



**FACULTAT DE FARMÀCIA
UNIVERSITAT DE BARCELONA**

**CARACTERITZACIÓ GENÈTICA I FUNCIONAL DE L'OPERÓ
glc, IMPLICAT EN EL METABOLISME DE GLICOLAT I
GLIOXILAT A *Escherichia coli***

MARIA TERESA PELLICER MOYA

2. EL LOCUS *glc* D'*Escherichia coli* TÉ UNA ESTRUCTURA GÈNICA D'OPERÓ

La primera aproximació a l'estudi de l'organització gènica del locus *glc* es va fer mitjançant Northern Blots. En aquests experiments es pretenia estudiar el patró d'expressió dels diferents gens que en formen part i el tamany dels seus respectius transcrits. Posteriorment es va procedir a l'estudi de l'expressió dels gens del sistema mitjançant fusions transcripcionals al gen marcador *lacZ*.

El transcrit de *glcC* és monocistrònic. El seu tamany concorda amb l'esperat a partir d'una transcripció individual del gen corresponent. S'expressa de manera molt similar en les diferents condicions assajades. Es pot observar, però, una lleu disminució en la intensitat del transcrit quan l'RNA prové de cultius induïts en presència de glioxilat, i sobretot de glicolat. Aquest patró d'expressió és típic de gens reguladors (Gallegos *et al*, 1996). En aquest sentit, cal remarcar l'alta similitud presentada pel producte gènica de *glcC* amb proteïnes reguladores de procariotes. Recentment, en el projecte d'assignació de funcions a les proteïnes d'*Escherichia coli* per recerca de motius funcionals i/o estructurals posterior a la seqüenciació total del genoma d'*Escherichia coli* (Blattner *et al*, 1997), GlcC ha estat classificada com a membre de la família GntR de reguladors transcripcionals (Reizer *et al*, 1991).

La fusió transcripcional de *glcC*, $\Phi(glcC-lacZ)$, va confirmar la informació obtinguda dels Northern Blots. Els nivells d'expressió de la fusió són baixos en creixements en hidrolitzat de caseïna i disminueixen quan en el medi s'addiciona glicolat o glioxilat. La determinació de l'inici de transcripció de *glcC* va permetre establir les seqüències dels hexàmers -10 i -35. Les seves seqüències són, respectivament, CATAAT i TTA ACT. Són bastant diferents al consens proposat per aquests elements en *cis*. La distància entre ells és de 19 pb, que tampoc és l'òptima perquè l'RNA polimerasa els pugui reconèixer amb facilitat (Pivec *et al*, 1985). En alguns sistemes gènics, aquesta separació entre els dos hexàmers ha estat descrita com un factor que disminueix l'expressió gènica; si aquesta distància s'escurça a 16-18 pb, s'aconsegueix un increment molt gran de l'expressió basal (Hidalgo i Demple, 1997).

La baixa expressió dels gens reguladors procariotes és un mecanisme de control de l'expressió dels gens que regulen. El baix nombre de molècules de GlcC presents en la cèl.lula són suficients per estimular la transcripció dels gens estructurals del sistema *glc* en les condicions externes adients. La baixa expressió del gen regulador es correlaciona amb el baix nivell d'expressió de la fusió transcripcional, indicant la menor sensibilitat del transcrit de *glcC* als diferents mecanismes de degradació de l'mRNA. Com que *glcC* i *glcD* comparteixen la

mateixa regió intergènica promotora (en direccions oposades), és possible que durant l'activació dels gens estructurals sigui més efectiu deixar tot el promotor lliure perquè aquests s'expressin sense l'impediment que podria suposar la transcripció conjunta amb el gen regulador. Per aquest motiu l'estabilitat de l'mRNA pot ajudar a augmentar el nombre de molècules presents sense necessitat d'augmentar la transcripció.

Pel que fa a la resta de gens del sistema *glc*, els resultats de Northern Blot van aportar informació respecte al seu patró d'inducció. La transcripció dels gens estructurals està induïda per la presència en el medi de glicolat i en menor grau de glioxilat (la inducció és aproximadament la meitat). Aquesta observació implica que o bé el glicolat addicionat al medi, o el glioxilat format per acció de la glicolat oxidasa o qualsevol altre metabolit generat en aquestes condicions és capaç d'activar la transcripció dels gens estructurals.

Els tamany dels transcrits observats no coincideixen amb el tamany dels gens individuals ni amb la suma total o parcial dels transcrits de diferents gens. Possiblement la inestabilitat dels mRNAs policistrònics procariòtics és la causa de la impossibilitat de trobar un transcrit sense. El transcrit policistrònic que inclou tots els gens estructurals del sistema *glc* sembla ser *per se* molt inestable, ja que és molt llarg (aproximadament 8.5 kb) i pot presentar diversos punts de trencament.

L'estabilitat dels transcrits depèn en part de la longitud dels mateixos, tot i que s'ha observat que transcrits de tamany similar tenen estabilitats molt diferents (Ehretsmann, 1992). Aquest fet implica que el trencament dels mRNAs no es produeix a l'atzar. Existeixen llocs de trencament del mRNA on comença la degradació: aquests llocs poden ser regions situades a les zones 3' i 5' no traduïdes, a les zones intergèniques en els missatgers policistrònics i a vegades en les zones codificants (Chen i Belasco, 1990). L'alta inestabilitat del transcrit de *glcDEFGBT* podria ser un mecanisme que permetés la ràpida adaptació als canvis produïts a l'entorn.

Davant la impossibilitat de conèixer l'organització gènica mitjançant Northern Blot es van construir fusions transcripcionals de tots els gens del sistema. Aquests experiments van revelar que el sistema té una estructura gènica d'operó. Només existeix un promotor funcional situat a 5' de *glcD*, en la zona intergènica entre *glcC* i *glcD*. La localització del punt d'inici de la transcripció va permetre identificar els dos hexàmers -10 i -35, de seqüència TAATAA i TAGACG respectivament i separats per 17 pb, que és la distància òptima de separació entre aquests dos elements *in cis*. Tant les seqüències dels dos hexàmers, que estan bastant conservades respecte el consens, com la distància que els separa provoca que en condicions inductores l'expressió de *glcDEFGBT* sigui elevada.

L'existència d'aquest únic promotor implica l'expressió conjunta de glicolat oxidasa i malat sintasa G. Encara que en alguns estats metabòlics no es requereixi en principi la participació de la glicolat oxidasa, p.e. creixement en glioxilat de la soca salvatge o creixement en acetat d'una soca mutant en malat sintasa A, aquest enzim estarà present si és necessària la malat sintasa G.

3. L'EXPRESSIÓ DELS GENS ESTRUCTURALS *glcDEFGBT* ÉS TOTALMENT DEPENDENT DELS FACTORS DE TRANSCRIPCIÓ GlcC I IHF

L'estudi de l'expressió dels gens estructurals del sistema *glc* ha mostrat que aquests s'expressen exclusivament quan estan presents les proteïnes GlcC i IHF. En absència d'algun d'aquests dos factors de transcripció els nivells d'expressió s'aboleixen per complet. Aquest fet indica que per formar el complex nucleoproteic que inicia la transcripció i per establir els contactes entre aquest complex i l'RNA polimerasa és imprescindible la presència d'ambdues proteïnes reguladores.

3.1. Paper de GlcC en la transcripció de *glcDEFGBT*

La proteïna GlcC és un activador transcripcional dels gens estructurals del sistema *glc*. La seva funció s'ha pogut assignar mitjançant un gran nombre d'experiments. En primer lloc, l'expressió de $\Phi(glcD-lacZ)$ està totalment abolida en un mutant *glcC*. La nul·la transcripció d'aquests gens concorda amb l'absència de les proteïnes codificades (glicolat oxidasa i malat sintasa) determinades en els extractes cel·lulars de la soca JA154 (*glcC*) per mesura de les activitats enzimàtiques, i en el cas de la malat sintasa G, també per electroimmunodetecció. De fet, la soca JA154 és incapaç de créixer a expenses de glicolat com a única font de carboni i energia.

L'estudi de la seqüència del promotor de *glcD* no ha mostrat una clara regió palindròmica de simetria invertida ("inverted repeat"). Aquestes zones acostumen a ser llocs d'unió de proteïnes reguladores; aquest és el cas de la unió de CRP (Ebright *et al*, 1989) o d'FNR (Green i Guest, 1994) als promotors dels gens que regulen. De totes maneres, el no haver trobat aquestes seqüències en l'anàlisi *in silico* no implica que la proteïna no es pugui unir (Hénaut i Dauchin, 1996). Experiments *in vivo* han permès localitzar el possible lloc d'unió de

la proteïna reguladora al promotor. Les delecions del promotor de *glcD* des del seu extrem 5' es van dur a terme amb la finalitat de poder localitzar el lloc d'unió de GlcC.

Els primers assajos de l'activitat promotora es van realitzar amb un fragment que incloïa 303 pb per davant de l'inici de transcripció i que correspon a la fusió emprada en diferents aspectes d'aquest treball anomenada $\Phi(\textit{glcD-lacZ})$. L'eliminació de seqüències fins la posició -184 no modifica el patró d'expressió d'aquesta fusió transcripcional, assolint-se la inducció completa en presència de glicolat. Aquest fet indica que per davant de la posició -184 no existeix cap element en *cis* al qual s'uneixi cap element en *trans* necessari per la transcripció. S'observa un increment del 25 % de l'expressió en presència de glicolat de la delecio f-184 respecte la delecio f-192. L'expressió d'f-177 està disminuïda un 40 % respecte f-192. La manca absoluta d'expressió de la fusió f-170 indica la pèrdua, en aquesta construcció, d'un element en *cis* indispensable per la transcripció de l'operó *glc* i que podria correspondre a la seqüència d'unió de la proteïna activadora GlcC. A aquesta seqüència (posicions -184 a -177), a causa del seu efecte activador i de la seva localització respecte els hexàmers -10 i -35, se li ha assignat el nom d'UAS (upstream activator site).

A mesura que es va deleccionant el promotor l'expressió de la fusió corresponent és nul·la fins a arribar a la delecio f-54. En aquesta delecio, el nivell d'expressió obtingut en condicions no inductores (creixements en hidrolitzat de caseïna) és igual que el del promotor *sensu*, si bé no es veu augmentat per la presència de glicolat. Aquesta expressió és independent de la presència de GlcC i d'IHF, la qual cosa indica que l'expressió únicament depèn de la unió de l'RNA polimerasa al promotor i descarta la unió d'algun factor en *trans* a 3' de la posició -54 que influeixi l'expressió aeròbica de *glcDEFGBT*. Aquesta expressió basal dependent només del reconeixement de la subunitat σ de l'RNA polimerasa a les seqüències promotores -10 i -35 queda abolida quan s'elimina l'hexàmer -35 (fusió f-27), la qual cosa demostra la correcta assignació del *tsp* així com la seva funcionalitat en la transcripció de *glcDEFGBT*.

El fet que la delecio f-88 no mostri el mateix patró d'expressió que la delecio f-54 suggereix l'existència d'algun element que impedeix l'expressió d'f-88 i que quan és deleccionat permet l'expressió basal i no induïble d'f-54. L'anàlisi *in silico* de la regió de DNA entre les posicions -88 i -54 no detecta cap possible seqüència coneguda d'unió a factors de transcripció globals, si bé conté una regió denominada com dA+dT, de seqüència 5' -AAAAAAAAT- 3', que presenta una gran facilitat per formar zones de desaparellament de bases de DNA. És probable que aquesta regió dA+dT impedeixi l'expressió basal observada en el cas d'f-54 per formació d'una estructura secundària que dificulti l'acostament i/o unió de l'RNA polimerasa al promotor.

La unió de GlcC a la seqüència UAS es va estudiar mitjançant la tècnica de la retardació de la mobilitat electroforètica en gel. Els estudis es van dur a terme amb extractes crus de la soca salvatge MC4100 i de la soca JA154 (*glcC::cat*). Ja que s'havia localitzat la UAS entre les posicions -184 i -177, es va escollir un fragment de DNA que contingués aquesta regió. Es va usar el fragment P, que abarca les posicions -247 a -134. Els resultats obtinguts revelen diversos aspectes interessants tant de la manera com la proteïna s'uneix al DNA com del mecanisme d'inducció del sistema.

En primer lloc, cal resaltar que s'observa una banda de retardació feble en els carrils corresponents al fragment de DNA lliure. Aquest fenomen s'ha repetit en tots els assajos, fins i tot després d'haver canviat de procediment de purificació de la sonda. Com que aquest DNA no ha estat en contacte amb proteïna, s'atribueix la formació d'aquesta banda de retardació a l'estructura intrínseca d'aquesta regió de DNA. En realitat aquesta banda, anomenada banda B ("Bending", curvatura), mostra la curvatura del DNA en aquesta regió. El % de dA+dT en aquesta regió no és especialment elevat (50 %) tot i que existeixen molts dímers dA-dA que provoquen la formació d'una molècula de DNA intrínsecament curvada (Pérez-Martín *et al*, 1994). Existeix un equilibri dinàmic entre les formes curvades i no curvades. La mobilitat electroforètica d'un fragment de DNA depèn, no només del seu pes molecular (sigui DNA lliure o unit a proteïnes), sinó també de la seva curvatura. DNAs amb angles de curvatura elevats migren menys en una electroforesi que el mateix fragment linial (Lane *et al*, 1992). La banda B és minoritària respecte la forma linial, possiblement perquè les condicions de l'assaig afavoreixen aquesta topologia.

Quan s'utilitzen els extractes crus provinents de la soca MC4100 crescuda en condicions inductores (hidrolitzat de caseïna més glicolat), es pot observar l'aparició de dues noves bandes de retardació. La banda B es pot observar encara que de forma molt difosa. La banda majoritària, anomenada banda C, presenta una baixa mobilitat electroforètica, que no varia d'intensitat per l'addició de glicolat suplementari a les mesclades de reacció. Aquesta banda podria correspondre a la unió de GlcC al DNA, possiblement en forma de dímer, la qual cosa estabilitzaria la unió. Aquesta unió, facilitaria *in vivo* la formació del complex actiu que permet la transcripció del *glcD*. La banda minoritària, banda A, presenta una mobilitat electroforètica més elevada que la banda intrínseca B i podria correspondre a la unió de GlcC en forma de monòmer.

Quan s'utilitzen els extractes crus provinents de la soca MC4100 crescuda en hidrolitzat de caseïna el patró varia. S'observa la presència de la banda A, amb una intensitat major que la dels extractes de cultius induïts. Es pot apreciar una gran disminució de la banda C respecte als cultius crescuts en glicolat. La desaparició de la banda C està lligada a un descens de la taxa de transcripció, per la qual cosa es confirmaria que aquesta banda correspon

a la unió amb GlcC, que provoca una inducció transcripcional. L'addició de glicolat a les mescles no intensifica aquesta banda. Aquest fet implica que o bé el glicolat no és l'inductor o bé que les condicions de l'assaig no són les òptimes per apreciar l'efecte d'aquest compost. També s'observa la presència de la banda B, aquesta molt més intensa que en el cas de la sonda lliure i en els cultius en presència de glicolat. L'aparició d'aquesta banda pot ser causada pel plegament que la unió de GlcC en forma de monòmer comporta. Aquest efecte no és visible en el cas dels creixements en glicolat perquè la majoria de la proteïna es trobaria en forma de dímer. La dimerització de la proteïna podria ser un efecte directe o indirecte de l'addició de glicolat en el medi. Aquest és el cas del regulador XylS, que canvia de forma inactiva a forma activa per l'addició de benzoats. Aquests compostos s'uneixen a la proteïna i són els inductors del canvi de la conformació inactiva a l'activa (Ramos *et al*, 1997).

El fet que les bandes C i A siguin pràcticament indetectables en els assajos de retardació de la mobilitat electroforètica duts a terme amb extractes cel·lulars de la soca JA154 (*glcC*) corrobora l'assignació d'aquestes bandes a la unió de GlcC. Aquest patró de bandes és paral·lel a la nul·la transcripció de *glcDEFGHT* en el mutant JA154. El resultat observat és similar independentment que els extractes cel·lulars fossin obtinguts a partir de creixements no induïts de la soca JA154, o bé induïts per la presència de glicolat en el medi de cultiu, la qual cosa indica que sigui quina sigui la molècula efectora, la unió de GlcC és imprescindible per induir la transcripció dels gens estructurals del sistema *glc*.

D'aquests experiments *in vitro* no es pot saber quina és la molècula efectora de la inducció dels gens estructurals del sistema *glc*, encara que sí s'observa que la formació del complexe estable només s'aconsegueix amb els extractes crus obtinguts en creixements en presència de glicolat.

Existeix una certa probabilitat que la molècula inductora sigui el glicolat. S'ha intentat purificar la proteïna GlcC mitjançant el sistema de l'addició d'una cua de sis histidines a l'extrem N-terminal. La proteïna va ser purificada en condicions denaturants i després de la renaturalització es van dur a terme assajos de retardació de la mobilitat electroforètica amb el fragment P. La proteïna no era gaire activa, bé pel tipus de condicions de purificació massa dràstiques, bé perquè l'addició de les histidines a l'extrem N-terminal (on es troba el motiu hèlix-volta-hèlix) interferia en la seva capacitat d'unió al DNA. Es requerien concentracions molt elevades de proteïna purificada (de l'ordre de micrograms) per aconseguir una retardació apreciable, si bé cal destacar que la capacitat de retardació del fragment de DNA es veia incrementada per l'addició de glicolat a les mescles de reacció. D'aquestes observacions no es pot concloure, però, que el glicolat sigui la molècula efectora, ja que la gran quantitat de proteïna afegida a les mescles de reacció podria formar complexos no del tot específics amb el fragment de DNA.

Un cop analitzada la unió de GlcC al seu promotor es va pretendre estudiar l'efecte dosi de la proteïna sobre l'expressió dels gens estructurals. D'una banda es va estudiar l'efecte del segrestament de la proteïna activadora i de l'altra la sobreexpressió de la mateixa.

