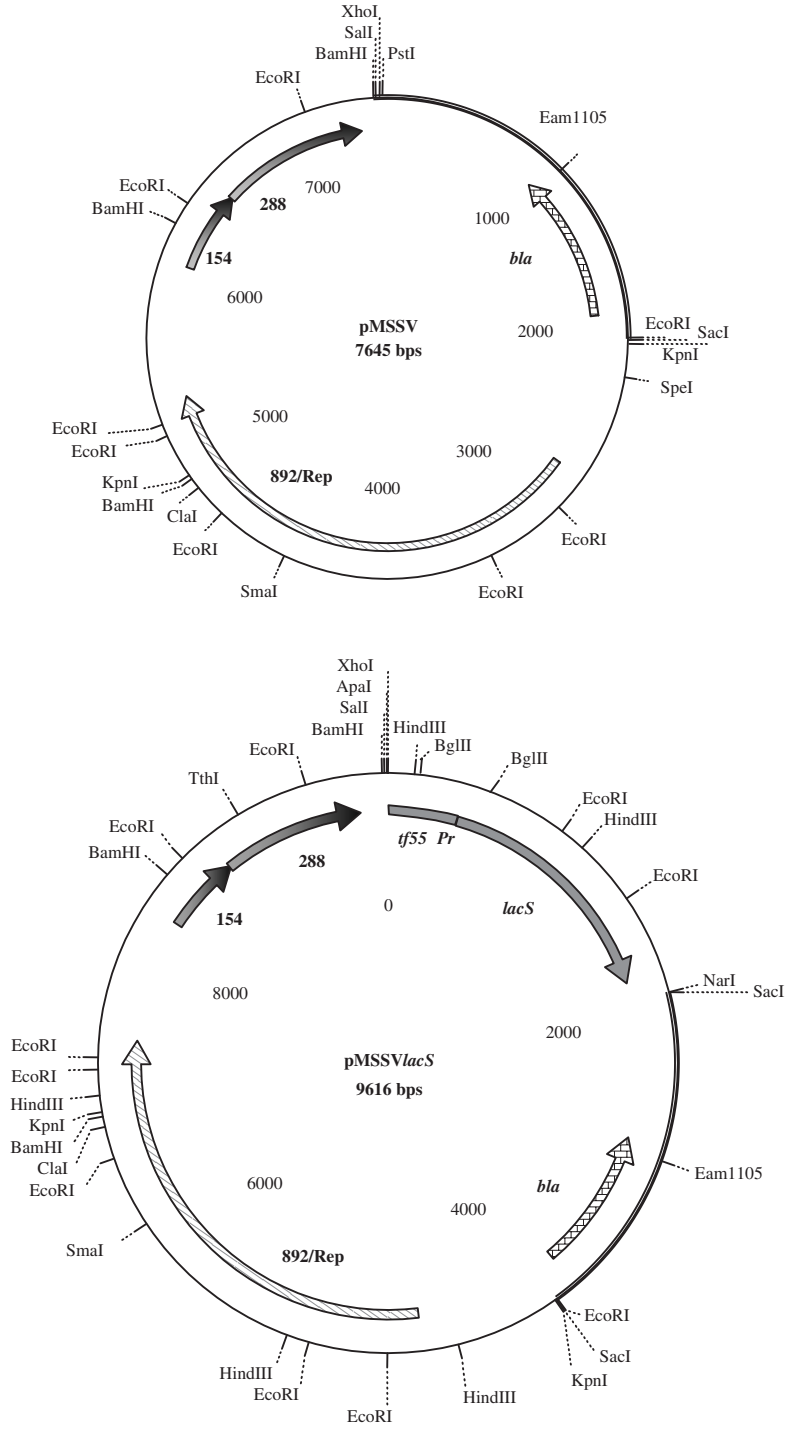

Figure 5. Plasmid maps of the minimal plasmid pMSSV and of the β-glycosidase expression vector pMSSV*lacS*. The bacterial moiety (indicated as a solid bar) of the plasmid pSSVrt was reduced in size, by eliminating every redundant sequence from the pUC19 vector and maintaining its ColE1 replicon and the ampicillin resistance marker (*bla*), to generate a new shuttle vector named pMSSV. An expression cassette containing the *lacS* gene fused to the thermosome tf55α subunit promoter was inserted into the newly constructed vector producing the expression vector pMSSV*lacS*.