Els primers experiments mostren que el segrestament efectiu de la proteïna GlcC només es produeix quan estan presents les regions del promotor contingudes entre les posicions -177 fins a +94. El fragment P, que conté les posicions -247 a -134 no segresta el factor de transcripció, tot i que conté la seqüència UAS. Aquest fet semblaria indicar que GlcC podria unir-se també en alguna zona més propera al *tsp*, i per tant l'efecte de segrestament total no seria visible a no ser que es disposi del promotor *senser*. Recollint la informació obtinguda dels assajos d'expressió en les diferents deleccions dels promotor això no sembla ser així, a no ser que l'efecte de la regió UAS fós *in vivo* molt més important que el d'aquesta segona caixa UAS. Els resultats del segrestament obren la porta a l'existència d'algun factor addicional que també sigui segrestable pel promotor *senser* però no per algun dels seus fragments. Aquest factor seria el responsable de la unió de GlcC al promotor *senser* (no al promotor parcial, que per assajos *in vitro* de retardació en gel ja s'ha demostrat que s'uneix). La topologia intrínscament curvada del promotor podria requerir d'un factor de modificació estructural del DNA que permetés que GlcC s'unís al DNA.

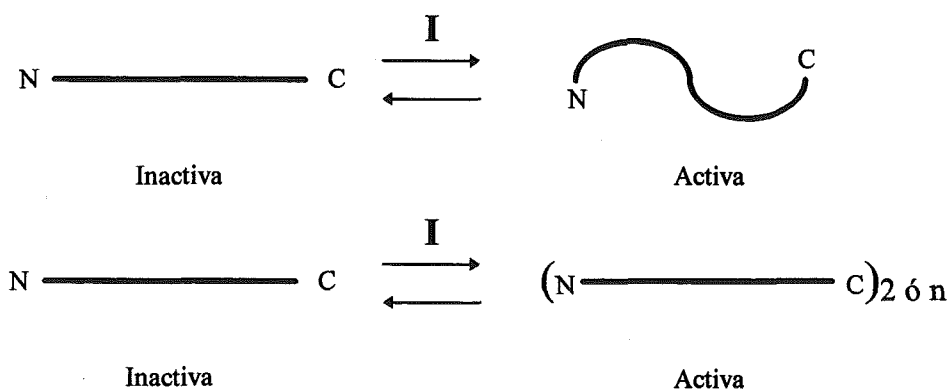


Figura 2. Desplaçament de l'equilibri entre la forma activa i la inactiva de proteïnes reguladores per efecte de l'inductor. N: extrem aminoterminal; C: extrem carboxiterminal; I: inductor o molècula efectora. En la part superior es mostra l'acció de l'inductor mitjançant un canvi conformacional. En la part inferior es mostra l'efecte d'oligomerització per acció de la molècula inductora.

L'efecte de l'increment en la quantitat de GlcC present en el citosol és clarament un efecte d'activació. L'activació de la transcripció de l'operó *glcDEFGBT* en la soca MC4100 transformada amb un plasmidi que sobreexpressa *glcC* és independent de la presència de

l'inductor en el medi de cultiu: l'activació s'observa en cultius d'hidrolitzat de caseïna. Aquesta observació demostra d'una banda que GlcC és una proteïna reguladora amb funció activadora. D'altra banda, també es pot deduir d'aquests resultats que GlcC existeix en el citosol en dues formes: una activa i una inactiva. Existeix un equilibri entre les dues formes, que és desplaçat cap a la forma activa en presència de l'inductor (Figura 2). L'inductor pot ser el causant de l'activació al provocar un canvi estructural de la proteïna o bé la dimerització del factor de transcripció, que ara podrà unir-se fàcilment al promotor i activar els gens estructurals (Ramos *et al*, 1997) (Figura 2).

La sobreproducció de proteïna reguladora a causa de la introducció del gen *glcC* en forma de vector multicòpia provoca un desplaçament cap a la forma activa de l'equilibri forma inactiva/forma activa, i per tant l'acció, que en condicions fisiològiques és duta a terme per l'inductor, queda superada pel gran nombre de còpies de GlcC en el citoplasma (Ramos *et al*, 1997) (Figura 2).

3.2. Paper de GlcC en la transcripció de *glcC*

L'estudi del paper de la proteïna GlcC sobre la transcripció del gen que la codifica va mostrar que aquesta és un repressor de la seva pròpia transcripció. Tres resultats corroboren aquest rol de la proteïna GlcC: i) les condicions que provoquen un inducció dels gens estructurals són repressores per l'expressió de *glcC*, com ho demostra el fet que $\Phi(\textit{glcC-lacZ})$ està reprimida en creixements en hidrolitzat de caseïna més glicolat, quan els gens estructurals estan en la seva màxima expressió, mentre que en creixements en hidrolitzat de caseïna l'expressió de $\Phi(\textit{glcC-lacZ})$ està augmentada; ii) en mutants *glcC* existeix un augment dels nivells de transcripció de $\Phi(\textit{glcC-lacZ})$ i iii) la introducció d'un alt nombre de còpies de GlcC en la soca MC4100 $\Phi(\textit{glcC-lacZ})$ provoca una disminució dels nivells d'activitat β -galactosidasa.

Aquestes observacions indiquen que quan s'estan transcrivint els gens *glcDEFGBT* la transcripció de *glcC* està disminuïda. En aquestes condicions no sembla necessari sintetitzar més molècules de GlcC perquè l'activació dels gens estructurals ja s'està produïnt. D'aquesta manera s'aconsegueix limitar l'efecte de l'activació dels gens estructurals a la presència de l'inductor. La conseqüència de l'autoregulació de GlcC és que la seva concentració intracel·lular es manté a un nivell baix bastant constant, i aquest fet pot ser essencial per controlar l'especificitat de l'activació dels seus gens diana (Jovanovic i Model, 1997).

D'aquesta manera, GlcC es pot descriure com una proteïna amb funció dual d'activador i de repressor simultani.

Aquesta dualitat ha estat descrita amb anterioritat per diferents factors de transcripció procariotes (Bertoni *et al.*, 1997). La causa d'aquest efecte dual és el solapament entre els promotors de *glcDEFGBT* i *glcC*. L'arranjament de gens reguladors transcrits divergentment dels operons als quals regulen és usual en l'expressió gènica procariota i també en la de bacteriòfags i de genomes nuclears, mitocondries, cloroplasts i virus eucariotes (Beck i Warren, 1988). Un exemple d'aquesta disposició és l'operó *tdc* d'*Escherichia coli*, que codifica els enzims de la deshidratació dependent de piridoxal-fosfat de l'L-treonina i l'L-serina (Wu i Datta, 1992). En el sistema *glc*, la zona UAS, definida com el lloc d'unió de GlcC al promotor dels gens estructurals, està localitzada a la posició +44 del promotor de *glcC*. Aquesta situació pot provocar un impediment estèric perquè l'RNA polimerasa transcriuigui correctament i sense entrebancs *glcC* mentre s'estan transcrivint *glcDEFGBT*. Per tant, la zona UAS pels gens estructurals pot ser considerada com una seqüència inhibidora ó operador del promotor de *glcC*.

3.3. Paper d'IHF en la transcripció de *glcDEFGBT*

El factor de transcripció IHF ha demostrat ser imprescindible per l'activació transcripcional dels gens estructurals del sistema *glc*. Mutants independents en les dues subunitats d'IHF no expressen $\Phi(glcD-lacZ)$ i són incapaçs de créixer en glicolat com a única font de carboni i energia. Tant les mutacions en el gen *himA* com en *himD* reprimeixen per complet la transcripció de *glcDEFGBT*. Això vol dir que els homodímers d'IHF que es formen no poden activar la transcripció, tot i que aquest fenomen ha estat descrit per algun sistema (Jovanovic i Model, 1997). L'efecte de la manca d'aquest factor de transcripció pleiotròpic s'exerceix directament sobre el promotor dels gens estructurals i no sobre sobre l'expressió del gen regulador. L'expressió de $\Phi(glcC-lacZ)$ no es veu afectada per la presència o absència d'IHF. Aquest també és el cas de l'operó *tdc* d'*Escherichia coli* (Wu i Datta, 1992). L'efecte dràstic que mutacions en IHF tenen sobre l'expressió de $\Phi(glcD-lacZ)$ indica que la transcripció dels gens estructurals del sistema *glc* està estrictament regulada.

Tenint en compte l'organització dels dos promotors solapats els resultats són esperables. IHF té un paper en la correcta disposició estructural de tots els factors que permeten la formació del complex transcripcionalment actiu. El promotor de *glcC*, en el qual

no s'ha identificat cap proteïna que reguli la seva expressió, no necessita arranjar la seva topologia i per activar-se només necessita el reconeixement per part de l'RNA polimerasa.

Per contra, en el promotor de *glcD* els elements en *cis* localitzats i implicats en l'activació de la transcripció estan allunyats els uns dels altres. En l'anàlisi *in silico* s'han trobat dues seqüències d'unió al dímer IHF (Goodrich *et al*, 1990). Aquestes seqüències estan localitzades en el promotor de *glcD* entre la zona UAS i la zona dA+dT. Aquesta localització entre la zona on s'uneix la proteïna activadora i la zona d'unió de l'RNA polimerasa ha estat descrita per diversos promotors procariotes: p.e. en el promotor de l'operó *narGHJI* que codifica la nitrat reductasa (Schröder *et al*, 1993) i en el promotor de l'operó *psp* de la proteïna de shock fàgic (Jovanovic i Model, 1997). En aquests exemples, igual que en el cas de *glcDEFGBT*, la funció d'IHF és la d'actuar com a activador transcripcional. Aquesta funció és conseqüència de la localització del consens per IHF en els promotors. En els casos en els quals IHF actua com a repressor de la transcripció, les zones d'unió d'aquest factor al promotor es localitzen solapades amb els hexàmers -35 i -10, impedit d'aquesta manera la unió de l'RNA polimerasa al promotor o bé sobre les zones UAS, impedit la unió dels activadors (Goosen i van de Putte, 1995).

A part de la localització de les possibles zones d'unió a IHF, el promotor de *glcD* conté unes característiques que el fan molt afí a IHF. Està descrit que la unió d'IHF als promotors que controla està afavorida si aquests tenen un alt contingut en dA+dT, ja que són promotors amb una més gran facilitat per curvar-se. En el cas del promotor de *glcD*, amb un % de dA+dT del 56 %, aquest efecte es veu augmentat pel fet que existeixen fragments flanquejant la possible seqüència d'unió a IHF en els quals existeix un alta concentració d'aquestes bases nitrogenades que s'uneixen entre elles per només dos ponts d'hidrogen. Trobem segments on existeixen 9 pb dA+dT (posicions -69 a -61, que és la zona anomenada dA+dT) i altres on en trobem 7 (posicions de -110 a -104 i de -168 a -162).

En la majoria de sistemes estudiats, l'absència d'IHF provoca canvis de dos a cinc cops en la taxa de transcripció. Aquests efectes es consideren modestos (Nash, 1996). Per contra, les mutacions en IHF provoquen una reducció total de l'expressió de l'operó *glc*. Han estat descrits pocs exemples en els quals l'efecte de mutacions en IHF sigui tant dràstic. Entre ells trobem la inducció de l'operó *nifHDK* pel regulador *nifA*, que codifica les subunitats de la dinitrogenasa i dinitrogenasa reductasa (Molina-López *et al*, 1994); la inducció dels gens de la nitrat reductasa pel regulador NarL (Schröder *et al*, 1993) i la inducció del gen que codifica la treonina deshidratasa per part del regulador *tdcR* (Wu i Datta, 1992).

L'efecte d'IHF sobre la transcripció podria ser conseqüència de la fase de creixement cel·lular, ja que està descrit que la quantitat d'IHF es duplica en fase estacionària respecte a la fase exponencial de creixement (Ditto *et al*, 1994). Aquest no sembla ser l'efecte exercit sobre

l'operó *glc*, ja que les determinacions de les activitats β -galactosidasa s'han dut a terme tant en els mutants com en la soca salvatge en els mateixos límits de la fase exponencial. És possible que l'expressió de l'operó *glc* estigui augmentada en la fase estacionària de creixement (en la qual existeix una gran degradació de purines); l'increment de la taxa de transcripció que hauria de succeir en aquestes condicions podria venir modulad per l'increment en la producció de molècules d'IHF (Ditto *et al*, 1994).

El mecanisme a través del qual IHF exerceix el seu efecte sobre la transcripció de *glcDEFGBT* es pot deduir dels assajos de retardació de la mobilitat electroforètica efectuats amb extractes crus de la soca salvatge MC4100 i dels mutants IHF. El fragment usat per dur a terme aquests assajos de retardació va ser el fragment G. Aquest engloba tot el promotor actiu de *glcDEFGBT* (posicions -247 a +94). Amb els extractes crus de la soca MC4100 crescuda en hidrolitzat de caseïna es pot observar l'aparició de dues bandes de retardació: la banda Y i la banda X. La banda X, la de menor mobilitat electroforètica, correspondria a tot el complex activador de la transcripció, que inclou IHF i la proteïna reguladora GlcC, possiblement en forma de dímer, en la conformació activa que permet els contactes físics entre el regulador i l'RNA polimerasa. La mobilitat electroforètica de la banda X és possiblement conseqüència no només de l'efecte de l'augment de pes molecular provocat per la unió de GlcC i IHF sinó també del canvi en la topologia del DNA. La banda Y, de major mobilitat electroforètica, correspondria a la unió de GlcC al DNA. Aquesta unió semblaria ser poc estable ja que els extractes crus obtinguts en aquestes condicions no inductores tindrien poques molècules de GlcC en forma activa (forma dimèrica).

Amb els extractes crus de la soca MC4100 crescuda en presència de glicolat també es pot observar l'aparició de les dues bandes de retardació. La banda X apareix més intensa, indicant que la formació del complex transcripcionalment actiu està afavorida en aquestes condicions, que corresponen a la màxima inducció *in vivo*. La banda Y també està més intensa i definida, ja que en aquestes condicions existeixen molts dímers de GlcC al citosol per l'acció de l'inductor (glicolat o algun metabolit derivat).

Amb els extractes crus de la soca JA162 (mutant IHF) crescuda en hidrolitzat de caseïna no s'observa cap banda de retardació. Tot i existir GlcC, aquest és incapaç en aquestes condicions d'unir-se al promotor. Possiblement les poques molècules dimeritzades que existeixen no són suficients per unir-se al promotor en absència d'IHF. Aquest resultat sembla indicar que IHF ajuda d'alguna manera a la unió de GlcC al seu promotor. L'absència de bandes de retardació és congruent amb la manca d'expressió de l'operó *glc* en mutants d'IHF. Quan s'usen extractes del mutant JA162 obtinguts a partir de cultius induïts per creixement en presència de glicolat només s'observa la banda Y. Això indicaria que quan el nombre de dímers o de GlcC activa és suficientment nombrós, algunes molècules poden unir-se al promotor,

sense que això comporti una activació de la transcripció, com es veu *in vivo*. En aquestes condicions no es forma la banda X, assignada al complex actiu.

De l'anàlisi de les retardacions en gel amb mutants IHF es desprèn que el paper d'IHF en l'expressió de l'operó *glc* pot estar relacionat, no només amb la curvatura del DNA, sinó també amb la facilitació de la unió de GlcC al promotor de *glcD*. El fet que quan IHF és absent la unió de l'activador es vegi severament disminuïda suporta aquesta hipòtesi. Sembla doncs que la unió d'IHF, en certa manera, provoca un fenomen cooperatiu respecte la unió de GlcC (Figura 3). El fenomen de la cooperativitat ha estat descrit entre PspF i IHF en el promotor *psp* a *Escherichia coli* (Jovanovic i Model, 1997). En aquest exemple, molt semblant al sistema *glc* en quant a organització gènica, el gen regulador *pspF* és regulat negativament pel seu producte gènic i els gens estructurals *pspA-E* estan regulats positivament per PspF gràcies a la presència d'IHF. En aquest sistema afavoreix sis cops la unió del regulador a les seves UASs i el contacte entre la proteïna unida i l'RNA polimerasa. En aquest cas la cooperativitat és recíproca i la unió de PspF a la seves UASs ajuda a la unió d'IHF.

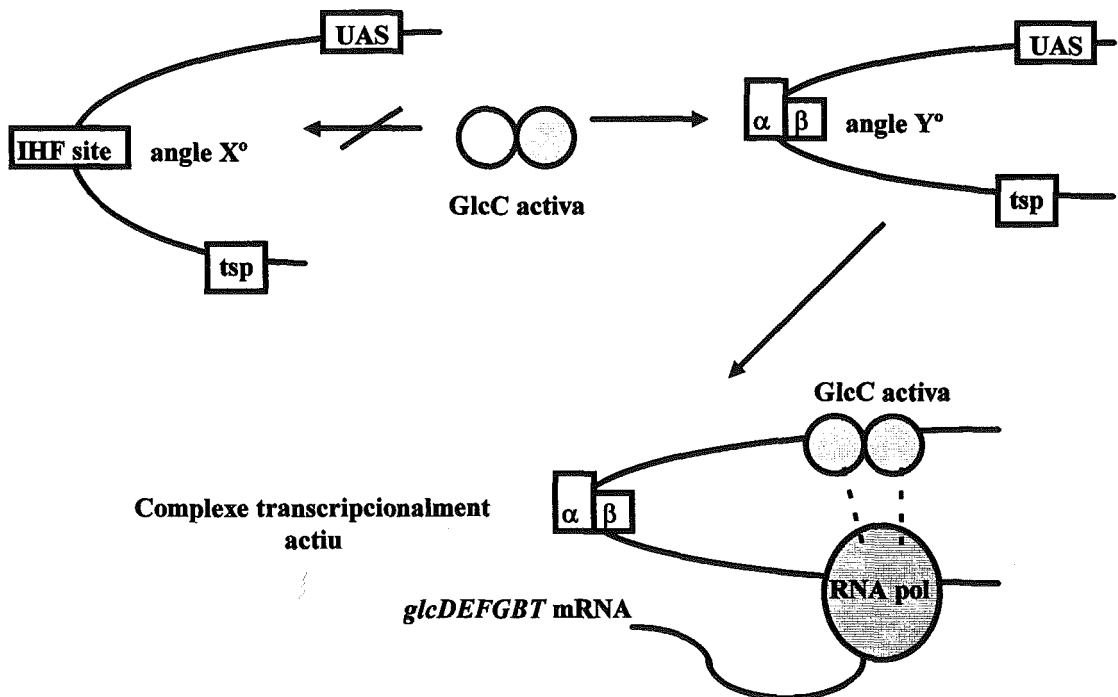


Figura 3. Model proposat per la formació del complex transcripcionalment actiu per l'expressió dels gens estructurals *glcDEFGHT*. La formació del complex transcripcionalment actiu per l'expressió de *glcDEFGHT* requereix la unió prèvia d'IHF (simbolitzat amb dos rectangles, α i β) al DNA intrínscament curvat (amb un angle X°); la proteïna GlcC activa (en forma de dímer, indicada amb dos cercles) pot unir-se al promotor que té un angle diferent (Y°) per acció d'IHF i formar aleshores el complex que permet iniciar la transcripció. Les interaccions entre GlcC activa unida al promotor i l'RNA polimerasa es simbolitzen amb dues línies discontinües. UAS: upstream activator site; IHF site: lloc d'unió d'IHF; RNAPol: RNA polimerasa.