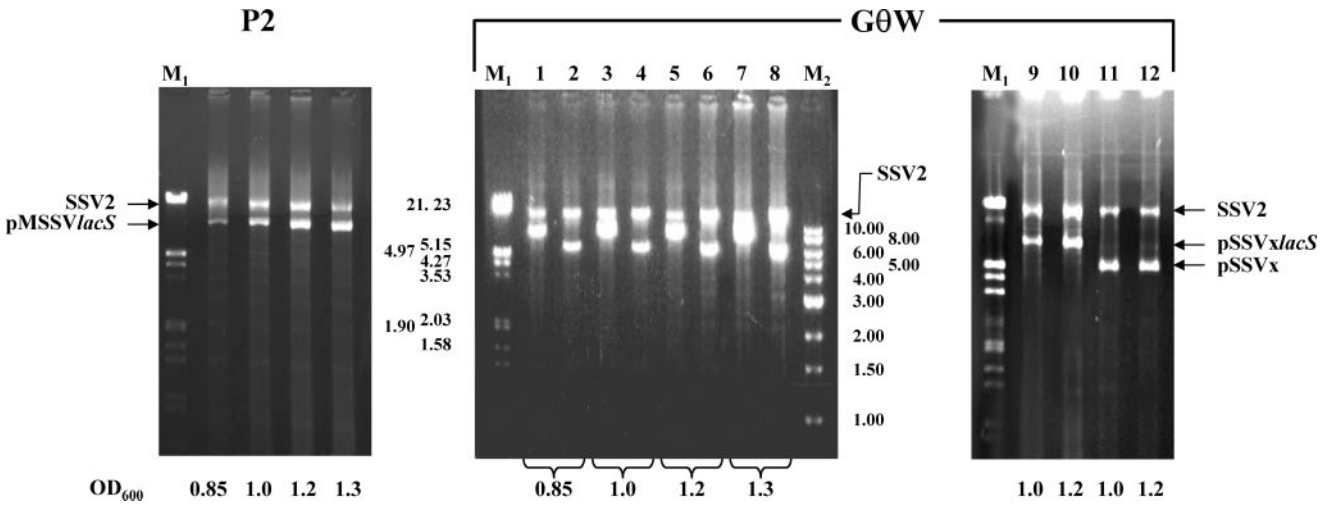

Figure 6. Transformation of *S.solfataricus* with the vectors pMSSV and pMSSV*lacS*. Competent cells of *S.solfataricus* P2 and GθW infected with the SSV2 virus were electroporated with the vector carrying the β-glycosidase gene pMSSV*lacS*. Strain GθW was also transformed with pSSVx*lacS*, a pMSSV*lacS*-derived plasmid obtained by excision of the *E.coli*-specific sequences and re-ligation. pMSSV and wild-type pSSVx were also transferred into the GθW/SSV2 lysogen for comparison of vector transfer and propagation efficiency in the presence and absence of the *lacS* gene. Extrachromosomal DNAs extracted from transformed cells after propagation of the cultures were checked by agarose electrophoresis for the presence of pMSSV*lacS* (all lanes in P2 and lanes 1, 3, 5 and 7 in GθW) and pMSSV (lanes 2, 4, 6 and 8 in GθW) at the different growth stages indicated as cell densities (OD_{600}). Similarly, recovery of the pSSVx*lacS* plasmid (SacI cut, lanes 9 and 10) was monitored in comparison to the parental pSSVx (SmaI cut, lanes 11 and 12). Mobility of the virus SSV2 is highlighted. M_1 and M_2: molecular weight markers (kb).