Un model semblant al presentat en la Figura 3 pot ser el que regeixi la interacció entre IHF i GlcC en el promotor de *glcDEFGBT*. Per unir-se al promotor, GlcC necessita trobar el DNA amb una topologia específica, que és la que li confereix IHF. Com que el promotor és intrínscament curvat, en cas que existeixin molts dímers de GlcC (GlcC activa), una petita part del dímer s'unirà al promotor tot i no existir la proteïna modificadora de la topologia del DNA. La unió de la forma activa de GlcC al promotor no es tradueix en una activació de la transcripció ja que IHF té un paper addicional en el procés d'inici de la transcripció. La distància existent entre la UAS i el lloc d'unió de l'RNA polimerasa requereix que el DNA es curvi considerablement per tal de facilitar els contactes proteïna-proteïna. Aquesta conformació la pot dur a terme IHF, que és capaç de distorsionar l'hèlix de DNA fins a provocar angles superiors a 140°. En el cas del sistema *glc*, l'efecte de GlcC sobre la unió d'IHF al promotor no ha estat demostrat.

El requeriment de l'acció d'IHF per la unió de GlcC al promotor de *glcDEFGBT* constitueix un mecanisme que pot conferir gran especificitat al procés d'inici de la transcripció. En aquest cas existeix un mecanisme de control que provoca que el reconeixement de la zona UAS per part de GlcC (reconeixement de la seqüència de nucleòtids) no sigui suficient perquè s'iniciï la transcripció; la presència d'IHF és imprescindible perquè es pugui produir la unió entre activador i promotor. S'ha descrit que la participació d'IHF en processos d'unió de proteïnes reguladores als seus promotors és un mecanisme que permet suprimir l'activació transcripcional produïda per reguladors heteròlegs (Pérez-Martín i de Lorenzo, 1995). Així en el sistema *glc*, la dependència total d'IHF afavoriria la resposta del regulador GlcC que té una posició òptima dins el promotor i desafavoriria l'acció d'altres activadors heteròlegs al sistema.

4. LA REPRESSIÓ ANAERÒBICA DE L'OPERO *glcDEFGBT* ESTÀ MEDIADA PEL SISTEMA DE DOS COMPONENTS ArcB/ArcA

L'expressió dels gens estructurals del sistema *glc* està reprimida en condicions anaeròbiques. L'efecte és directe sobre els gens estructurals, ja que la taxa de transcripció del gen regulador *glcC* no es veu modificada per la disponibilitat d'oxigen. Aquesta repressió s'exerceix a nivell transcripcional per acció del factor de transcripció ArcA, pertanyent al sistema de transducció de senyals de dos components ArcB/ArcA. La repressió mediada per aquest sistema global de control de la transcripció té un efecte pleiotròpic sobre l'acció de la proteïna activadora específica del sistema (GlcC). Aquest fenomen s'observa també en altres

sistemes. Així, l'activació de *sodA* (codifica la superòxid dismutasa dependent de manganès d'*Escherichia coli*) requereix la participació de l'activador SoxS. SoxS es sintetitza en condicions anaeròbiques, però *sodA* està reprimida, ja que l'efecte repressor d'ArcA és superior a l'exercit per l'activador específic (Compan i Touati, 1993).

L'expressió anaeròbica és de l'ordre del 3.5 % dels nivells aeròbics, tant a nivell de l'expressió no induïda (creixements en hidrolitzat de caseïna) com induïda (creixements en hidrolitzat de caseïna més glicolat). La repressió és superable per addició de nitrat en el medi. La presència d'un acceptor final d'electrons permet la respiració anaeròbica cel·lular que desreprimeix la transcripció; el component sensor ArcB en presència de nitrat deixa de rebre el senyal d'anaerobiosi i per tant no s'autofosforila. Això provoca que ArcA no sigui fosforilada i que per tant no actui com a repressor de la transcripció. L'addició de piruvat als cultius provoca una disminució de la desrepressió exercida pel nitrat. Està descrit que el piruvat indueix la fosforilació d'ArcB (Iuchi, 1993), i per tant indirectament d'ArcA, que en forma d'ArcA-P regula la transcripció.

La confirmació que un sistema està sota el control d'ArcB/ArcA s'aconsegueix analitzant l'expressió del sistema en un fons genètic mutant en ArcA. Els mutants ArcB no s'utilitzen amb aquesta finalitat, ja que aquesta proteïna de membrana no és l'única molècula capaç de fosforilar ArcA (encara que és la majoritària en condicions de creixement anaeròbic). ArcA pot ser fosforilada també per acetil-fosfat. Per tant l'efecte d'una mutació en ArcA és més eficaç per esbrinar si un operó forma part del moduló Arc.

L'assaig de l'activitat β -galactosidasa en el mutant JA164 Φ (*glcD-lacZ*) (*arcA2*) mostra la desaparició de la repressió anaeròbica observada en la soca salvatge MC4100 Φ (*glcD-lacZ*), indicant la participació directa d'aquest factor de transcripció en la repressió de l'expressió de la fusió transcripcional en cèl·lules creixudes anaeròbicament en absència de cap acceptor final d'electrons alternatiu. En el fons genètic de la soca JA164 (mutant *arcA*) l'expressió aeròbica de Φ (*glcD-lacZ*) està elevada respecte la soca salvatge. En condicions aeròbiques existeix una mínima taxa de transcripció d'*arcA*; la proteïna derivada no pot ser fosforilada per ArcB, que en condicions d'aerobiosi està inactiva, però sí per una altra molècula, per exemple l'acetil-fosfat. Aquesta petita quantitat d'ArcA-P en cèl·lules creixent aeròbicament és suficient per explicar la disminució del 14 % observada en els nivells d'expressió de *glcDEFGHT* en la soca salvatge MC4100 respecte la soca JA164.

La repressió exercida pel sistema Arc en la transcripció dels gens estructurals de l'operó *glc* està mediada per la unió de la proteïna ArcA al promotor. Experiments de retardació de la mobilitat electroforètica en gel mostren que la manera d'unir-se a aquest promotor és la mateixa que ha estat descrita per altres sistemes pertanyents al moduló Arc, com són *lctPRD* o *sdhCDAB* (Lynch i Lin, 1996a). ArcA s'uneix específicament al promotor

en forma fosforilada. La fosforilació incrementa l'especificitat i l'afinitat d'ArcA pel promotor de *glcD*. S'observen dos complexos de retardació, complexe I i II, segons la quantitat de proteïna. Existeix un límit en la concentració de proteïna a partir del qual no es formen complexos de menor mobilitat electroforètica. La formació d'aquests complexos sembla indicar que o bé existeix més d'un lloc d'unió de la proteïna al llarg del DNA usat en els experiments (que inclou les seqüències -247 a +94) o bé la proteïna és capaç d'oligomeritzar-se i per tant de formar complexos de més alt pes molecular i de menor mobilitat electroforètica.

La primera hipòtesi va ser confirmada en els assajos de protecció a la DNasa I. Es pot observar com l'addició d'ArcA no protegeix cap regió del promotor, mentre que ArcA-P protegeix dues regions. Aquestes regions, anomenades ArcA site I (posicions -45 a +31) i ArcA site II (posicions -106 a -161), contenen diferents seqüències amb alta similitud al consens proposat per Lynch i Lin (1996a) i confirmat en aquest estudi (totes tenen 7/10 posicions conservades). Aquestes seqüències es localitzen a les posicions -7 a -16 (a ArcA site I, cadena no codificant) i a les posicions -102 a -111 i -110 a -119 (a ArcA site II, cadena no codificant i codificant respectivament). El fet que la regió protegida per la DNasa I sigui molt extensa per una proteïna que té un Pm de 28000 suggereix la unió de la mateixa en forma d'oligòmer. Aquest tipus de protecció extensa ja ha estat descrit per altres sistemes (Shen i Gunsalus, 1997). Al llarg de les regions protegides per ArcA-P s'han localitzat diverses posicions hipersensibles a la DNasa I, la qual cosa suggereix que la unió d'ArcA-P promou un canvi topològic al DNA (Shen i Gunsalus, 1997).

La contribució individual de cadascuna de les tres seqüències a la regulació anaeròbica de *glc* no ha estat establerta, ja que tant els assajos *in vivo* com *in vitro* s'han realitzat amb el promotor sense (sense deletar cap de les tres seqüències). Tot i això la localització de les mateixes respecte als altres elements reguladors del promotor permeten pensar en el paper que cadascuna d'elles exerceix.

La seqüència localitzada en ArcA site I està situada sobre l'hexàmer -10 del promotor, igual que succeeix en el cas d'*aldA*. Possiblement l'efecte repressor d'ArcA en interaccionar amb aquesta posició sigui el d'impedir que l'RNA polimerasa s'hi unexi i per tant d'aquesta manera s'evita l'inici de la transcripció. L'hexàmer -35 i l'inici de transcripció també es troben en la zona més fortament protegida per ArcA-P en ArcA site I.

Respecte a les seqüències localitzades a ArcA site II, aquestes es solapen amb la zona d'unió d'IHF. La total dependència de l'expressió del promotor de *glcD* a la unió d'IHF (demostrada en condicions aeròbiques) podria ser el mecanisme pel qual la unió d'ArcA-P a aquesta regió provoqui la repressió impeding la unió del factor de transcripció IHF. Aquest efecte repressor es veuria amplificat pel fet que sense la unió prèvia d'IHF la unió de

l'activador específic GlcC està seriosament impedida. Estan descrits diversos sistemes anaeròbics que tenen una activació dependent d'IHF (p.e. l'operó *tdcABC* (Wu i Datta, 1992) i l'operó *narGHJI* (Schröder *et al*, 1993)), per tant la quantitat d'IHF present en condicions d'anaerobiosi no ha de ser el factor limitant per la transcripció de *glcD*. La curvatura que ArcA-P sembla exercir sobre el promotor de *glcD* podria tenir una orientació diferent a la promoguda per IHF o si més no tenir un angle diferent a l'induït per IHF. Per tant, la unió d'ArcA-P a la zona II impediria la unió d'IHF i distorsionaria el DNA de manera que aquest tingués una curvatura incompatible amb l'activació transcripcional. Segons aquest model, la repressió anaeròbica del sistema *glc* estaria assegurada per l'impediment de la unió de l'RNA polimerasa i d'IHF.

La repressió anaeròbica de l'operó *glc* és congruent amb el metabolisme del glicolat. Encara que es dongués la transcripció dels gens estructurals i la traducció de les proteïnes codificades, l'enzim glicolat oxidasa no seria funcional. El seu mecanisme catalític implica la cessió dels electrons a la cadena de transport electrònic. Com que el transport d'electrons està aturat en condicions d'anaerobiosi, és coherent que un enzim eminentment aeròbic no es sintetitzi sota una atmosfera anòxica. Quan en el medi s'afegeix un acceptor final d'electrons, la repressió desapareix. Aquest efecte únicament s'ha observat amb nitrat com a acceptor alternatiu a l'oxigen. Ni el fumarat ni el DMSO restauren els nivells aeròbics. Possiblement la transferència dels electrons despresos en l'oxidació del glicolat a glioxilat requereix un acceptor d'elevat potencial redox, com és el del nitrat, i no es produiria quan en el medi estan presents altres acceptors de menor potencial (fumarat o DMSO).

5. LA MUTAGÈNESI DIRIGIDA DE LA SEQÜÈNCIA D'UNIÓ D'ArcA AL PROMOTOR d'*aldA* CONFIRMA LA FUNCIONALITAT DE LA SEQÜÈNCIA CONSENS PROPOSADA PER LYNCH I LIN

La funcionalitat de la seqüència d'unió d'ArcA al promotor de *glcD* es va estudiar en el promotor del gen *aldA*. Aquest gen mostra una seqüència 100 % similar al consens d'unió d'ArcA proposat (Lynch i Lin, 1996a) i una gran repressió anaeròbica. Ha estat descrit que *aldA* està reprimit per l'acció del sistema de dos components ArcB/ArcA (Limon *et al*, 1997). Aquests autors van localitzar la seqüència d'unió d'ArcA al promotor d'*aldA*, entre les posicions -13 i -4 de la cadena codificant, en base a la seva similitut amb el consens proposat per Lynch i Lin (1996a). En la cadena no codificant també es troba una seqüència que té 10/10

posicions conservades respecte al consens (posicions -2 a -11). Aquestes dues seqüències, que són 100 % similars al consens, han estat mutades dirigidament per tal de poder establir a nivell general si la seqüència proposada és en realitat la zona on s'uneix el regulador transcripcional ArcA i si és funcional *in vivo*.

Els assajos de retardació de la mobilitat electroforètica efectuats amb el promotor salvatge mostren que ArcA s'uneix específicament al promotor sempre i quan es trobi en forma fosforilada. La dependència de la fosforilació per la unió de la proteïna ha estat descrita pels promotors d'*lctPRD*, *gltA*, *sdhCDAB* i *cydAB* (Lynch i Lin, 1996a).

Els assajos de protecció a la DNasa I confirmen la localització de les dues caixes 10/10 en la zona anomenada ArcA site I. A més a més s'ha localitzat una segona regió anomenada ArcA site II, que està protegida davant la digestió de la DNasa I. En aquesta segona caixa existeix una seqüència 7/10 a la cadena codificant entre les posicions -138 i -129. Els "footprints" van ser realitzats amb el promotor sense, que conté els dos centres ArcA site I i ArcA site II. La protecció davant la digestió era visible en ambdues zones alhora, indicant que la unió a una caixa no impedeix la unió a l'altra. En els mateixos assajos realitzats amb concentracions menors de proteïna es pot observar que mentre ArcA site I continua protegida, la protecció d'ArcA site II no és apreciable. Aquest fet suggereix una més alta afinitat d'unió amb la caixa I i implica que la unió d'ArcA a aquesta caixa no requereix la unió a la caixa II. Per conèixer realment la mecànica de la unió a les dues zones i si existeix cooperativitat en quan a la unió d'ArcA a les dues regions s'haurien de fer els assajos de retardació en gel i de footprinting amb fragments parcials del promotor que només continguessin una de les caixes.

La localització d'una caixa d'unió per ArcA solapada amb l'hexàmer -10 (ArcA site I) explica l'efecte repressor sobre la transcripció d'*aldA*. En condicions anaeròbiques, ArcA es fosforila per acció d'ArcB. Un cop fosforilada, pot interaccionar amb aquesta seqüència, i provocar un impediment estèric per la correcta disposició de l'RNA polimerasa i per tant impedir la transcripció.

La segona zona protegida per ArcA-P (ArcA site II) està localitzada sobre la seqüència anomenada URS (upstream regulatory site) per Limon *et al* (1997). La zona URS sembla intervenir en la inducció de l'operó per aldehyds. És possible que en aquestes condicions inductores, ArcA-P s'uneixi a totes dues zones permetent un efecte cooperatiu entre els dos complexos units al DNA. Per analitzar aquesta hipòtesi s'haurien de construir mutants en aquesta caixa i veure l'efecte *in vivo* en condicions aeròbiques i anaeròbiques, però s'hauria de tenir en compte per no mutar cap posició imprescindible pel funcionament de URS.

La mutagènesi dirigida del promotor d'*aldA*, en el qual es van modificar dues posicions de la possible seqüència consens per ArcA en la regió ArcA site I (es conserven 10/10 residus

del consens proposat), mostra que el canvi de bases varia dràsticament el comportament *in vivo* i *in vitro*. A l'hora de decidir quines mutacions es volia introduir es va tenir en compte no mutar cap base essencial pel reconeixement per part de l'RNA polimerasa. Es van construir dos promotors mutants: *ald2* i *ald3*. El promotor mutat *ald2* té una G canviada per una C, i una C canviada per una G. El seu contingut en C+G queda doncs inalterat. Contràriament, en el cas del promotor *ald3*, la G és transformada en una A i la C en una T. El contingut de C+G disminueix, i a més es forma una zona molt rica d'As i d'Ts que té una gran facilitat per curvar-se. Per assegurar que l'elecció havia estat correcte, es va estudiar el patró d'expressió aeròbica basal dels mutants construïts. Es va poder comprovar que el sistema es continuava expressant, tot i que amb una certa disminució respecte el promotor salvatge (87 % per *ald2* i 77 % per *ald3*). En el promotor salvatge, l'expressió del sistema està reprimida en condicions de creixement anaeròbiques (en absència d'acceptor final d'electrons); aquesta repressió desapareix quan es transfereix una mutació *arcA2* en el fons genètic MC4100 Φ (*ald1-lacZ*), indicant el paper del sistema de dos components ArcB/ArcA en la repressió anaeròbica d'*aldA*.