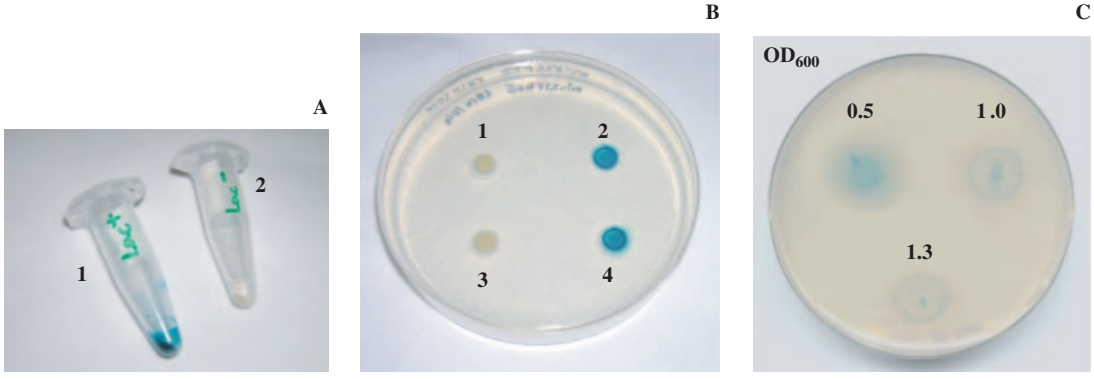

Figure 7. Complementation of the β-glycosidase mutation in *S.solfataricus* GθW and spreading of the pMSSV*lacs* vector. The expression vector pMSSV*lacS* was transferred into *S.solfataricus* GθW cells lysogens for SSV2. Successful transformation was checked after propagation for several generation by direct exposure of cell pellets from liquid cultures to X-Gal and development of the blue color [**A** (1) Lac^+, pMSSV*lacS* transformed cells; (2) Lac^-, cells transfected with pSSVx]. Maintenance of the plasmid was also confirmed for colonized areas on plates and X-Gal test [**B** (1) Strain GθW; (2) wild-type strain Gθ; (3) SSV2/ pMSSV-infected GθW; (4) SSV2/ pMSSV*lacS* -infected GθW]. **C.** 4-µl aliquots of culture supernatant of SSV2/pMSSV*lacS*-infected GθW, withdrawn at the different cell densities indicated, were spotted onto a continuous lawn of uninfected GθW cells. The spreading of the recombinant satellite particles was revealed by the blue color developed on the plaques (primary infected cells) upon X-Gal exposure. Stain intensity depended on the extent of cell growth retardation in the plaques and was progressively less intense for increasingly higher virus titers.

cells by electroporation, whereas pMSSV*lacS* was also transferred into cells of the P2/SSV2 strain. After culture propagation, extrachromosomal DNAs were prepared and analyzed by agarose electrophoresis (Figure 6). The presence and growth-dependent accumulation of both vectors in GθW/ SSV2 and of pMSSV*lacS* in P2/SSV2 revealed successful transformation and DNA replication, confirming that the plasmid size was critical for particle formation and spreading; moreover the *lacS* gene is harmless and does not induce recombination of the vector also in the P2 strain which already contains a wild-type chromosomal copy of the gene. Identical results were obtained when SSV2 lysogens of *Sulfolobus* GθW were transformed with the plasmid pSSVx*lacS* lacking the *E.coli* sequences and obtained by cleavage with *Sac*I and re-ligation (Figure 6). All pSSVx derived plasmid failed to transform *Sulfolobus* if linearized prior to transfer. In fact, X-Gal staining test was positive on *Sulfolobus* GθW/SSV2 transformed with pMSSV*lacS* (or pSSVx*lacS*, data not shown) on liquid-cultured (Figure 7A), plated (Figure 7B), and primary infected cells (Figure 7C). Plaques depicted in Figure 7C stained after incubation with the chromogenic substrate as a result of propagation of the engineered satellite virus. Since cell growth retardation in the plaques is directly proportional to the SSV2 virus titer

which is maximal in the stationary phase cultures (P. Contursi, unpublished data), turbidity and color faded progressively with increasing cell density.

Therefore, the β-galactosidase was expressed from the recombinant plasmid and the engineered satellite virus had spread efficiently throughout the culture.

pMSSV*lacS* DNA prepared from *S.solfataricus* could be re-transferred into *E.coli* without suffering recombinational rearrangements. Plasmid preparations and total DNA from *Sulfolobus* transformants were analysed by restriction analyses and Southern hybridization that confirmed the maintenance of the vector at the same levels of the parental pSSVrt. No vector integration into the host chromosome occurred since the recombinant vector exhibited the same restriction pattern as the DNA prepared from *E.coli* and no signal relative to the plasmid could be detected on genomic DNAs in the Southern blots (Figure 8). Bands corresponding to chromosomal *tf55*α promoter (in all strains tested) and *lac*S gene (absent in GθW) could also be visualized as internal controls for detection of single copy chromosomal sequences. Signal assignment of the restriction fragments of *tf55*α gene was performed on the basis of the sequence and localization on the P2 strain genome, whereas the mapping performed by Bartolucci *et al*. (32) allowed the identification of *lacS* gene specific signals. A similar Southern analysis confirmed that the pMSSV*lacS* copy number varied from 10–15 (mid-log phase) to ∼130 molecules per cell (stationary phase).

Isolation of single pMSSV*lacS* transfectants

Single colonies formed by suitably diluted mid exponential pMSSV*lacS* transformant cells (0.3 OD_{600}) on rich solid medium (M182, glucose 0.1%) resulted positive to X-Gal staining only in the fraction of $1–3 \times 10^{-3}$; below this cell density value the presence of the plasmid was undetectable in plated cells. The fraction of positive clones could be increased to 40% when stationary phase cultures were plated, namely when the plasmid had reached its maximum copy number per cell. Unfortunately, blue stained clones loosed the recombinant plasmid when suspended and directly propagated in liquid cultures as indicated by the analysis on extrachromosomal DNA and negative staining with X-Gal. This result was nevertheless expected, since also wild-type pSSVx has been demonstrated to be lost in cells of single colonies on plates and/or of progressively diluted cultures (9); indeed we confirmed and took advantage of this feature for the SSV2 lysogen selection of GθW and P2.

In order to stabilize the pMSSV*lacS* transformants, selection on minimal media containing lactose as the only nutrient source was performed (cells not complemented for *lacS* function are unable to grow because of the lack of any β-galactosidase activity) (32). Streaking (and/or seeding of suitable dilutions) revealed that the culture had 100% colony forming efficiency on lactose. Moreover, all colonies resulted positive in the X-Gal test (Figure 9A), demonstrating that they were able to retain the plasmid under selective nutrient conditions. Interestingly, cells plated as spots after resuspension and immediate re-seeding onto rich medium, maintained the β-galactosidase activity (for the analysis of 10 independent clones see Figure 9B). This procedure (resuspension and immediate re-seeding) overcame critical dilution and was successful also for the smaller fraction of positive transformants isolated from solid rich medium.