L'expressió dels promotors mutats en la caixa ArcA site I van mostrar el mateix patró. Tant l'expressió anaeròbica de Φ (*ald2-lacZ*) com la de Φ (*ald3-lacZ*) es veu desreprimida per la introducció d'aquestes mutacions (és un 80 % de l'aeròbica). Aquest resultat mostra que el consens que havia estat proposat per Lynch i Lin (1996a) en base a l'estudi *in vitro* de diferents promotors regulats per ArcA, és realment el lloc d'unió de la proteïna, i que les posicions són essencials per la unió del factor de transcripció.

Per tal de poder veure si aquests canvis observats *in vivo* també es donaven *in vitro* es van realitzar assajos de retardació de la mobilitat electroforètica usant com a fragment de DNA els promotors mutats. La longitud del promotor salvatge i dels mutats era la mateixa. Els assajos es van realitzar en paral·lel de manera que els canvis observats només fossin atribuïbles al canvi en els dos parells de bases.

La unió inespecífica d'ArcA al promotor no distingeix entre els promotors mutats i el tipus salvatge. Aquesta observació indica que la unió inespecífica d'ArcA no és mitjançada pel reconeixement de bases sinó un tipus d'interacció dèbil derivada del fet que ArcA és una proteïna d'unió al DNA. Qualsevol proteïna és capaç d'interaccionar amb el DNA de forma inespecífica, encara que només sigui per les interaccions electrostàtiques que s'estableixen entre els grups fosfat carregats negativament i situats a l'exterior de l'hèlix de DNA i els residus proteics bàsics.

Per contra, la retardació específica dels diferents promotors amb ArcA-P sí va mostrar un patró diferent en funció de la seqüència. Pel promotor tipus salvatge s'han definit tres complexos, que per ordre decreixent de mobilitat electroforètica són: complexe I, II i III. El promotor *ald2* forma un complexe similar al complexe I a concentracions superiors que el tipus

salvatge. No s'observa la formació de cap complexe similar a II. És capaç de formar un complexe semblant al complexe III a la mateixa concentració que ho forma el promotor salvatge. Significativament, els complexos formats tenen un aspecte més laxe i poc definit, demostrant que el canvi de les bases nitrogenades afecta l'afinitat en la unió de la proteïna al promotor. Per tant, l'observada disminució de la repressió anaeròbica en els mutants és un efecte de la menor unió d'ArcA-P al promotor a conseqüència d'una disminució de l'afinitat.

El promotor *ald3* té un comportament electroforètic més similar al promotor salvatge en tant que és capaç de formar complexos similars als complexos I, II i III. Els complexos I i II es generen a concentracions superiors que en el cas del promotor salvatge i el complexe III es forma a la mateixa concentració. L'aspecte de les bandes de retardació, tot i no ser tant definides com les del promotor salvatge, és menys laxe que el de les del promotor *ald2*. Possiblement aquest fet és una conseqüència d'haver augmentat el contingut en dA+dT del promotor. Aquest augment provoca una major capacitat de distorsió del DNA. Molts factors de transcripció exerceixen la seva funció provocant curvatura en el seu DNA diana (p.e. l'RNA polimerasa, el factor CRP, IHF, etc.). És possible que ArcA també provoqui alguna distorsió en la topologia del DNA i per tant la seva funció es veuria afavorida per la presència d'un DNA més moldejable. D'aquesta manera, la disminució de l'afinitat en els promotors *ald2* i *ald3* com a conseqüència del canvi de bases és parcialment superada *in vitro* en el promotor *ald3*.

Les diferències observades *in vitro* entre els promotors *ald2* i *ald3* no són detectables *in vivo*. Aquest fet demostra que la interacció de múltiples factors de transcripció regula l'expressió d'un promotor, i que per tant els estudis *in vitro* amb una proteïna purificada només permeten veure les interaccions entre la proteïna en qüestió i el DNA, sense poder apreciar la contribució que altres factors poden tenir en la formació del complexe nucleoproteic.

Els assajos de protecció a la DNasa I amb els promotors mutats van permetre observar que la disminució de l'afinitat no anava associada a una disminució de l'especificitat en la unió d'ArcA-P. Tot i que a concentracions deu vegades més elevades que en el promotor salvatge, ArcA-P continuava protegint les mateixes seqüències de DNA.

La validació total de la seqüència del consens per ArcA requereix una anàlisi mutacional més àmplia, en la resta de les posicions que el conformen. Aquesta tasca requereix una anàlisi prèvia per saber quines de les posicions són modificables sense que s'impedeixi la unió a l'RNA polimerasa. Cal recordar que les caixes que han estat mutades en aquest estudi es troben localitzades sobre la posició -10 del promotor d'*aldA*. Per aquest motiu podria ser interessant fer mutagènesi de caixes per ArcA que no estiguin situades en posicions tant crítiques per l'expressió basal dels gens.

6. IMPORTÀNCIA DEL SISTEMA GÈNIC *glc* EN EL MANTENIMENT DELS NIVELLS INTRACEL·LULARS DE GLIOXILAT

6.1. L'estructura d'operó dels gens estructurals del sistema *glc* facilita el control dels nivells intracel·lulars de glioxilat

L'existència d'un únic promotor que regula conjuntament l'expressió de glicolat oxidasa i malat sintasa G provoca que la síntesi dels dos enzims sigui simultània. Per tant, encara que en alguns estats metabòlics no es requereixi en principi la participació de la glicolat oxidasa, aquesta estarà present si és necessària la malat sintasa G.

Aquesta organització pot respondre a un mecanisme de seguretat en contra la toxicitat del glioxilat. El glioxilat és una molècula altament oxidant, i que pot provocar danys a estructures fonamentals per la cèl·lula (lípid, DNA, etc.) per formació de radicals lliures tòxics (superòxids, hidròxids, etc.). La cèl·lula disposa de mecanismes que intenten paliar els efectes nocius que els radicals lliures poden provocar. Per exemple, *Escherichia coli* disposa de dues superòxids dismutases que la protegeixen del dany oxidatiu provocat pels radicals superòxid (Beyer *et al*, 1991). Evidentment, però, el millor mecanisme de protecció és el control de la formació i acumulació de l'agent oxidant, que en aquest cas és un important metabolit intermediari al voltant del qual es connecten una gran varietat de vies (Figura 4).

El fet que glicolat oxidasa i malat sintasa G es sintetitzin conjuntament és important en aquest sentit per tal de modular els nivells intracel·lulars de glioxilat. Així, quan la glicolat oxidasa genera glioxilat per oxidació de glicolat, la cèl·lula s'assegura sempre la presència d'una malat sintasa (MSG) que convertirà aquest aldehyd en malat. No cal dependre de la malat sintasa A codificada en un locus diferent i mobilitzada majoritàriament en altres condicions fisiològiques.

La completa incorporació dels carbonis del glioxilat al metabolisme general requereix la participació dels enzims de la via del D-glicerat, els quals s'indueixen pel propi glioxilat. La transformació d'aquest compost a través d'aquesta via provoca també la devallada dels seus nivells intracel·lulars. D'altra banda, l'existència a *Escherichia coli* d'una activitat glioxilat reductasa constitutiva determina la formació de glicolat a partir de glioxilat en diferents condicions fisiològiques (creixements en glioxilat i acetat, degradació de purines, etc.) (Figura

4). Aquest fet permet mobilitzar la malat sintasa G en situacions que no impliquin directament la presència de glicolat exogen, assegurant així la posterior metabolització de l'aldehid tòxic.

De la mateixa manera que ha estat discutit pel cas de l'operó *glc*, l'estructura gènica del sistema *ace* en forma d'operó també assegura la detoxificació del glioxilat format. La síntesi d'isocitrat liasa va lligada a la síntesi de malat sintasa A i per tant, sempre que es genera glioxilat per acció de la isocitrat liasa, la cèl.lula disposa d'una activitat enzimàtica que el transforma en una molècula menys reactiva i per tant menys tòxica pel bacteri.

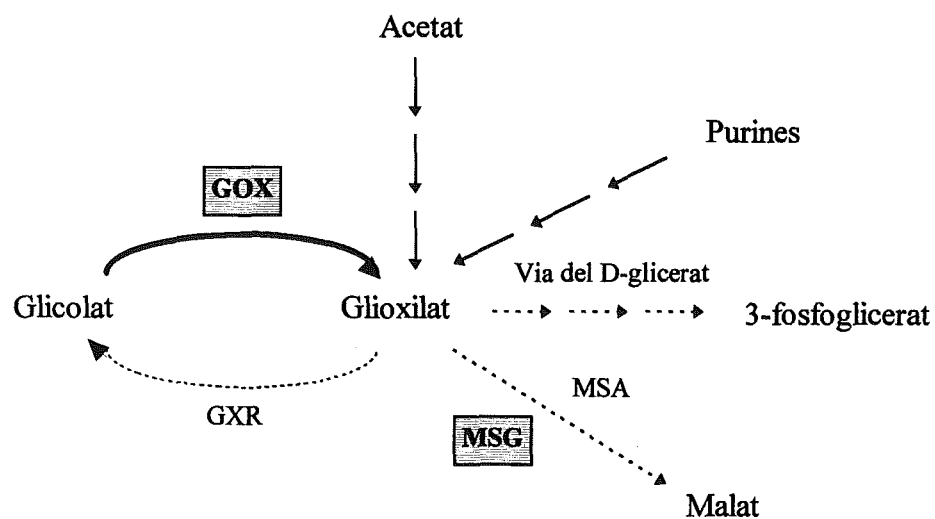


Figura 4. Encreuament metabòlic del glioxilat. Les vies de formació de glioxilat intracel.lular estan indicades amb fletxes de traç continu i les vies que comporten una disminució del glioxilat intern amb traç discontinu. GOX: glicolat oxidasa; GXR: glioxilat reductasa; MSG: malat sintasa G; MSA: malat sintasa A. Els enzims codificats en el locus *glc* estan emmarcats en un rectangle gris.

L'expressió de l'operó *glc* no és mai nul·la, tot i que en certes condicions de creixement s'expressa a nivells molt baixos (p.e. creixements en glucosa). Això implica que sempre existeix dins la cèl.lula un cert nivell de glicolat oxidasa i malat sintasa, i per tant es pot deduir que en contra del que es pensava en un principi, l'operó *glc* no està dissenyat únicament per permetre el creixement d'*Escherichia coli* en glicolat com a font exògena de carboni i energia sinó que és un operó implicat en el control de la concentració intracel.lular de l'intermediari metabòlic glioxilat.

6.2. Importància de la inducció creuada dels sistemes *glc* i *ace*

L'operó *ace*, involucrat en el metabolisme d'acetat, i l'operó *glc*, implicat en el metabolisme de glicolat mostren un cert nivell d'inducció creuada. Aquest entrecruament entre els dos sistemes permet la substitució metabòlica dels dos isoenzims de malat sintasa. L'activitat enzimàtica de les dues malat sintases, malat sintasa A (operó *ace*) i malat sintasa G (operó *glc*) es veu augmentada en unes condicions de creixement que no són les pròpies de cada sistema.

La inducció creuada entre el sistemes *glc* i *ace* es produeix perquè ambdós expressen una activitat enzimàtica del tipus malat sintasa: l'isoenzim malat sintasa G i la malat sintasa A. La capacitat del mutant JA159 (*glcB*, MSG⁻) per créixer en glicolat i la del mutant JA166 (*aceB*, MSA⁻) per créixer en acetat mostren la complementació funcional dels dos isoenzims, així com l'expressió dels seus respectius operons, en unes condicions per les quals en principi no han estat dissenyats. Es podria pensar que el creixement fós permès per la presència d'una tercera malat sintasa induïble indistintament per aquestes dues fonts de carboni. Tanmateix, no s'ha descrit cap altre isoenzim de malat sintasa a *Escherichia coli*, i es descarta la seva existència ja que la soca JA168, doble mutant de malat sintasa A i G és incapaç de créixer en glicolat o en acetat.

Els nivells d'activitat malat sintasa G de la soca salvatge MC4100 crescuda en acetat són sis vegades menors que els obtinguts creixent en glicolat. La mobilització de l'isoenzim G per part de l'acetat no és gaire elevada en la soca salvatge però es veu incrementada en un mutant *aceB*, la qual cosa permet suplir la deficiència en malat sintasa A. L'increment en els nivells d'activitat enzimàtica malat sintasa G es correlaciona amb un augment proporcional de la quantitat de proteïna, mesurada immunològicament. Aquesta correlació indicaria que no existeixen altres mecanismes de regulació de la transcripció addicionals al de la formació del complex nucleoproteic transcripcionalment actiu.

La manca d'activitat enzimàtica glicolat oxidasa en els extractes cel·lulars de la soca MC4100 crescuda en acetat es pot explicar en base als nivells d'expressió del sistema *glc* en aquestes condicions (sis vegades inferior a l'obtinguda en glicolat). La baixa activitat específica d'aquest enzim en els extractes de cèl·lules induïdes per creixement en glicolat fa que els seus nivells es situïn per sota dels límits de detecció en els creixements en acetat. Es pot especular, però, amb la possibilitat que la manca d'activitat glicolat oxidasa en aquestes condicions fós conseqüència de diferents mecanismes de regulació, que impliquessin una inactivació de

l'enzim sintetitzat, o alternativament una expressió gènica diferencial respecte el gen que codifica la malat sintasa G.

La possible presència d'un promotor alternatiu únicament per *glcB* induït per creixements en acetat està descartada ja que l'expressió de $\Phi(\textit{glcBA-lacZ})$ no mostra cap activitat promotora ni en presència d'acetat ni de glicolat, inductor propi del sistema *glc*. Per tant, la inducció creuada dels sistemes *glc* i *ace* s'exerceix sobre l'únic promotor actiu situat a 5' de *glcD*.

La inducció de l'operó *glcDEFGBT* per acetat està mediada per l'acció de la proteïna GlcC. Mutacions en el gen *glcC* impedeixen la inducció de *glcD* per acetat, per tant la proteïna és indispensable per poder activar el sistema en qualsevol condició de creixement. Quan els experiments de segrestament duts a terme amb la soca MC4100 $\Phi(\textit{glcD-lacZ})$ descrits anteriorment (Resultats 4.1.2.) es feien en presència d'acetat com a inductor, els resultats eren els mateixos que en glicolat. Els nivells cromosòmics disminueixen al 20 % quan la soca és transformada amb un plasmidi multicòpia que conté tot el promotor *glcD* (posicions -177 a +94), indicant que el fragment està segrestant un factor necessari per la inducció creuada mediada per acetat.

Per analitzar si la molècula efectora de la inducció de l'expressió de la fusió $\Phi(\textit{glcD-lacZ})$ en creixements en acetat és el propi acetat o bé un altre metabolit format en el seu metabolisme, la fusió transcripcional de l'operó *glc* va ser transferida a diferents fons genètics i assajada en base a l'activitat β -galactosidasa expressada. L'anàlisi paral.lela de la malat sintasa G mitjançant electroimmunodifusió, permet observar que els seus nivells augmenten o disminueixen paral.lelament amb els nivells de transcripció a partir del promotor de *glcD*.

En els cultius en acetat de la soca JA165 (mutant d'isocitrat liasa) s'observa un nivell d'inducció de $\Phi(\textit{glcD-lacZ})$ del mateix ordre que en la soca salvatge. Aquesta soca només pot transformar l'acetat fins a isocitrat, i per tant indica que o bé l'acetat per sí mateix o algun dels metabolits que es formen fins a l'obtenció d'isocitrat pot ser l'inductor de l'operó *glc*. Per tant, sembla que no cal que es formi glioxilat per poder observar inducció del promotor de *glcD* en aquestes condicions.

Encara que no sigui el propi glioxilat qui promou la inducció observada en la soca JA165, l'increment en els seus nivells promou la inducció del sistema *glc*. Així, en la soca JA166 $\Phi(\textit{glcD-lacZ})$ (mutant de malat sintasa A) la inducció del promotor per creixement en presència d'acetat és 2.5 vegades major que la de la soca salvatge. La manca d'un isoenzim de malat sintasa determina segurament un increment en els nivells intracel.lulars de glioxilat i aquesta pot ser la causa de l'augment en l'expressió del sistema. Bé el glioxilat per sí mateix o

el glicolat format per l'activitat constitutiva glixilat reductasa poden ser els responsables de la inducció del sistema *glc* en aquestes condicions.

En la soca JA168 $\Phi(glcD-lacZ)$ (doble mutant de malat sintasa A i G) l'expressió està augmentada cinc cops respecte la soca salvatge, no només en acetat sinó també en hidrolitzat de caseïna. En aquesta soca, els nivells de glixilat són encara majors, ja que cap dels isoenzims de malat sintasa és funcional. El fet que la inducció es produeixi en creixements en hidrolitzat de caseïna implica la formació de glixilat es produeix en la degradació d'aminoàcids (directament o bé per formació d'acetat que es transformarà en glixilat).

En els mutants JA166 i JA168 la inducció en presència d'acetat està augmentada. En tots dos casos la quantitat de glixilat està incrementada intracel·lularment, i això es tradueix en un augment de la inducció creuada. Potser la toxicitat del glixilat és el senyal directe o indirecte que fa activar la transcripció de *glcDEFGBT* i que per tant media la inducció creuada. Pot ser un senyal directe si el glixilat és un efector de la proteïna activadora GlcC. Tantmateix, el fet que la soca JA165 (mutant isocitrat liasa que no forma glixilat a partir d'acetat) sigui capaç d'induir la transcripció de *glcD* sembla descartar o donar menys pes l'efecte directe del glixilat com a inductor de GlcC. És possible, doncs, que actui com un senyal indirecte a través de la seva transformació en glicolat, la qual està afavorida en condicions d'acumulació intracel·lular de glixilat. En aquest cas el paper de la glixilat reductasa seria essencial com a mecanisme de protecció i defensa contra l'oxidació de les estructures cel·lulars que promou el glixilat

La similitud estructural entre l'acetat i el glicolat permet dislumbrar el mecanisme molecular que media la inducció creuada dels operons *glc* i *ace*. A partir de l'estudi de la inducció de la fusió $\Phi(glcD-lacZ)$ per creixement de la soca tipus salvatge MC4100 en diferents fonts de carboni es pot deduir que la proteïna GlcC pot reconèixer diferents molècules estructuralment semblants al glicolat. Aquestes actuarien com a efectors provocant una activació de la proteïna, encara que amb una eficiència menor que el glicolat (com es reflecteix pels menors nivells d'inducció). Tots dos, acetat i glicolat, són àcids monocarboxílics de dos àtoms de carboni. Únicament es diferencien en la substitució d'un hidrogen del grup metil (acetat) per un grup hidroxil (glicolat).