These experiments demonstrate that cultures of pMSSV*lacS* transformants before selection on plates were homogeneously infected (all cells contained the plasmid) and that in diluted cell suspensions (such as those necessary for single colony formation on plates and/or obtained by transfer of single colonies in liquid medium), actively dividing cells loosed the vector, unless a selective pressure was imposed.

β-galactosidase assays on pMSSV*lacS* transfectants

Detection of β-galactosidase activity on denaturing gels by enzyme staining revealed that the *lacS* gene product was indistinguishable when expressed as heterologous in *E.coli* and xenologous in *S.solfataricus*, namely the *Sulfolobus* expression system did not interfere with the correct polypeptide syntesis (data not shown).

β-galactosidase activity was measured using a spectrophotometric assay with the specific substrate ONPG in crude cell extracts of both transfected mutant GθW strain and wild type Gθ. The mutant recipient strain GθW has been shown to exhibit no detectable activity (32), whereas the specific activities of pMSSV*lacS* transformants in primary transformation mixtures rose from undetectable levels to 1.2 U/mg protein; this value was ∼2-fold higher than that found in wild-type cells expressing *lacS* under the control of its own promoter. The activity remained stable in diluted and propagated cells cultured as described above, when assays were performed at the same growth phase. Under identical growth conditions, the value of endogenous β-galactosidase activity in the P2 wild-type strain (i.e. the natural source of both the *lacS* gene and *tf55*α promoter sequences inserted in the pMSSV*lacS*), reached 0.1 U/mg as the highest value. Therefore, using this value as reference the expression level in GθW/pMSSV*lacS* is ∼12-fold higher. For a single culture, activity increased up to late logarithmical growth phase and then maintained approximately the same value up to late stationary phase (Figure 10). This result confirmed that the recombinant satellite viral DNA was replicated inside the cells and that virus particle formed and spread throughout the culture; the increasing activity should therefore depend only on the copy number of the plasmid and hence on the number of the *lacS* gene copies per cell.

Heat shock of the stably transfected cells, shifting the culture temperature from 75° to 88°C, induced an increase of the specific activity up to 2.5-fold (3.0 U/mg) after 3 h and remained constant in cells exposed to thermal stress for 24 h.

DISCUSSION

In this paper, we have developed a relatively small-sized and high copy number shuttle vector for *S.solfataricus* based on the satellite virus pSSVx from *S.islandicus* REY15/4.

An extended intergenic region between the still uncharacterized ORFs c68 and 288 on the pSSVx genome sequence

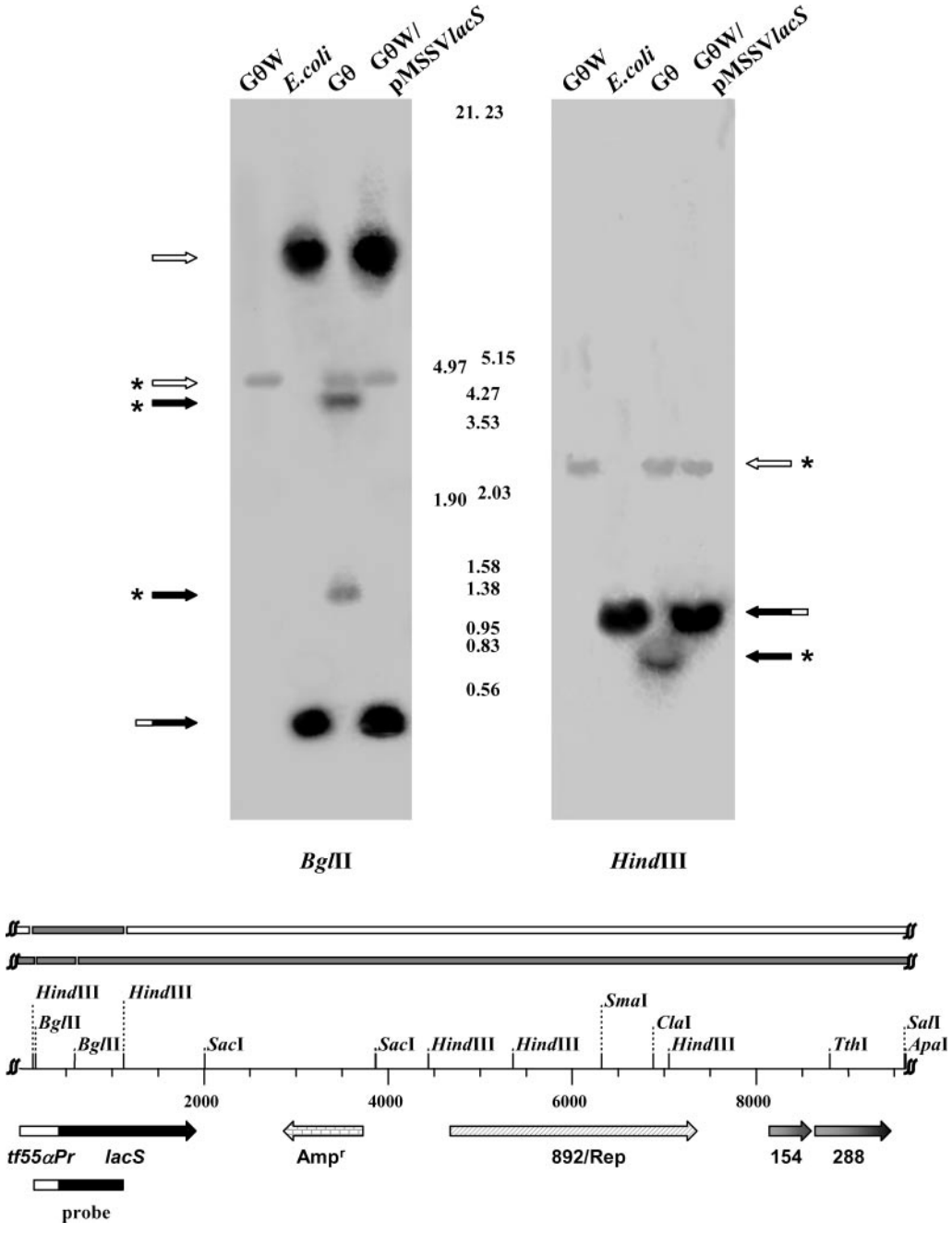

Figure 8. Southern analysis of pMSSV*lacS* transformants. For hybridization, total DNAs from transformant cells (GθW/pMSSV*lacS*) and from the recipient mutant strain GθW as well as from wild-type cells Gθ were cut with BglII and HindIII as indicated. The recombinant pMSSV*lacS* vector prepared from *E.coli* before transformation was used as a reference for correct restriction patterns (molecular weight standards are indicated). White and black arrows indicate hybridization to *tf55*α and *lacS* gene sequences, respectively, the asterisks distinguishing signals of the chromosomal copies. Signals of the *tf55*α–*lacS* gene fusion on the vector are highlighted by white/black arrows. The scheme on the bottom represents the restriction patterns for the two enzymes on the map of the linearized pMSSV*lacS* and the fragments producing positive signals for hybridization (grey bars).

appeared potentially useful for inserting foreign DNA. In fact, the vector pSSVrt clearly showed efficient replication and maintenance in SSV2 lysogens of the strains GθW and P2.

The insertion of foreign sequences into the pSSVrt vector allowed also to determine the limiting upper size (~11 kb) of the DNA to be accommodated in the virus particles. The additional sequences severely affected packaging and spreading rather than the transfer and/or the replication/maintenance of the viral DNA. These results also indicated that transport in viable virions and no other mechanism, such as conjugation (29,46–48), is responsible for cell-to-cell transfer of this genetic element.

A 'minimal' vector pMSSV was devised to accommodate inserts of a wider size range at least up to the 2.0 kb of a smaller *tf55*α/*lacS* expression cassette (35). In fact, the deletion of a region non-essential for replication and

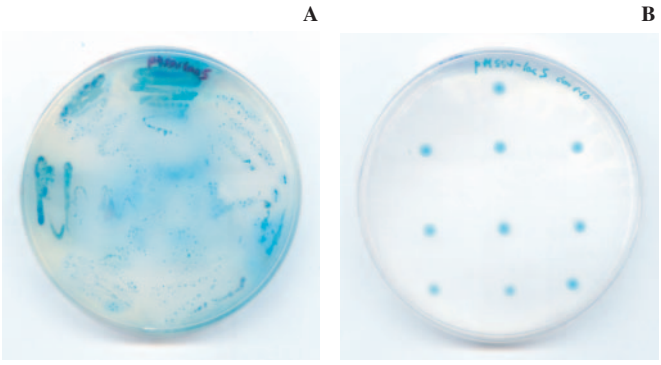

Figure 9. Selection of the pMSSV*lacs* transformants as independent clones. SSV2/GθW cells from a stationary phase culture (1.2 OD_{600}) grown in rich medium and containing pMSSV*lacS* at maximum copy number were streaked onto minimal 0.25% lactose medium to obtain single colonies. After blue color development with X-Gal at 75°C (**A**), 10 independent clones were picked, resuspended in liquid minimal medium and seeded drop wise onto rich solid medium for growth in circumscribed colonized areas. All clones maintained the plasmid as indicated by the positive X-Gal test (**B**).