El sistema *ace* ha estat estudiat extensament per diferents autors (Chung *et al*, 1988) (Sunnaborg *et al*, 1990) i l'inductor del sistema s'ha demostrat, per experiments de retardació de la mobilitat electroforètica en gel, que és el fosfoenolpiruvat (Cortay *et al*, 1991). La unió del fosfoenolpiruvat al repressor IclR impedeix que el factor de transcripció s'uneixi al promotor d'*aceBAK* i per tant l'expressió d'aquest operó està desreprimida. La formació de fosfoenolpiruvat per la via del D-glicerat (Ornston i Ornston, 1969) pot ser el mecanisme a través del qual el glicolat provoca la inducció creuada sobre l'operó *ace*.

L'expressió dels sistemes *glc* i *ace* està coordinada en diferents condicions cel·lulars que van més enllà de les que demostren la inducció creuada (creixements en glicolat i acetat). En condicions d'anaerobiosi, igual que succeeix en l'operó *glc*, l'operó *ace* està reprimit. El sistema *ace* és un sistema aeròbic (Clark i Cronan, 1996): l'acetat no pot ser metabolitzat sense el cicle de Krebs, que és eminentment aeròbic. En aquest sistema, s'ha suggerit que la repressió anaeròbica està mediada pel sistema de dos components ArcB/ArcA (Iuchi i Lin, 1988). Per tant, en absència d'oxigen els dos operons estan igualment reprimits i aquest fet implica que cap d'ells no funcionarà sense la presència de l'altre.

El factor IHF activa l'expressió de l'operó *ace* (Resnik *et al*, 1996), igual que succeeix en l'operó *glc*, si bé el mecanisme a través del qual actua és diferent del proposat en aquest treball per l'operó *glc*. Qualsevol variació en el medi extern que comporti un canvi en la concentració intracel·lular d'IHF afectarà per igual als dos sistemes, i per tant aquests continuaran estant coordinadament transcrits i regulats.

El fet que dos sistemes globals de control de la transcripció (ArcB/ArcA i IHF) tinguin els mateixos efectes sobre l'expressió de *glcDEFGBT* i *aceBAK* possibilita que la inducció creuada que presenten tots dos operons varïi en el mateix sentit enfront els canvis als quals es veu sotmès el bacteri.

6.3. L'expressió de *glcDEFGBT* depèn del fluxe de carbonis

El sistema *glc* està induït en cultius de la soca salvatge MC4100 crescuda en presència de glicolat, glioxilat, acetat, D,L-glicerat i D-lactat.

Ja s'ha comentat anteriorment que la inducció de l'operó *glc* promoguda pel glioxilat pot ser una conseqüència indirecta de la seva transformació en glicolat catalitzada per la glioxilat reductasa. Tantmateix, cal destacar l'elevada similitut estructural d'aquests dos compostos de dos àtoms de carboni, la qual cosa podria suggerir el reconeixement del glioxilat per part de la proteïna activadora GlcC. D'altra banda, D-lactat i glicerat són compostos de tres carbonis, que igual que el glicolat tenen estructura d' α -hidroxiàcids. En base als nivells d'activitat β -galactosidasa es podria suggerir que són capaços d'interaccionar amb GlcC i induïr el canvi d'aquesta proteïna a la seva forma transcripcionalment activa. En aquest cas la inducció es podria explicar només en base a la similitut estructural de les molècules inductores, ja que el metabolisme d'aquests compostos no genera cap intermediari comú amb els derivats del glicolat.

Curiosament, els creixements en L-lactat no estimulen la inducció del sistema *glc*, indicant que la disposició espacial del grup hidroxil és important pel reconeixement. Tot i que el glicolat no presenta estereoisòmers, el reconeixement de l'inductor per part de l'activador GlcC sembla ser estereoselectiu (Figura 5).

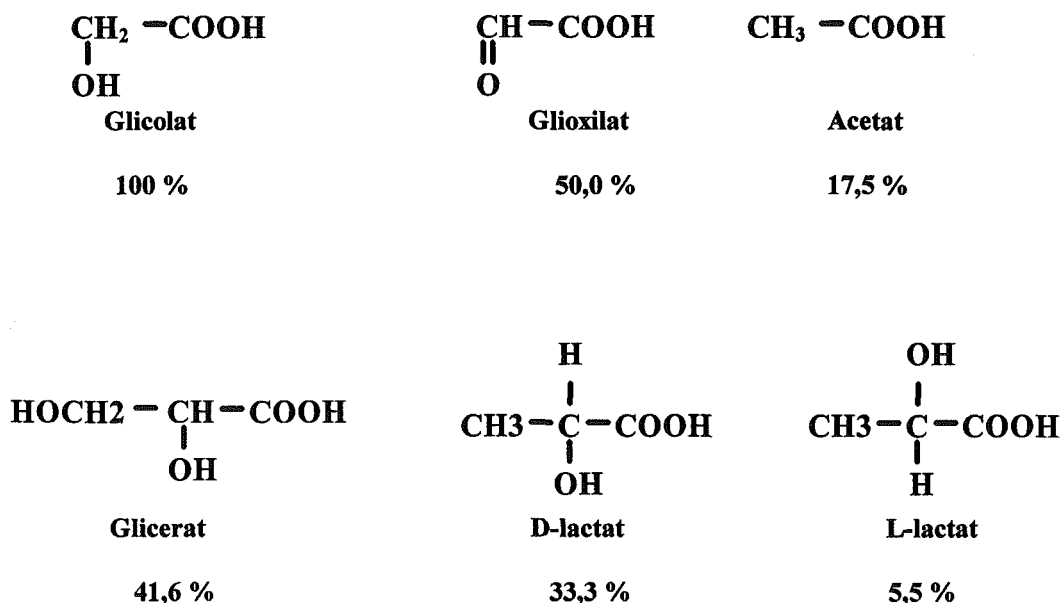


Figura 5. Compostos estructuralment relacionats amb el glicolat que indueixen l'expressió de $\Phi(\text{glcD-lacZ})$ en la soca salvatge MC4100. Es mostra l'estructura de diferents compostos relacionats amb el glicolat i el % d'inducció de la fusió $\Phi(\text{glcD-lacZ})$ (respecte als nivells assolits amb glicolat) en creixements de la soca MC4100 en hidrolitzat de caseïna al 0.05 % en presència de cadascun dels compostos.

L'expressió dels gens estructurals del sistema *glc* està reprimida en creixements en presència de glucosa (20 unitats Miller d'activitat β -galactosidasa). Aquests nivells d'activitat són menors que els observats per altres fonts de carboni no inductores, com són glicerol o xilosa.

L'addició de cAMP al medi no reverteix la repressió exercida per glucosa. Això indica que l'efecte repressor de la glucosa no està mediat per la proteïna CRP i no sembla respondre al model de repressió catabòlica revertible per cAMP. En aquest sentit cal esmentar que l'anàlisi *in silico* del promotor de *glcD* no mostra cap caixa d'unió a CRP.

Cal tenir present, però, que la glucosa pot exercir repressió catabòlica per un mecanisme diferent que no depèn de CRP-cAPM, conegut com a "exclusió de l'inductor". Recentment s'ha descrit que en aquest mecanisme està implicat el producte del gen *ptsG*, el qual en la seva forma defosforilada és un inhibidor de la lactosa permeasa (Kiwata *et al.*, 1997). D'aquesta manera la glucosa promou una activació del gen *ptsG* que condueix a la inactivació d'algunes permeases. Aquest podria ser el mecanisme o un dels mecanismes implicats en la repressió per glucosa exercida sobre l'operó *glc*. Una inactivació del(s) transportador(s) del glicolat evitaria l'entrada de l'inductor a l'interior de les cèl·lules mentre en el medi la glucosa hi sigui present. El fet que el moduló CRP no controli l'expressió del sistema *glc* sembla indicar que la seva funció principal no és la de créixer a expenses de glicolat extern, i apunta la funció del sistema *glc* en el metabolisme intermediari d'*Escherichia coli*.

Els cultius en presència de glicerol també tenen un efecte repressor sobre l'expressió de *glcDEFGHT*, encara que potser no tant marcat com la glucosa (50 unitats Miller d'activitat β -galactosidasa). La repressió que diferents fonts de carboni exerceixen sobre els gens estructurals del sistema *glc* suggereixen la regulació del mateix per un sistema global de control de la transcripció que dirigeix el fluxe metabòlic de compostos carbonats i que provoca una repressió per catabolít independent de la concentració intracel·lular de cAMP. Aquest sistema global de la transcripció podria ser el sistema CRA. Creixements en succinat, que és una font de carboni gluconeogènica, mostren uns nivells de transcripció similars als de l'hidrolitzat de caseïna, la qual cosa indica que en aquest cas no existeix una repressió catabòlica. Això és indicatiu que el sistema *glc* pot estar sota control de CRA.

En la seqüència del promotor de *glcDEFGHT* no s'identifica cap seqüència consens per la unió a CRA. Estan descrits, però, operons que pertanyen al moduló CRA i que no tenen en els seus promotors seqüències d'unió al factors de transcripció, p.e. *cydAB* (Ramseier *et al.*, 1996). La demostració que el sistema *glc* està sota el control de CRA s'haurà de comprovar en mutants deficients en aquesta proteïna reguladora.

De tots els sistemes que mostren caixes d'unió a CRA, només els sistemes implicats en la utilització de fonts de carboni pertanyen a aquest moduló. L'acció de CRA com activador o repressor només es pot apreciar per assajos *in vivo*, ja que no sembla haver distinció entre les seqüències consens reconegudes quan aquesta proteïna actua com a activador que com a repressor. El factor que determina que CRA actui com a activador o repressor és la interacció que pugui establir amb d'altres proteïnes reguladores que actuïn sobre el promotor. Diferents sistemes que estan sota el control de CRA també estan regulats per IHF, i del resultat d'aquesta interacció sorgeix l'efecte positiu o negatiu de CRA.

El sistema *ace* està regulat positivament per CRA. Sembla lògic que els sistemes *glc* i *ace*, que presenten inducció creuada i que estan conjuntament regulats per IHF i ArcB/ArcA també estiguin regulats en quant a la repressió catabòlica i a la direcció del fluxe de carbonis.

6.4. Formació endògena de glicolat

El fet que el glicolat i el glioxilat siguin interconvertibles dins la cèl.lula per acció de la glicolat oxidasa i de la glioxilat reductasa fa difícil saber quin dels dos compostos és la molècula efectora de GlcC, l'activador transcripcional del sistema *glc*. L'estudi de l'expressió de la fusió $\Phi(glcD-lacZ)$ en els fons genètics de mutants deficientes en glicolat oxidasa (soques JA155(*glcD::cat*), JA156(*glcE::cat*) i JA157(*glcF::cat*)) evidencia que la incapacitat de les cèl.lules per transformar glicolat en glioxilat incrementa la transcripció d'aquesta fusió de manera molt considerable en condicions en les quals no hi ha glicolat exogen, com són els creixements en hidrolitzat de caseïna. Sembla, doncs, que el metabolisme d'aminoàcids genera glicolat que s'acumula en els mutants glicolat oxidasa. L'expressió de $\Phi(glcD-lacZ)$ en altres mutants del sistema (mutants de malat sintasa G, de GlcG i del possible transportador) no es veu modificada en creixements en hidrolitzat de caseïna respecte la soca salvatge MC4100, la qual cosa sembla donar pes a la hipòtesi que la formació i acúmulo de glicolat intracel.lular en creixements en hidrolitzat de caseïna pot ser el responsable de l'increment en la transcripció de la fusió de l'operó en aquests mutants.

Aquest efecte d'augment de l'expressió de l'operó *glc* en mutants glicolat oxidasa podria ser causat per un efecte dosi de la proteïna activadora GlcC. Aquesta possibilitat està descartada perquè els nivells d'expressió de la fusió $\Phi(glcC-lacZ)$ en els mutants de glicolat oxidasa no varia respecte la soca salvatge MC4100. Per tant, aquest resultat sembla confirmar que l'efecte observat en els mutants és derivat de l'augment dels nivells intracel.lulars de la molècula efectora.

Estan descrites dues vies per la formació de glicolat intracel.lular: a partir de l'oxidació de glicolaldehid per acció de l'aldehid deshidrogenasa (codificada pel gen *aldA*) o a partir de la reducció del glioxilat catalitzada per la glioxilat reductasa. L'estudi de l'expressió de l'operó *glc* en mutants d'aquests enzims pot ajudar a elucidar el mecanisme d'acumulació de glicolat en els mutants glicolat oxidasa creixent en hidrolitzat de caseïna.

El gen que codifica la glioxilat reductasa no ha estat ni localitzat ni descrit, per la qual cosa la introducció d'una mutació en el seu gen i les conseqüències que aquesta pot comportar no s'han pogut analitzar en aquest treball.

Una mutació en el gen *aldA* va ser transferida al fons genètic de la soca MC4100 i de les soques mutants glicolat oxidasa. En les soques resultants es va assajar l'activitat transcripcional de la fusió $\Phi(\textit{glcD-lacZ})$, i es va poder observar que la mutació en aquest gen no varia en absolut el patró d'expressió. Aquesta observació indica que en els mutants glicolat oxidasa el glicolat acumulat en la degradació d'aminoàcids no prové de l'oxidació de glicolaldehid.

Aquesta formació de glicolat podria ser produïda per una reacció enzimàtica no descrita o bé per reducció de glioxilat (catalitzada per la glioxilat reductasa). La segona hipòtesi està reforçada pel resultat del doble mutant de malat sintasa G i A (té molt disminuïda la capacitat d'eliminar glioxilat), en el qual s'observa un increment de l'expressió del sistema en creixements en hidrolitzat de caseïna. Aquest fet sembla indicar que existeix una connexió entre el sistema *glc* i el metabolit central glioxilat a través de l'enzim constitutiu glioxilat reductasa. Per tant, el sistema *glc* no sembla estar dissenyat únicament per créixer a expenses de glicolat com a font externa de carboni i energia sinó que sembla intervenir activament en el control dels nivells de glioxilat (quan aquests augmenten el sistema s'indueix per la seva acció directe com a molècula efectora o per la formació de glicolat). El rol del sistema *glc* en l'encreuament metabòlic del glioxilat mostra la importància d'aquest sistema en la fisiologia cel·lular d'*Escherichia coli*, i podria explicar el fet que l'expressió d'aquest sistema mai no és completament nul·la en cap font de carboni assajada.

Curiosament, els nivells d'expressió de la fusió $\Phi(\textit{glcD-lacZ})$ en un mutant glicolat oxidasa creixent en glioxilat són menors que en la soca tipus salvatge. Quan aquest mutant creix en hidrolitzat de caseïna, la degradació d'aminoàcids genera quantitats baixes de glioxilat, que per acció de la glioxilat reductasa es transformen en glicolat. El glicolat s'acumula i provoca la sobreexpressió de l'operó *glc*. La glioxilat reductasa possiblement és molt afí al glioxilat, i en aquestes condicions la concentració intracel·lular de l'aldehid no és limitant per l'activitat reductasa però sí per les altres vies de transformació de glioxilat. Quan el mutant es fa créixer en presència de glioxilat, la concentració intracel·lular del mateix es veu molt incrementada de forma ràpida i aquesta pot ser capaç d'induir l'expressió dels gens de la seva pròpia degradació (els enzims de la via del D-glicerat). La transformació de glioxilat a 3-fosfoglicerat podria eliminar pràcticament tot el glioxilat de l'interior cel·lular i disminuir la seva concentració a nivells inferiors dels catalíticament actius per la glioxilat reductasa. Això semblaria indicar que el glicolat seria l'inductor del sistema *glc* i que aquest es formaria a partir del glioxilat. Per tant, la relació entre les activitats enzimàtiques codificades en el locus *glc* i els nivells interns de glioxilat quedaria palesa en aquest resultat.

CONCLUSIONS

1. La seqüenciació de la regió del locus *glc* (min 64.5 del cromosoma d'*Escherichia coli*) que complementa l'activitat glicolat oxidasa deficient en el mutant JA151, mostra que en aquesta regió estan codificats cinc gens (*glcC*, *glcD*, *glcE*, *glcF* i *glcG*) relacionats amb el metabolisme de glicolat i glioxilat.

2. L'anàlisi de la seqüència aminoacídica derivada dels gens que codifiquen la glicolat oxidasa, l'expressió dels mateixos en un sistema de maxicèl.lules, la seva disrupció per inserció d'un casset CAT i la localització subcel.lular dels seus productes d'expressió permeten concloure que la glicolat oxidasa és un enzim heteroligomèric, format per tres subunitats (GlcD, GlcE i GlcF), i lligat a la membrana plasmàtica.

3. Els estudis d'expressió per Northern Blot i les fusions transcripcionals al gen *lacZ* dels diferents gens del locus *glc* indiquen que el sistema *glc* està organitzat en dues unitats transcripcionals, una de les quals pertany al gen regulador i l'altra als gens estructurals, els quals s'organitzen en forma d'operó (*glcDEFGBT*).

4. L'anàlisi de la seqüència aminoacídica derivada del gen *glcC*, la seva disrupció per inserció d'un casset CAT, l'efecte dosi de la proteïna en la transcripció i els estudis de retardació de la mobilitat electroforètica amb extractes crus del mutant JA154 (*glcC::cat*) mostren que GlcC és l'activador transcripcional específic de l'operó *glcDEFGBT* i que és alhora un repressor de la seva pròpia transcripció.