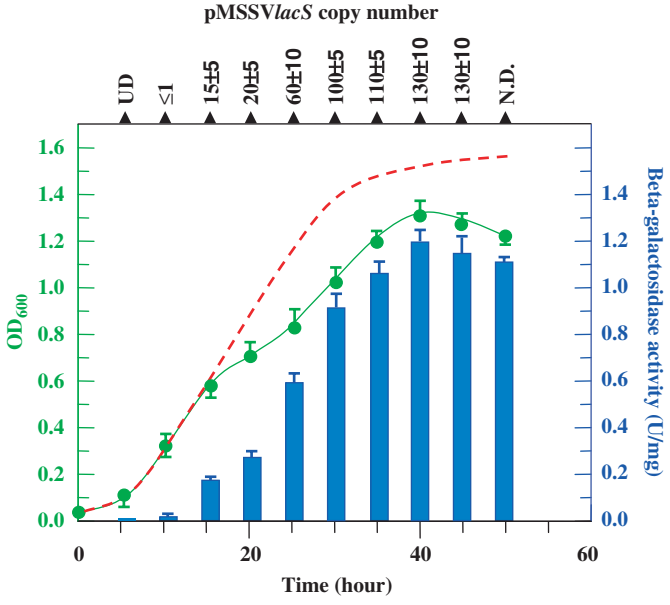

Figure 10. β-galactosidase activity and plasmid copy number in cultures of pMSSV*lacS* transformants. The β-galactosidase activity was measured in crude extracts of pMSSV*lacS* transformants harvested at different optical densities from a single culture. Specific activity (indicated by solid bars) is plotted together with the growth curve and the relative plasmid copy number. All values are the average of independent experiments performed on three different cultures and standard deviation is indicated for optical density and activity as well as for plasmid copy number. The dashed red line indicates a typical growth curve of the empty strain GθW. UD, undectectable; N.D., not determined.

selection in *E.coli* and a shortened *lacS* 3′-UTR resulted in an efficient shuttle vector able to transform *S.solfataricus* and to spread in primary cultures. The vector did not suffer either integration into the host chromosome or rearrangements and resided with undetectable to 130 copies per chromosome, the number depending on the cell growth phase. Interestingly, β-galactosidase activity in complemented mutant cells followed an increasing trend which correlated to the plasmid accumulation up to stationary phase and was not dependent on the number of the generations. *lacS* gene expression was inducible by thermal stress, namely it was confirmed to be transcriptionally regulated by the chaperonin gene promoter. However, heat shock did not produce the same β-galactosidase activity increase (∼10-fold) described by Jonuscheit *et al* (35), although maximum gene expression resulted at comparable levels in the two host/vector systems (3 versus 5 U/mg).

Maximum size for DNA insertion into pMSSV*lacS* was determined to be ∼1.3 kb (data not shown). In principle longer inserts up to 3.1 kb can be accommodated after cloning in *E.coli* and excision of the bacterial sequences prior to the transfer into *Sulfolobus*.

Maintenance of pMSSV*lacS* in early exponential grown and/or plated cells seemed to be the critical point for the efficacy of this system, because of both the low copy number (this work) and the reduced viral production in metabolically active cells (P. Contursi, unpublished data). In fact, segregation of the plasmid in an actively dividing cell (in which SSV2 is also maintained at low copy number and hence hardly sustains active replication and/or particle proliferation of pSSVx and derivatives) can be very asymmetrical until it is lost. We were able to overcome this bottleneck; as expected, efficient selection and stabilization of single pMSSV*lacS* transformants was indeed obtained by plating cells grown up to stationary phase (maximum copy number and highest viral production), by avoiding critical dilutions in the resuspension of single colonies and/or by imposing selective metabolic pressure in lactose minimal media.

This system presents many advantages compared with others already mentioned: (i) after DNA transfer, the vector propagates efficiently throughout the culture as a virus, overcoming the usual low transformation efficiency of *Sulfolobus* cells; (ii) the vector is stably maintained at high-copy number with no integration into the chromosome and hence no reduction in the number of episomal molecules and (iii) in stable transfectants, the β-galactosidase activity is dependent only on the copy number of the vector.

The availability of this new two-element transformation systems based on SSV2 and the engineered pSSVx will contribute to clarify the mechanisms responsible for the satellite/helper dependence as well as for replication, gene regulation and packaging of the episomal DNAs.

Further work will explore the use of this vector for the expression of both homologous and heterologous genes in *S.solfataricus* as well as for testing *Sulfolobus* regulatory sequences.

ACKNOWLEDGEMENTS

The authors are grateful to Prof. Quinxin She for providing us with the strain REY 15/4 of *S.islandicus*. This work was funded by the Ministero dell'Università e della Ricerca Scientifica (Decreto Direttoriale Prot. N. 1105/2002). Funding to pay the Open Access publication charges for this article was provided by 'MIUR-Decreto Direttoriale prot. n. 1105/2002' Project.