5. L'anàlisi de l'expressió de la fusió transcripcional de l'operó *glcDEFGBT* en fons genètics de mutants en IHF i els estudis de retardació de la mobilitat electroforètica amb extractes crus dels mutants JA162 (*himA::cat*) i JA163 (*himD3::cat*) permeten concloure que la transcripció de l'operó *glc* és totalment dependent del factor de transcripció global IHF.

6. La repressió de l'operó *glcDEFGBT* en creixements en glucosa, la incapacitat de revertir aquest efecte repressor per addició de cAMP en el medi i la inexistència de seqüències consens per la unió de cAMP-CRP en el seu promotor mostren que el sistema *glc* no forma part del moduló CRP.

7. L'expressió de la fusió transcripcional Φ (*glcD-lacZ*) en condicions de creixement anaeròbiques (en presència o absència d'acceptors finals d'electrons), l'anàlisi de la mateixa en mutants *arcA*, els estudis de retardació de la mobilitat electroforètica i els estudis de protecció

a la DNasa I amb proteïna ArcA purificada permeten concloure que la repressió de l'operó *glc* en condicions d'anaerobiosi està exercida pel sistema de dos components ArcB/ArcA.

8. L'expressió del sistema *aldA* en mutants de la caixa ArcA-P en la regió -10 del seu promotor, els estudis de retardació de la mobilitat electroforètica i els de protecció a la DNasa I amb proteïna ArcA purificada amb els promotors mutats indiquen que el consens proposat per Lynch i Lin és la seqüència a la qual s'uneix ArcA i que està realment implicada en la repressió anaeròbica mediada per aquest factor de transcripció global.

9. L'anàlisi genètica de diferents mutants del sistema *glc* i del sistema *ace*, l'anàlisi bioquímica de les activitats enzimàtiques codificades en el locus *glc* creixent en glicolat, acetat i glioxilat, i l'estudi de l'expressió de l'operó *glcDEFGBT* per acetat mitjançant fusions transcripcionals mostren que els sistemes *glc* i *ace* presenten una inducció creuada.

10. L'expressió de l'operó *glcDEFGBT* i de *glcC* en els fons genètics de mutants glicolat oxidasa i en dobles mutants de glicolat oxidasa i d'aldehid deshidrogenasa, creixent en hidrolitzat de caseïna, permeten concloure que el glicolat s'acumula intracel·lularment en condicions en les quals no es troba present en el medi de cultiu, i que aquest no prové de l'oxidació de glicolaldehid.

11. L'estructura d'operó dels gens estructurals del sistema *glc*, la seva sobreexpressió en mutants de malat sintasa A i/o de malat sintasa G en creixements en hidrolitzat de caseïna, l'existència de l'activitat enzimàtica constitutiva glioxilat reductasa i el fet que l'expressió de l'operó mai no és nul·la indiquen que el sistema *glc* juga un paper molt important en el control de les concentracions de l'intermediari metabòlic glioxilat.

BIBLIOGRAFIA

VII. BIBLIOGRAFIA

Altendorf, K., Siebers, A. i W. Epstein. (1992). "The KDP ATPase of *Escherichia coli*". *Ann. N. Y. Acad. Sci.* **617**: 228-243.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. i K. Struhl. (1991). "Current protocols in molecular biology". Ed. John Wiley & Sons, New York.

Badía, J., Baldomà, L., Aguilar, J. i A. Boronat. (1989). "Identification of the *rhaA*, *rhaB* and *rhaD* gene products from *Escherichia coli*". *FEMS Microbiol. Lett.* **65**: 253-258.

Baldomà, L. i J. Aguilar. (1987). "Involvement of lactaldehyde dehydrogenase in several metabolic pathways of *Escherichia coli*". *J. Biol. Chem.* **262**: 13991-13996.

Baldomà, L. i J. Aguilar. (1988). "Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic-anaerobic regulation of L-lactaldehyde dissimilation". *J. Bacteriol.* **170**: 416-421.

Bartkus, J. M. i R. P. Mortlock. (1986). "Construction of an improved D-arabinose pathway in *Escherichia coli* K12". *J. Bacteriol.* **165**: 704-709.

Beck, C. F. i R. A. Warren. (1988). "Divergent promoters, a common form of gene organization". *Microbiol. Rev.* **52**: 318-326.

Bell, A. i S. Busby. (1994). "Location and orientation of an activating region in the *Escherichia coli* transcription factor, FNR". *Mol. Microbiol.* **11**: 383-390.

Berger, D. K., Narberhaus, F., Lee, H.-S. i S. Kustu. (1995). "In vitro studies of the domains of the nitrogen fixation regulatory protein NIFA". *J. Bacteriol.* **177**: 191-199.

- Bertoni, G., Pérez-Martín J. i V. de Lorenzo. (1997). "Genetic evidences of separate repressor and activator activities of the XylR regulator of the TOL plasmid, pWW0, of *Pseudomonas putida*". *Mol. Microbiol.* **23**: 1221-1227.
- Beyer, W. F., Imlay, J. i I. Fridovich. (1991). "Superoxide dismutases". *Prog. Nucleic Acid Res. Mol. Biol.* **40**: 221-253.
- Birnboim, H. C. i J. Doly. (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA". *Nucleic Acids Res.* **7**: 1513-1523.
- Blattner, F. R., Plunkett, G. III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. i Y. Shao. (1997). "The complete genome sequence of *Escherichia coli* K-12". *Science.* **277**: 1153-1174.
- Boronat, A., Caballero, E. i J. Aguilar. (1983). "Experimental evolution of a metabolic pathway for ethylene glycol utilization by *Escherichia coli*". *J. Bacteriol.* **153**: 134-139.
- Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. i H. Buc. (1989). "Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*". *EMBO J.* **8**: 4289-4296.
- Branden, C. i J. Tooze. (1991). "Introduction to protein structure". p. 148-150. Garland Publishing, New York.
- Brown, T. D. K., Jones-Mortimer, M. C. i H. L. Kornberg. (1977). "The enzymatic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*". *J. Gen. Microbiol.* **102**: 327-336.
- Busby, S. i R. H. Ebright (1994). "Promoter structure, promoter recognition, and transcription activation in prokaryotes". *Cell.* **79**: 743-746.
- Busby, S. i A. Kolb. (1996). "The CAP modulon", p. 255-279. In *Regulation of gene expression in Escherichia coli*. Ed. E. C. C. Lin and A. S. Lynch. Landes Co., Austin, TX.
- Calb, R., Davidovitch, A., Koby, S., Giladi, H., Goldenberg, S., Margalit, H., Holtel, A., Timmis, K., Sanchez-Romero, J. M., de Lorenzo, V. i A. B. Oppenheim. (1996). "Structure and function of the *Pseudomonas putida* integration host factor". *J. Bacteriol.* **178**: 6319-6326.

- Casadaban, M. J. (1976). "Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu". *J. Mol. Biol.* **104**: 541-555.
- Cedar, H. (1988). "DNA methylation and gene activity". *Cell.* **48**: 5-6.
- Chang, Y.-Y., Wang, A.-Y. i J. E. Cronan, Jr. (1993). "Molecular cloning, DNA sequencing, and biochemical analyses of *Escherichia coli* glyoxylate carboligase". *J. Biol. Chem.* **268**: 3911-3919.
- Chen, C. A. i J. G. Belasco. (1990). "Degradation of *pufLMX* mRNA in *R. capsulatus* is initiated by nonrandom endonucleolytic cleavage". *J. Bacteriol.* **172**: 4578-4586.
- Chen, Y. M., Zhu, Y. i E. C. C. Lin. (1987). "NAD-linked aldehyde dehydrogenase for aerobic utilization of L-fucose and L-rhamnose by *Escherichia coli*". *J. Bacteriol.* **169**: 3289-3294.
- Chin, A. M., Feucht, B. U. i M. H. Saier, Jr. (1987). "Evidence for the regulation of gluconeogenesis by the fructose phosphotransferase system in *Salmonella typhimurium*". *J. Bacteriol.* **169**: 897-899.
- Chung, T., Klumpp, D. J. i D. C. LaPorte. (1988). "Glyoxylate bypass operon of *Escherichia coli*: cloning and determination of the functional map". *J. Bacteriol.* **170**: 386-392.
- Chung, T., Klumpp, D. J. i D. C. LaPorte. (1993). "Relative expression of the products of glyoxylate bypass operon: contributions to transcription and translation". *J. Bacteriol.* **175**: 4572-4575.
- Cocks, G. T., Aguilar, J. i E. C. C. Lin. (1974). "Evolution of L-1,2-propanediol catabolism in *Escherichia coli* by recruitment of enzymes for L-fucose and L-lactate metabolism". *J. Bacteriol.* **118**: 83-84.
- Cohen, S. N. i K. J. McDowall. (1997). "RNase E: still a wonderfully mysterious enzyme". *Mol. Microbiol.* **23**: 1099-1106.
- Cole, S. T. (1982). "Nucleotide sequence coding for the flavoprotein subunit of the fumarate reductase of *Escherichia coli*". *Eur. J. Biochem.* **122**: 479-484.
- Cole, S. T., Eiglmeier, K., Ahmed, S., Honore, N., Elmes, L, Anderson, W. F. i J. H. Weiner. (1988). "Nucleotide sequence and gene-polypeptide relationships of the *glpABC* operon encoding the anaerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12" *J. Bacteriol.* **170**: 2448-2456.

Cole, S. T., Grundtröm, T., Jaurin, B., Robinson, J. J. i J. H. Weiner. (1982). "Location and nucleotide sequence of *frB*, the gene encoding for the iron-sulphur protein subunit of the fumarate reductase of *Escherichia coli*". *Eur. J. Biochem.* **126**: 211-216.

Compan, I. i D. Touati. (1993). "Interactions of six global transcription regulators in expression of manganese superoxide dismutase in *Escherichia coli* K-12". *J. Bacteriol.* **175**: 1687-1696.

Compan, I. i D. Touati. (1994). "Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA". *Mol. Microbiol.* **11**: 955-964.

Cortay, J., Bleicher, F., Rieul, C., Reeves, H. C. i A. J. Cozzone. (1988). "Nucleotide sequence and expression of the *aceK* gene encoding for isocitrate dehydrogenase kinase/phosphatase in *Escherichia coli*". *J. Bacteriol.* **170**: 88-97.

Cortay, J. C., Nègre, D., Galinier, A., Duclos, B., Perrière, G. i A. J. Cozzone. (1991). "Regulation of the acetate operon in *Escherichia coli*: purification and functional characterization of the IclR repressor". *EMBO J.* **10**: 675-679.

Cotter, P. A., Chepuri, V., Gennis, R. B. i R. P. Gunsalus. (1990). "Cytochrome *o* (*cyoABCDE*) and *d* (*cydAB*) oxidase gene expression in *Escherichia coli* is regulated by oxygen, pH, and the *fnr* gene product" *J. Bacteriol.* **172**: 6333-6338.

Cronan, J. E. Jr. i D. C. LaPorte. "Tricarboxylic acid cycle and glyoxylate bypass". (1996). p. 206-216. In *Escherichia coli and Salmonella. Cellular and molecular biology*. Ed. Frederick C. Neihardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press. Washington D. C.

Daniel, R., Stuert, K. i G. Gottschalk. (1995). "Biochemical and molecular characterization of the oxidative branch of the glycerol utilization by *Citrobacter freundii*". *J. Bacteriol.* **177**: 4392-4401.

Darlison, M. G. i J. R. Guest. (1984). "Nucleotide sequence encoding the iron-sulphur protein subunit of the succinate dehydrogenase of *Escherichia coli*". *Biochem. J.* **223**: 507-517.

Davis, R. W., Botstein, D. i J. R. Roth. (1980). "A manual for genetic engineering. Advanced bacterial genetics". Ed. Cold Spring Harbor Laboratory. Cold Spring Harbor, N. Y.

- Deretic, V., Leveau, H. J., Mohr, C. D. i N. S. Hibler. (1992). "In vitro phosphorylation of AlgR, a regulator of mucoidy in *Pseudomonas aeruginosa*, by a histidine protein kinase and effects of small phospho-donor molecules". *Mol. Microbiol.* **6**: 2761-2767.
- Ditto, M. D., Roberts, D. i R. A. Weisberg. (1994). "Growth phase variation in integration host factor in *Escherichia coli*". *J. Bacteriol.* **176**: 3738- 3748.
- Dombroski, A., Walter, W., Record, M., Siegele, D. i C. Gross. (1992). "Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA". *Cell.* **70**: 501-512.
- Dong, J. M., Taylor, J. S., Latour, D. J., Iuchi, S. i E. C. C. Lin. (1993). "Three overlapping *lct* genes involved in L-lactate utilization by *Escherichia coli*". *J. Bacteriol.* **175**: 6671-6678.
- Dove, S. L., Joung, J. K. i A. Hochschild. (1997). "Activation of prokaryotic transcription through arbitrary protein-protein contacts". *Nature.* **386**: 627-630.
- Drapal, N. i G. Sawers. (1995). "Purification of ArcA and analysis of its interaction with the *pfl* promoter regulatory region". *Mol. Microbiol.* **16**: 597-607.
- Durchslag, H. i P. Zipper. (1988). "Primary and post-irradiation inactivation of the sulphhydryl enzyme malate synthase. Correlation of protective effects of additives". *FEBS Lett.* **237**: 208-212.
- Ebright, R. H., Ebright, Y. M. i Gunasekera, A. (1989). "A Consensus DNA site for the *Escherichia coli* catabolite gen activator protein (CAP): CAP exhibits a 450 fold higher affinity for the DNA site than for the *Escherichia coli lac* site". *Nucl. Acids Res.* **17**: 10295-10305.
- Edenborn, H. i C. D. Lichfield. (1985). "Glycollate metabolism by *Pseudomonas sp.*, strain S227, isolated from a coastal marine sediment". *Mar. Biol.* **88**: 199-205.
- Egli, C., Thüer, M., Suter, D., Cook, A. M. i T. Leisinger. (1989). "Monochloro and dichloroacetic acids as carbon sources for a stable, methanogenic mixed culture". *Arch. Microbiol.* **152**: 218-223.
- Elliot, T. (1992). "A method for constructing single-copy lac fusions in *Salmonella typhimurium* and its application to the *hemA-prfA* operon". *J. Bacteriol.* **174**: 245-253.
- Ehretsmann, C. P., Carpousis, A. J. i H. M. Kirsch. (1992). "mRNA degradation in prokaryotes". *FASEB J.* **6**: 3186-3192.

- Falmagne, P., Vanderwinkel, E. i Wiame, J. M. (1965). "Mise en evidence de deux malate synthases chez *Escherichia coli*". *Biochim. Biophys. Acta.* **99**: 246-258.
- Feinberg, A. P. i B. Vogelstein. (1984). "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity". *Anal. Biochem.* **132**: 6-13.
- Finkel, S. E. i R. C. Johnson. "The Fis protein: it's not just for DNA inversion anymore". (1992). *Mol. Microbiol.* **6**: 3257-3265.
- Freestone, P., Grant, S., Toth, I. i V. Norris. (1995). "Identification of phosphoproteins in *Escherichia coli*". *Mol. Microbiol.* **15**: 573-580.
- Friedman, D. I. (1988). "Integration host factor: a protein for all reasons". *Cell.* **55**: 545-554.
- Friedrich, M. i B. Schink. (1991) "Fermentative degradation of glyoxylate by a new strictly anaerobic bacterium". *Arch. Microbiol.* **156**: 392-397.
- Friedrich, M., Ladeser, U. i B. Schink. (1991). "Fermentative degradation of glycolic acid by defined syntrophic cocultures". *Arch. Microbiol.* **156**: 517-534.
- Fuqua, W. C. (1992). "An improved chloramfenicol resistance gene cassette for site-directed marker replacement mutagenesis". *BioTechniques.* **12**: 223-225.
- Furuya, A. i J. A. Hayashi. (1963). "Glycolic acid oxidation by *Escherichia coli* adapted to glycolate". *J. Bacteriol.* **85**: 1129-1131.
- Gallegos, M. T., Marqués, S. i J. L. Ramos. (1996). "Expression of the TOL plasmid *xylS* gene in *Pseudomonas putida* occurs from a σ^{70} -dependent promoter or from a σ^{70} -and σ^{54} -dependent tandem promoters according to the aromatic compound used for growth". *J. Bacteriol.* **178**: 2356-2361.
- García-Martín, C., Baldomà, L., Badía, J. i J. Aguilar. (1992). "Nucleotide sequence of the *rhaR-sodA* interval specyfing *rhatT* in *Escherichia coli*" *J. Gen. Microbiol.* **138**: 1109-1116.
- Garnak, M. i H. C. Reeves. (1979). "Phosphorylation of isocitrate dehydrogenase of *Escherichia coli*". *Science.* **203**: 1111-1112.

Georgellis, D., Barlow, T., Arvidson, S. i A. von Gabain "Retarded mRNA turnover in *Escherichia coli*: a means of maintaining gene expression during anaerobiosis" (1993). *Mol. Microbiol.* **9**: 375-382.

Georgellis, D., Lynch, A. S. i E. C. C. Lin. (1997). "In vitro phosphorylation study of the Arc two-component system of *Escherichia coli*". *J. Bacteriol.* **179**: 5429-5435.

Gerlach, P., Valentin-Hansen, P. i E. Bremer. (1990). "Transcriptional regulation of the *cytR* repressor gene of *Escherichia coli*: autoregulation and positive control by cAMP/CAP complex". *Mol. Microbiol.* **4**: 479-488.

Ghalambor, M. i E. C. Heath. (1962). "The metabolism of L-fucose II. The enzymatic cleavage of L-fuculose-1-phosphate". *J. Biol. Chem.* **237**: 2427-2433.