Conflict of interest statement. None declared.

REFERENCES

1. Weinbauer,M.G. and Rassoulzadegan,F. (2004) Are viruses driving microbial diversification and diversity? *Environ. Microbiol*., **6**, 1–11.
2. Forterre,P. (2001) New viruses for the new millennium. *Trends Microbiol*, **9**, 114.
3. Wommack,K.E. and Colwell,R.R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev*., **64**, 69–114.
4. Zillig,W., Prangishvilli,D., Schleper,C., Elferink,M., Albers,S., Janekovic,D. and Gotz,D. (1996) Viruses, plasmids and other genetic elements of thermofilic and hypertermofilic *Archaea*. *FEMS Microbiol. Rev*., **18**, 225–236.
5. Prangishvili,D. (2003) Evolutionary insights from studies on viruses of hyperthermophilic Archaea. *Res. Microbiol*., **154**, 289–294.
6. Snyder,J.C., Stedman,K., Rice,G., Wiedenheft,B., Spuhler,J. and Young,M.J. (2003) Viruses of hyperthermophilic Archaea. *Res. Microbiol*., **154**, 474–482.
7. Arnold,H.P., Ziese,U. and Zillig,W. (2000) SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology*, **272**, 409–416.
8. Zillig,W., Arnold,H.P., Holz,I., Prangishvili,D., Schweier,A., Stedman,K., She,Q., Phan,H., Garrett,R. and Kristjansson,J.K. (1998) Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles*, **2**, 131–140.
9. Arnold,H.P., She,Q., Phan,H., Stedman,K., Prangishvilli,D., Holz,I., Kristjansson,J.K., Garret,R. and Zillig,W. (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol. Microbiol*., **34**, 217–226.
10. Rice,G., Stedman,K., Snyder,J., Wiedenheft,B., Willits,D., Brumfield,S., McDermott,T. and Young,M.J. (2001) Viruses from extreme thermal environments. *Proc. Natl Acad. Sci. USA*, **98**, 13341–1345.
11. Rice,G., Tang,L., Stedman,K., Roberto,F., Spuhler,J., Gillitzer,E., Johnson,J.E., Douglas,T. and Young,M. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl Acad. Sci. USA*, **101**, 7716–7720.
12. Wiedenheft,B., Stedman,K., Roberto,F., Willits,D., Gleske,A.K., Zoeller,L., Snyder,J., Douglas,T. and Young,M. (2004) Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. *J. Virol*., **78**, 1954–1961.
13. Martin,A., Yeats,S., Janekovic,D., Reiter,W.D., Aicher,W. and Zillig,W. (1984) SAV1, a temperature u.v.-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius*. *EMBO J*., **3**, 2165–2168.
14. Schleper,C., Kubo,K. and Zillig,W. (1992) The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc. Natl Acad. Sci. USA*, **89**, 7645–7649.
15. Palm,P., Schleper,C., Grampp,B., Yeats,S., McWilliam,P., Reiter,W.D. and Zillig,W. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology*, **185**, 245–250.
16. She,Q., Peng,X., Zillig,W. and Garret,R.A. (2001) Gene capture in archaeal chromosomes. *Nature*, **409**, 478.
17. Reiter,W.D., Zillig,W. and Palm,P. (1988) Archaebacterial viruses. *Adv. Virus Res*., **34**, 143–188.
18. She,Q., Brugger,K. and Chen,L. (2002) Archaeal integrative genetic elements and their impact on genome evolution. *Res Microbiol*., **153**, 325–332.
19. Stedman,K., She,Q., Phan,H., Arnold,H.P., Holtz,I., Garret,R. and Zillig,W. (2003) Relationships between fuselloviruses infecting the extremely thermofilic archaeon *Sulfolobus*: SSV1 and SSV2. *Res. Microbiol*., **154**, 295–302.
20. Kletzin,A., Lieke,A., Urich,T., Charlebois,R.L. and Sensen,C.W. (1999) Molecular analysis of pDL10 from *Acidianus ambivalens* from extremely thermophilic and acidophilic Archaea. *Genetics*, **152**, 1307–1314.
21. Peng,X., Holz,I., Zillig,W., Garrett,R.A. and She,Q. (2000) Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol*., **303**, 449–454.
22. Zillig,W., Kletzin,A., Schlepr,C., Holz,I., Janekovic,D., Nain,J., Lanzerdorfer,M. and Kristjanson,J.K. (1994) Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. *Syst. Appl. Microbiol*., **16**, 609–628.
23. Cannio,R., Contursi,P., Rossi,M. and Bartolucci,S. (2001) Thermoadaptation of a mesophilic hygromicin B phosphotransferase by directed evolution in hyperthermophilic Archaea: selection of a stable genetic marker for DNA transfer into *Sulfolobus solfataricus*. *Extremophiles*, **5**, 153–159.
24. Contursi,P., Cannio,R., Prato,S., Fiorentino,G., Rossi,M. and Bartolucci,S. (2003) Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. *FEMS Microbiol. Lett*., **218**, 115–120.
25. She,Q., Singh,R.K., Confalonieri,F., Zivanovic,Y., Allard,G., Awayez,M.J., Chan-Weiher,C.C.Y., Clausen,I.G., Curtis,B.A., De Moors,A. *et al*. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl Acad. Sci. USA*, **98**, 7835–7840.
26. Pisani,F.M. and Rossi,M. (1994) Evidence that an archaeal alpha-like DNA polymerase has a modular organization of its associated catalytic activities. *J. Biol. Chem*., **269**, 7887–7892.
27. Chong,P.K. and Wright,P.C. (2005) Identification and characterization of the *Sulfolobus solfataricus* P2 proteome. *J. Proteome Res*., **4**, 1789–1798.
28. Snijders,A.P., de Koning,B. and Wright,P.C. (2005) Perturbation and interpretation of nitrogen isotope distribution patterns in proteomics. *J. Proteome Res*., **4**, 2185–2191.
29. Elferink,M., Schleper,C. and Zillig,W. (1996) Transformation of the extremely thermoacidophilic archaeon *Sulfolobus solfataricus* via a self-spreading vector. *FEMS Microbiol. Lett*., **137**, 31–35.
30. Aravalli,R.N. and Garret,R. (1997) Shuttle vectors for hyperthermophilic archaea. *Extremophiles*, **1**, 183–191.
31. Cannio,R., Contursi,P., Rossi,M. and Bartolucci,S. (1998) An autonomously replicating vector for *Sulfolobus solfataricus*. *J. Bacteriol*., **180**, 3237–3240.
32. Bartolucci,S., Rossi,M. and Cannio,R. (2003) Characterization and Functional complementation of a nonlethal deletion in the cromosome of a β-glycosidase mutant of *Sulfolobus solfataricus*. *J. Bacteriol*., **185**, 3948–3957.
33. Worthington,P., Hoang,V., Perez-Pomares,F. and Blum,P. (2003) Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus J. Bacteriol*., **185**, 482–488.
34. Albers,S.V., Jonuscheit,M., Dinkelaker,S., Urich,T., Kletzin,A., Tampe,R., Driessen,AJ. and Schleper,C. (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl. Environ. Microbiol*., **72**, 102–111.
35. Jonuscheit,M., Martusewitsch,E., Stedman,K. and Schleper,C. (2003) A reporter gene system for the hypertermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol*., **48**, 1241–1252.

36. Grogan,D.W. (1989) Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. *J. Bacteriol.*, **171**, 6710–6719.
37. Stedman,K., Schleper,C., Rumpf,E. and Zillig,W. (1999) Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics*, **152**, 1397–1405.
38. Cubellis,M.V., Rozzo,C., Monteucchi,P. and Rossi,M. (1990) Isolation and sequencing of a new beta-glycosidase-encoding archaebacterial gene. *Gene*, **94**, 89–94.
39. Sambrook,J. and Russel,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
40. Pisani,F.M., Rella,R., Raia,C.A., Rozzo,C., Nucci,R., Gambacorta,A., De Rosa,M. and Rossi,M. (1990) Thermostable beta-galactosidase from the archaebacterium *Sulfolobus solfataricus*. Purification and properties. *Eur. J. Biochem.*, **187**, 321–328.
41. Reiter,W.D., Palm,P. and Zillig,W. (1988) Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucleic Acids Res.*, **16**, 2445–2459.
42. Ermolaeva,M.D., Khalak,H.G., White,O., Smith,H.O. and Salzberg,S.L. (2000) Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.*, **301**, 27–33.
43. de Hoon,M.J.L., Makita,Y., Nakai,K. and Miyano,S. (2005) Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comput. Biol.*, **1**, e25.
44. Grogan,D.W. (1991) Evidence that beta-galactosidase of *Sulfolobus solfataricus* is only one of several activities of a thermostable beta-D-glycosidase. *Appl. Environ. Microbiol.*, **57**, 1644–1649.
45. Schleper,C., Roder,R., Singer,T. and Zillig,W. (1994) An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous β-galactosidase gene. *Mol. Gen. Genet.*, **243**, 91–96.
46. Prangishvilli,D., Albers,S.V., Holz,I., Arnold,H.P., Stedman,K., Klein,T., Singh,H., Hiort,J., Schwiver,J., Kristjansson,J.K. *et al.* (1998) Conjugation in Archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid*, **40**, 190–202.
47. She,Q., Phan,H., Garret,R., Albers,S., Stedman,K. and Zillig,W. (1998) Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. *Extremophiles*, **2**, 417–425.
48. Stedman,K., She,Q., Phan,H., Holz,H., Singh,H., Prangishvilli,D., Garret,R. and Zillig,W. (2000) pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J. Bacteriol.*, **182**, 7014–7020.