Goodrich, J. A., Schwartz, M. L. i W. R. McClure. (1990). "Searching for and predicting the activity of sites for DNA binding proteins: compilations and analysis of the binding sites for *Escherichia coli* integration host factor (IHF)". *Nucleic Acids Res.* **18**: 4993-5000.

Goosen, N. i P. van der Putte. (1995). "The regulation of transcription initiation by integration host factor". *Mol. Microbiol.* **16**: 1-7.

Gralla, J. D. i J. Collado-Vides. (1996). "Organization and function of transcription regulatory elements", p. 1232-1245. In *Escherichia coli and Salmonella. Cellular and molecular biology*. Ed. Frederick C. Neihardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press. Washington D. C.

Granston, B. E. i H. A. Nash. (1993). "Characterization of a set of integration host factor mutants deficient for DNA binding". *J. Mol. Biol.* **234**: 45-59.

Gray, C. T., Wimpenny, J. W. T., Hugues, D. E. i M. R. Mossman. (1966). "Regulation of metabolism in facultative bacteria. I. Structural and functional changes in *Escherichia coli* associated with shifts between the aerobic and anaerobic states". *Biochim. Biophys. Acta.* **117**: 22-32.

Green, M. i S. S. Cohen. (1956). "The enzymatic conversion of L-fucose to L-fuculose". *J. Biol. Chem.* **219**: 557-568.

Green, J. i J. R. Guest. (1993). "Activation of FNR-dependent transcription by iron: an *in vitro* switch for FNR". *FEMS Microbiol. Lett.* **113**: 219-222.

- Green, J. i J. R. Guest. (1994). "Regulation of transcription at the *ndh* promoter of *Escherichia coli* by FNR and novel factors". *Mol. Microbiol.* **12**: 433-444.
- Gualerzi, C. i C. L. S. Pon. (1990). "Initiation of mRNA translation in prokaryotes". *Biochem.* **29**: 5881-5889.
- Gunsalus, R. P. (1992). "Control of electron flow in *Escherichia coli*: coordinated transcription of respiratory pathway genes". *J. Bacteriol.* **174**: 7069-7074.
- Hacking, A. J. i E. C. C. Lin. (1977). "Regulatory changes in the evolution of a catabolic pathway for propanediol in *Escherichia coli*". *J. Bacteriol.* **130**: 832-838.
- Hanahan, D. (1994). "Technique for transformation of *Escherichia coli*". Vol. 1, p. 1-35. In *DNA cloning: a practical approach*. Ed. D. M. Glover and B. D. Hames. Second Edition. IRL Press, Oxford.
- Hansen, M. J., Chen, L.-C., Fejzo, M. L. i J. G. Belasco. "The *ompA* 5' untranslated region impedes a major pathway for mRNA degradation in *Escherichia coli*". (1994). *Mol. Microbiol.* **12**: 707-716.
- Hansen, R. W. i J. A. Hayashi. (1961). "Glycolate metabolism in *Escherichia coli*". *J. Bacteriol.* **83**: 679-687.
- Harman J. G., McKenney, K., Peterkofsky, A. (1986). "Structure-function analysis of three cAMP-independent forms of the cAMP receptor protein". *J. Biol. Chem.* **261**: 16332-16339.
- Harrington, R. E. (1992). "DNA curving and bending in protein-DNA recognition". *Mol. Microbiol.* **6**: 2549-2555.
- Haydon, D. J., Quail, M. A. i J. R. Guest. (1993). "A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex in *Escherichia coli* is located within the *pdhR* gene". *FEBS Lett.* **336**: 43-47.
- Heath E. C. i M. Ghalambor. (1962) "The metabolism of L-fucose I. The purification and properties of L-fuculose kinase". *J. Biol. Chem.* **237**: 2423-2426.
- Hénaut, A. i A. Dauchin. (1996). "Analysis and predictions from *Escherichia coli* sequences, or *Escherichia coli* In Silico". p. 2047-1066. In *Escherichia coli and Salmonella. Cellular and molecular biology*. Ed. Frederick C. Neihardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin,

K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press. Washington D. C.

Hidalgo, E., Chen, Y. M., Lin, E. C. C. i J. Aguilar. (1991). "Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 *ald* gene encoding aldehyde dehydrogenase". *J. Bacteriol.* **173**: 6118-6123.

Hidalgo, E. i B. Dimple. (1994). "An iron-sulfur center essential for transcriptional activation by the redox-sensing SoxR protein". *EMBO J.* **13**: 138-146.

Hidalgo, E., Limon, A. i J. Aguilar. (1996). "A second *Escherichia coli* gene with similarity to *gapA*". *Microbiología Sem.* **12**: 99-106.

Hikida, M., Atomi, H., Fukuda, Y., Aoki, A., Hishida, T., Teranishi, Y., Ueda, M. i A. Tanaka. (1991). "Presence of two transcribed malate synthase genes in an *n*-alkane utilizing yeast *Candida tropicalis*". *J. Biochem.* **110**: 909-914.

Ho, C., Pratt, E. A. i G. S. Rule. (1989). "Membrane-bound D-lactate dehydrogenase of *Escherichia coli*: a model for protein interactions in membranes". *Biochim. Biophys. Acta.* **988**: 173-184.

Hoch, J. A. i T. J. Silhavy. (1995). "Two-component signal transduction". ASM Press, Washington, D. C.

Holmes, D. S. i M. Quigley. (1981). "A rapid boiling method for the preparation of bacterial plasmids". *Anal. Biochem.* **114**: 43-47.

Houman, F., Díaz-Torres, M. R. i A. Wright. (1990). "Transcriptional antitermination in the *bgl* operon of *Escherichia coli* is modulated by specific RNA binding protein". *Cell.* **62**: 1153-1163.

Hurley, J. H., Dean, A. M., Sohl, J. L., Kosland, D. E. J. i R. M. Stroud. (1990). "Regulation of an enzyme by phosphorylation at the active site". *Science.* **249**: 1012-1016.

Igarashi, K. i A. Ishihama. (1991). "Bipartite functional map of the *E. coli* RNA polymerase α subunit: involvement of the c-terminal region in transcription activation by cAMP-CRP". *Cell.* **65**: 1015-1022.

- Ikemura, T. "Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: proposal for a synonymous codon choice that is optimal for the *Escherichia coli* translational system". (1981). *J. Mol. Biol.* **151**: 389-409.
- Ingledeew, W. J. i R. K. Poole. (1984). "The respiratory chains of *Escherichia coli*". *Microbiol. Rev.* **48**: 222-271.
- Ishihama, A. (1993). "Protein-protein communication within the transcription apparatus". *J. Bacteriol.* **175**: 2483-2489.
- Ishizuka, H., Hanamura, A., Inada, T. i H. Aiba. (1994). "Mechanism of down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene". *EMBO J.* **13**: 3077-3082.
- Iuchi, S. (1993). "Phosphorylation/dephosphorylation of the receiver module at the conserved aspartate residue controls transphosphorylation activity at histidine kinase in sensor protein ArcB of *Escherichia coli*". *J. Biol. Chem.* **268**: 23972-23980.
- Iuchi, S., Aristarkhov, A., Dong, J. M., Taylor, J. S. i E. C. C. Lin. (1994). "Effects of nitrate respiration on expression of the Arc-controlled operons encoding succinate dehydrogenase and flavin-linked L-lactate dehydrogenase". *J. Bacteriol.* **176**: 1695-1701.
- Iuchi, S., Furlong, D. i E. C. C. Lin. (1989). "Differentiation of *arcA*, and *cpxA* mutant phenotypes of *Escherichia coli* by sex pilus formation and enzyme regulation". *J. Bacteriol.* **171**: 2889-2893.
- Iuchi, S. i E. C. C. Lin. (1987). "The *narL* gene product activates the nitrate reductase operon and represses the fumarate reductase and trimethylamine N-oxide reductase operons in *Escherichia coli*". *Proc. Natl. Acad. Sci. USA.* **84**: 3901-3905.
- Iuchi, S. i E. C. C. Lin. (1988). "*arcA* (*dye*), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in anaerobic pathways". *Proc. Natl. Acad. Sci. USA.* **85**: 1888-1892.
- Iuchi, S. i E. C. C. Lin. (1991). "Adaptation of *Escherichia coli* to respiratory conditions: regulation of gene expression". *Cell.* **66**: 5-7.
- Iuchi, S., Matsuda, Z., Fujiwara, T. i E. C. C. Lin. (1990). "The *arcB* gene of *Escherichia coli* encodes a sensor-regulator protein for anaerobic repression of the *arc* modulon". *Mol. Microbiol.* **4**: 715-727.

- Iuchi, S. i L. Weiner. (1996). "Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments". *J. Biochem.* **120**: 1055-1063.
- Jones, H. i R. P. Gunsalus. (1987). "Regulation of *Escherichia coli* fumarate reductase (*frdABCD*) operon expression by alternative electron acceptors and the *fnr* gene product". *Mol. Microbiol.* **15**: 473-482.
- Jovanovic, G. i P. Model. (1997). "PspF and IHF bind co-operatively in the *psp* promoter-regulatory region of *Escherichia coli*". *Mol. Microbiol.* **25**: 473-481.
- Kaiser, D. i R. Losick. (1993). "How and why bacteria talk to each other". *Cell.* **73**: 873-885.
- Kimata, K., Takahashi, H., Inada, T., Postma, P i H. Aiba. (1997). "cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in *Escherichia coli*". *Proc. Natl. Acad. Sci. USA.* **94**: 12914-12919.
- Kohara, Y., Akiyama, K. i K. Isono. (1987). "The physical map of the whole *Escherichia coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library". *Cell.* **50**: 495-508.
- Kornberg, H. L. (1967). "The role and control of the glyoxylate cycle in *Escherichia coli*". *Biochem. J.* **99**: 1-11.
- Kornberg, H. L. i A. M. Gotto. (1961). "The metabolism of C₂ compounds in microorganisms. Synthesis of cell constituents from glycollate by *Pseudomonas sp*". *Biochem. J.* **78**: 69-82.
- Kornberg, H. L. i H. A. Krebs. (1957). "Synthesis of cell constituents from C₂-units by a modified tricarboxilic acid cycle". *Nature.* **179**: 988-991.
- Kuhnke, G., Fritz, H.-J. i R. Ehring. (1987). "Unusual properties of promoter up mutations in the *Escherichia coli* galactose operon and evidence suggesting RNA polymerase-induced DNA bending". *EMBO J.* **6**: 507-513.
- Kuritzkes, D. R., Zhang, X.-Y. i E. C. C. Lin. (1984). "Use of Φ (*glp-lac*) in studies of respiratory regulation of the *Escherichia coli* anaerobic *sn*-glycerol-3-phosphate dehydrogenase genes (*glpAB*)". *J. Bacteriol.* **157**: 591-598.

- Laemli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of the bacteriophage T4". *Nature*. **227**: 680-685.
- Lane, D., Prentki, P. i M. Chandler. (1992). "Use of gel retardation to analyze protein-nucleic acid interactions". *Microbiol. Rev.* **56**: 509-528.
- LaPorte, D. C., Thorsness, P. E. i D. E. Koshland, Jr. (1984). "The branch point effect: ultrasensitivity and ubsensitivity to metabolic control". *J. Biol. Chem.* **259**: 14068-14075.
- LaPorte, D. C., Thorsness, P. E. i D. E. Koshland, Jr. (1985). "Compensatory phosphorylation of ICDH: a mechanism for adaptation to the intracellular environment". *J. Mol. Biol.* **260**: 10563-10568.
- Laurell, C. B. (1966). "Quantitative stimation of proteins by electrophoresis in agarose gel containing antibodies". *Anal. Biochem.* **15**: 45-52.
- Lemire, B. D., Robinson, J. J. i J. H. Weiner. (1982). "Identification of membrane anchor polypeptides of *Escherichia coli* fumarate reductase". *J. Bacteriol.* **152**: 1126-1131.
- Lemire, B. D., Robinson, J. J., Bradley, R. D., Scraba, D. G. i J. H. Weiner. (1983). "Structure of fumarate reductase on the cytoplasmic membrane of *Escherichia coli*". *J. Bacteriol.* **155**: 391-397.
- Lin, E. C. C. (1996). "Dissimilatory pathways for sugars, polyols, and carbohydrates". p. 307-342. In *Escherichia coli and Salmonella. Cellular and molecular biology*. Ed. Frederick C. Neihardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press. Washington D. C.
- Lodi, T., and I. Ferrero. (1993). "Isolation of the DLD gene of *Saccharomyces cerevisiae* encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase". *Mol. Gen. Genet.* **238**: 315-324.
- Lodi, T., O'Connor, D., Goffrini, P. i I. Ferrero. (1994). "Carbon catabolite repression in *Kluyveromyces lactis*: isolation and characterization of the KIDL gene encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase". *Mol. Gen. Genet.* **244**: 622-629.
- Lord, J. M. (1972). "Glycolate oxidoreductase in *Escherichia coli*". *Biochim. Biophys. Acta.* **267**: 227-237.
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. i R. J. Randall. (1951). "Protein measurement with the Folin phenol reagent". *J. Biol. Chem.* **193**: 265-275.

- Ludtke, D., Larson, T. J., Beck, C. i W. Boos. (1982). "Only one gene is required for the *glpT*-dependent transport of *sn*-glycerol-3-phosphate in *Escherichia coli*". *Mol. Gen. Genet.* **186**: 540-547.
- Lukat, G. S., McCleary, W. R., Stock, A. M. i J. B. Stock. (1992). "Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors". *Proc. Natl. Acad. Sci. USA.* **89**: 718-722.
- Lynch, A. S. i E. C. C. Lin. (1996). "Regulation of aerobic and anaerobic metabolism by the Arc system". p. 361-373. In *Regulation of gene expression in Escherichia coli*. Ed. E. C. C. Lin and A. S. Lynch. Landes Co., Austin, TX.
- Lynch, A. S. i E. C. C. Lin. (1996a). "Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters". *J. Bacteriol.* **178**: 6238-6249.
- Macheroux, P., Kieweg, V., Massey, V., Soderlind, E., Stenberg, K. i V. Lindqvist. (1993). "Role of tyrosine 129 in the active site of spinach glycolate oxidase". *Eur. J. Biochem.* **213**: 1047-1054
- Matsubara, H., Inove, K., Hase, T., Hiura, H., Kakuno, T. i J. Yamashita. (1983). "Structure of the extracellular ferredoxin from *Rhodospirillum rubrum*: close similarity to clostridial ferredoxins". *J. Biochem.* **93**: 1385-1390.
- Mel, S. F. and J. J. Mekalanos. (1996). "Modulation of horizontal gene transfer in pathogenic bacteria by in vivo signals". *Cell.* **87**: 795-798.
- Miki, K. i E. C. C. Lin. (1973). "Enzyme complex which couples glycerol-3-phosphate dehydrogenation to fumarate reduction in *Escherichia coli*". *J. Bacteriol.* **114**: 767-771.
- Miller, J. H. (1972). "Experiments in molecular genetics". Cold Spring Harbor Laboratory, Cold spring Harbor, N. Y.
- Molina, I. Pellicer, M.-T., Badia, J., Aguilar, J. i L. Baldomà. (1994). "Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme". *Eur. J. Biochem.* **224**: 541-548.
- Molina-López, J., Govantes, F. i E. J. Santero. (1994). "Geometry of the process of transcription activation al the σ_{54} -dependent *nifH* promoter of *Klebsiella pneumoniae*". *J. Biol. Chem.* **269**: 22657-22662.

- Mori, K. i H. Aiba. (1985). "Evidence for negative control of *cya* transcription by cAMP and cAMP receptor protein in intact *Escherichia coli* cells". *J. Biol. Chem.* **260**: 3063-3070.
- Mowbray, S. L. i L. B. Cole. (1992). "1.7 Å X-ray structure of the periplasmic ribose receptor from *Escherichia coli*". *J. Mol. Biol.* **225**: 155-175.
- Moxon, E. R. i C. F. Higgins (1997). "A blueprint for life". *Nature.* **389**: 120-121.
- Mozola, M. A. i D. I. Friedman. (1985). "A ϕ 80 function inhibitory for growth of lamboid phage in *him* mutants of *Escherichia coli* deficient in IHF. Genetic analysis of the Rha phenotype". *J. Virol.* **140**: 313-327.
- Nash, H. A. (1996). "The HU and IHF proteins: accessory factors for complex protein-DNA assemblies" p. 149-179. In *Regulation of gene expression in Escherichia coli*. Ed. E. C. C. Lin and A. S. Lynch. Landes Co., Austin, TX.
- Nègre, D., Bonod-Bidaud, C., Geourjon, C., Deléage, G., Cozzone, A. J. i J. C. Cortay. (1996). "Definition of a consensus DNA-binding site for the *Escherichia coli* pleiotropic regulatory protein, FruR". *Mol. Microbiol.* **21**: 257-266
- Nishimura, M., Akhmedov, Y. D., Strzalka, K. i T. Akazawa. (1983). "Purification and characterization of glycolate oxidase from pumpkin cotyledons". *Arch. Biochem. Biophys.* **222**: 397-402.
- Nyström, T., Larsson, C. i L. Gustafsson. (1996). "Bacterial defense against aging: role of the *Escherichia coli* ArcA regulator in gene expression, readjusted energy flux and survival during stasis". *EMBO J.* **15**: 3219-3228.
- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins". *J. Biol. Chem.* **250**: 4007-4021.
- Ornston, L. N. i M. K. Ornston. (1969). "Regulation of glyoxylate metabolism in *Escherichia coli* K-12". *J. Bacteriol.* **98**: 1098-1108.
- Paġel, J. M. i G. W. Hatfield. (1991). "Integration host factor-mediated expression of the *ilvGMEDA* operon of *Escherichia coli*" *J. Biol. Chem.* **266**: 1895-1996.

- Parekh, B. S. i G. W. Hatfield. (1996). "Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model". *Proc. Natl. Acad. Sci. USA*. **93**: 1173-1177.
- Park, S.-J., Tseng, C.-P. i R. P. Gunsalus. (1995). "Regulation of succinate dehydrogenase (*sdhCDAB*) operon expression in *Escherichia coli* in response to carbon supply and anaerobiosis: role of ArcA and Fnr". *Mol. Microbiol.* **15**: 473-482.
- Parkinson, J. S. (1993). "Signal transduction schemes of bacteria". *Cell*. **73**: 857-871.
- Pérez-Martín, J. i V. de Lorenzo. (1995). "Integration host factor supresses promiscuous activation of the σ^{54} -dependent promoter *Pu* of *Pseudomonas putida*". *Proc. Natl. Acad. Sci. USA*. **92**: 7277-7281.
- Pérez-Martín, J., Rojo, F. i V. de Lorenzo. (1994). "Promoters responsive to DNA bending: a common theme in prokaryotic gene expression". *Microbial. Rev.* **58**: 268-289.
- Pérez-Martín, J., Timmis, K. N. i V. de Lorenzo (1994a). "Co-regulation by bent DNA" *J. Biol. Chem.* **296**: 22657-22662.
- Pivec, L., Rozkot, F., Sázelová, P. i Vitek, A. (1985). "Consensus simmetry pattern in *Escherichia coli* promoter sequences". *Folia Biol.* **31**: 213-234.
- Prag, G., Greenberg, S. i A. B. Oppenheim. (1997). "Structural principles of prokaryotic gene regulatory proteins and the evolution of repressors and gene activators". *Mol. Microbiol.* **26**: 919-920.
- Price, N. T., Jackson, V. N. i A. P. Halestrap. (1998). "Cloning and sequencing of four new mammalian monocarboxylate transporter (MCT) homologues confirms the existence of a transporter family with ancient past". *Biochem. J.* **329**: 321-328.
- Pruss, G. J. i Drlica. K. (1989). "DNA supercoliling and prokaryotic transcription". *Cell*. **56**: 521-523.
- Ptashne, M. i A. Gann. (1997). "Transcriptional activation by recruitment". *Nature*. **386**: 569-577.
- Quintillà, F. X., Baldomà, L., Badía, J. i J. Aguilar. (1991). "Aldehyde dehydrogenase induction by glutamate in *Escherichia coli*: role of the 2-oxoglutarate". *Eur. J. Biochem.* **202**: 1321-1325.

Ramos, J. L., Marqués, S. i K. N. Timmis. (1997). "Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators". *Annu. Rev. Microbiol.* **51**: 341-373.

Ramseier, T. M., Chien, S. Y. i M. H. Saier, Jr. (1996). "Cooperative interaction between Cra and Fnr in the regulation of *cydAB* operon of *Escherichia coli*". *Curr. Microbiol.* **33**: 270-274.

Ramseier, T. M., Bledig, S., Michotey, V., Feghali, R. i M. H. Saier, Jr. (1995). "The global regulatory protein FruR modulates direction of carbon flow in *Escherichia coli*". *Mol. Microbiol.* **16**: 1157-1169.

Ramseier, T. M., Nègre, D., Cortay, J. C., Scarabel, M., Cozzone, A. J. i M. H. Saier, Jr. (1993). "In vitro binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru*, *pps*, *ace*, *pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*". *J. Mol. Biol.* **234**: 128-144.

Reizer, A., Deutscher, J., Saier, M. H. Jr. i J. Reizer. (1991). "Analysis of the gluconate (*gnt*) operon of *Bacillus subtilis*". *Mol. Microbiol.* **5**: 1081-1089.

Resnik, E., Pan, B., Ramani, N., Freundlich, M. i D. C. LaPorte. (1996). "Integration host factor amplifies the induction of the *aceBAK* operon of *Escherichia coli* by relieving IclR repression". *J. Bacteriol.* **178**: 2715-2717.

Rice, P. A., Yang, S.-W., Mizuchi, K. i H. A. Nash. (1996). "Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn". *Cell.* **87**: 1295-1306.

Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. i R. L. Gourse. "A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase". (1993). *Science.* **262**: 1407-1413.

Rule, G. S., Pratt, E. A., Chin, C. C., Wold, F. i C. Ho. (1985). "Overproduction and nucleotide sequence of respiratory D-lactate dehydrogenase of *Escherichia coli*". *J. Bacteriol.* **161**: 1059-1068.

Saier, M. H., Jr. (1996). "Cyclic AMP-independent catabolite repression in bacteria". *FEMS Microbiol. Lett.* **138**: 97-103.

- Saier, M. H. Jr. i T. M. Ramseier. (1996). "The catabolite repressor/activator (Cra) protein of enteric bacteria". *J. Bacteriol.* **178**: 3411-3417.
- Sallal, A.-K. J. i N. A. Nimer. (1989). "The intracellular localization of glycolate oxidoreductase in *Escherichia coli*". *FEBS Lett.* **258**: 277-280.
- Sambrook, J., Fritsch, E. F. i T. Maniatis. (1989). "Molecular cloning: a laboratory manual". 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
- Sandalova, T. i Y. Lindqvist. (1993). "Crystal structure of apo-glycolate oxidase". *FEBS Lett.* **327**: 361-365.
- Sanger, F., Nicklen, S. i A. R. Coulson. (1977). "DNA sequencing with chain-terminating inhibitors". *Proc. Natl. Acad. Sci. USA.* **174**: 5463-5467.
- Sawada, H. i Y. Takagi. (1964). "The metabolism of L-rhamnose in *Escherichia coli*. III. L-rhamnose-1-phosphate aldolase". *Biochim. Biophys. Acta.* **92**: 26-32.
- Scarabel, M., Penin, F., Bonod-Bidaud, C., Nègre, D., Cozzone, A. J. i J. C. Cortay. (1995). "Overproduction, purification and structural characterization of the functional N-terminal DNA-binding domain of the *fru* repressor from *Escherichia coli* K-12". *Gene.* **153**: 9-15.
- Schröder, I., Darie, S and R. P. Gunsalus. (1993). "Activation of the *Escherichia coli* nitrate reductase (*narGHJI*) operon by NarL and Fnr requires integration host factor". *J. Biol. Chem.* **268**: 771-774.
- Schryvers, A. i J. H. Weiner. (1981). "The anaerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli*". *J. Bacteriol.* **256**: 9959-9965.
- Schwam, H., Michelson, S., Randall, W.C., Sondey, J.M. i R. Hirschmann. (1979). "Purification and characterization of human liver glycolate oxidase. Molecular weight, subunit and kinetic properties". *Biochem.* **18**: 2828-2833.
- Seip, J. E., Fager, S. K., Gavagam, J. E., Gosser, L. W., Anton, D. L. i R. DiCosimo. (1993) "Biocatalytic production of glycolic acid". *J. Org. Chem.* **58**: 2253-2259.
- Shapiro, L. (1993). "Protein localization and assymetry in the bacterial cell". *Cell.* **73**: 841-855.
- Shapiro, L., Kaiser, D and R. Losick. (1993). "Development and behavior in bacteria". *Cell.* **73**: 835-836.

- Shen, J. i R. P. Gunsalus. (1997). "Role of multiple ArcA recognition sites in anaerobic regulation of succinate dehydrogenase (*sdhCDAB*) gene expression in *Escherichia coli*". *Mol. Microbiol.* **26**: 223-236.
- Smith, M. W. i F. C. Neidhart. (1983). "Proteins induced by aerobiosis in *Escherichia coli*". *J. Bacteriol.* **154**: 344-350.
- Spurio, R., Falconi, M., Brandi, A., Pon, C. L. i G. O. Gualerzi. (1997). "The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending". *EMBO J.* **16**: 1795-1805.
- Sridhara, S. i T. T. Wu. (1969). "Purification and properties of lactaldehyde dehydrogenase from *Escherichia coli*". *J. Biol. Chem.* **244**: 5233-5238.
- Stewart, V. (1988). "Nitrate respiration in relation to facultative metabolism in enterobacteria". *Microbiol. Rev.* **52**: 190-232.
- Stoker, N. G., Pratt, J. M. i H. B. Holland. (1984). "In vivo gene expression systems in prokaryotes". p. 154-172 in *Transcription and translation: a practical approach*. ed. B. D. Hames and S. H. Higgins. IRL Press Ltd., Oxford.
- Sunnarborg, A., Klumpp, D., Chung, T. i D. C. LaPorte. (1990). "Regulation of the glyoxylate bypass operon: cloning and characterization of *iclR*". *J. Bacteriol.* **172**: 2642: 2649.
- Takagi, Y. i H. Sawada. (1964). "The metabolism of L-rhamnose in *Escherichia coli*. I. L-rhamnose isomerase". *Biochim. Biophys. Acta.* **92**: 10-17.
- Takagi, Y. i H. Sawada. (1964a). "The metabolism of L-rhamnose in *Escherichia coli*. II. L-rhamnose kinase". *Biochim. Biophys. Acta.* **93**: 18-25.
- Tanaka, K., Haniu, M., Yasunobu, K. T., Himes, R. H. i J. M. Akagi. (1973). "The primary structure of the *Clostridium thermosaccharolyticum* ferredoxin, a heat stable ferredoxin". *J. Biol. Chem.* **248**: 2215-2217.
- Tanaka, K., Mikame, E. i T. Suzuki. (1986). "Methane fermentation of 2-methoxyethanol by mesophilic digesting sludge". *J. Ferment. Technol.* **64**: 305-309.
- Tardat, B. i D. Touati. (1993). "Iron and oxygen regulation of *Escherichia coli* MnSOD expression: competition between global regulators Fur and ArcA for binding to DNA". *Mol. Microbiol.* **9**: 53-63.

Tsunehiro, A. i S. Adhya. "Repressor induced site-specific binding of HU for transcriptional regulation" (1997). *EMBO J.* **16**: 3666-3674.

Vanderwinkel, E. i M. DeVlieghere. (1968). "Physiologie et génétique de l'isocitrate et des malate synthases chez *Escherichia coli*". *Eur. J. Biochem.* **5**: 81-90.

Vogels, G. D. i C. Van der Drift. (1976). "Degradation of purines and pyrimidines by microorganisms". *Bacteriol. Rev.* **40**: 403-468.

Webster, C., Gaston, K. i S. Busby. (1988). "Transcription from the *Escherichia coli melR* promoter is dependent on the cyclic AMP receptor protein". *Gene.* **68**: 297-305.

Weickert, M. J. i S. Adhya. (1992). "A family of bacterial regulators homologous to Gal and Lac repressors". *J. Biol. Chem.* **267**: 15869-15874.

Weiner, J. H. i P. Dickie. (1979). "Fumarate reductase of *Escherichia coli*. Elucidation of the covalent-flavin component". *J. Biol. Chem.* **254**: 8590-8593.

Werner, M. H. i S. K. Burley. (1997). "Architectural transcription factors: proteins that remodel DNA". *Cell.* **88**: 733-736.

Winans, S. C., Elledge, S. J., Krueger, J. H. i G. C. Waljer. (1985). "Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*". *J. Bacteriol.* **161**: 1219-1221.

Wood, D., Darlinson, M., Wilde, R. i J. R. Guest. (1984). "Nucleotide sequence encoding the flavoprotein and hydrophobic subunits of the succinate dehydrogenase of *Escherichia coli*". *Biochem. J.* **222**: 519-534.

Wu, Y. i P. Datta. (1992). "Integration host factor is required for positive regulation of the *tdc* operon of *Escherichia coli*". *J. Bacteriol.* **174**: 133-140.

Yuan, H. S., Finkel, S. E., Feng, J. A., Kaczor-Grzeskowiak, M., Johnson, R. C. i R. E. Dickerson. (1991). "The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding". *Proc. Natl. Acad. Sci. USA.* **88**: 9558-9562.

Zabarovsky, E. R. and G. Winberg. (1990). "High efficiency electroporation of ligated DNA into bacteria". *Nucleic Acids Res.* **18**: 5912.

Zarebinski, C. M., Hoyt, J. C. i H. C. Reeves. (1991). "Properties of isocitrate lyase from *Escherichia coli* K12 grown on acetate or glycolate". *Curr. Microbiol.* **22**: 65-68.

Zhu, Y. i E. C. C. Lin. (1986). "An envolvant of *Escherichia coli* that employ L-fucose pathway also for growth on L-galactose and D-arabinose". *J. Mol. Evol.* **23**: 259-266.

Zulianello, L., van Ulsen, P., van de Putte, P. i N. Goosen. (1995). "Participation of the flank regions of the integration host factor protein in the specificity and stability af DNA bending". *J. Biol. Chem.* **270**: 17902-17907.

Molecular characterization of *Escherichia coli* malate synthase G Differentiation with the malate synthase A isoenzyme

Isabel MOLINA, María-Teresa PELLICER, Josefa BADIA, Juan AGUILAR and Laura BALDOMA
Department of Biochemistry, School of Pharmacy, University of Barcelona, Spain

(Received April 18/June 13, 1994) – EJB 94 0548/4

Two genes encoding the enzymes malate synthase G and glycolate oxidase, have been linked to locus *glc* (64.5 min), responsible for glycolate utilization in *Escherichia coli*. The gene encoding malate synthase G, for which we propose the notation *glcB*, has been cloned, sequenced and found to correspond to a 2262-nucleotide open-reading frame, which can encode a 723-amino-acid polypeptide, clearly different from the isoenzyme malate synthase A, which has 533 amino acids. Northern-blot experiments indicate that *glcB* was expressed as an apparently monocistronic transcript, inducible by glycolate. Malate synthase G was purified to near homogeneity. The molecular mass determined by gel filtration yielded a value of 82 kDa for the purified enzyme and the same value as for the crude extract enzyme, indicating a monomeric structure. Despite the lower sequence similarity between malate synthase G and the other reported malate synthases, three out of nine consensus boxes defined in most of these enzymes are conserved in addition to a cysteine residue that has been reported to be important for the catalytic mechanisms.

Two isoenzymes of malate synthase have been described for the metabolism of glyoxylate in *Escherichia coli*. Malate synthase A (MSA), involved in the glyoxylate by-pass reaction, is encoded by the gene *aceB* belonging to the *ace* operon located at 91 min on the genetic map [1], and is predominant (60% total activity) in cells metabolizing the glyoxylate formed in the dissimilation of acetate [2]. The other isoenzyme, malate synthase G (MSG), encoded by a gene mapped in the locus *glc*, accounts for almost the entire malate-synthesizing activity in cells metabolizing the glyoxylate formed during growth on glycolate [2]. Both isoenzymes catalyze the condensation of glyoxylate and acetyl-CoA to yield malate, display very similar kinetic parameters for the substrate [3], and are distinguished on the basis of their chromatographic separation, thermostability and sensitivity to inhibitors such as glycolate and oxalate [3].

In glycolate metabolism, this compound is first oxidized to glyoxylate by the action of glycolate oxidase [4, 5]. Glyoxylate is subsequently metabolized by two divergent condensation reactions; one reaction is catalyzed by glyoxylate carboligase, which converts two glyoxylate molecules to tartronic semialdehyde and carbon dioxide [6], and the other reaction is catalyzed by MSG. The latter process can be considered a reaction of an oxidative pathway if glyoxylate is metabolized through the dicarboxylic cycle, or a reaction of

an anaplerotic route if the citric acid cycle is involved in the oxidation of glyoxylate [7].

While the molecular organization and regulation of the gene *aceB* have been extensively studied [1], nothing has been reported about the gene encoding MSG in *E. coli*. In this study we report the cloning and DNA sequencing of the gene encoding MSG and the molecular characterization of the purified enzyme.

MATERIALS AND METHODS

Bacterial strains and growth conditions

All strains were derivatives of *E. coli* K12. Strain W3110, used by Kohara et al. [8] to prepare the library from which *glcB* was cloned in this study, was also used as a source of MSG in the purification procedure. Strain DV021A05 containing a deletion *speC-glc* [2] was made *ppc*⁺ by transduction of this marker from the wild type strain and selection of transductants on glucose (strain JA150). Strain JA151 was obtained by transduction of the *recA* marker from strain BW11656 [9] into strain JA150 and selection of transductants on Luria-Bertani medium/chloramphenicol plates. Two glycolate-negative mutant strains (JA152, JA153) were derived from the *E. coli* strain CGSC 4833. Strain JA152 was obtained by mutagenesis with ethyl methane sulfonate, selection on glucose plates and replica plating on glycolate. Strain JA153 was obtained by insertion of Tn5, selection on Luria-Bertani medium/kanamycin plates and replica plating on glycolate. Both strains retained the ability to grow on glyoxylate. Strain S ϕ 1023 (*nupG::Tn10*) and strain NK6027 (*metC::Tn10*) [10] were used in transduction experiments for the linkage of their markers to the *glc* locus.

Cells were grown aerobically as described previously [11] on Luria-Bertani medium or minimal medium. In the latter case, carbon sources were added at the following con-

Correspondence to L. Baldomá, Department of Biochemistry, School of Pharmacy, University of Barcelona, Av. Diagonal 643, E-08028 Barcelona, Spain

Fax: +34 3 402 1896.

Abbreviations. MSA, malate synthase A; MSG, malate synthase G.

Enzymes. Malate synthase (EC 4.1.3.2); glycolate oxidase (EC 1.1.3.15); glyoxylate carboligase (EC 4.1.1.47).

Note. The nucleotide sequence reported in this paper has been submitted to the EMBL/GenBank sequence data bank with accession number X74547.

centrations: 0.5% casein acid hydrolysate, or 1% in cultures of transformed cells; 10 mM glucose; 30 mM glycolate; 30 mM glyoxylate. For solid medium cultures, 15 g/l agar was added. Where indicated, kanamycin to 25 µg/ml, tetracycline to 15 µg/ml, chloramphenicol to 50 µg/ml, or ampicillin to 100 µg/ml were added.

Preparation of cell extracts and enzyme activities

Malate-synthase activity was measured in extracts of cells grown with the indicated conditions and harvested at the end of the exponential phase. Cell extracts were prepared as described previously [12] in 10 mM Tris/HCl, pH 8.0, containing 1 mM MgCl₂. Determination of total activity and the relative concentrations of the two forms of malate synthase, A and G, were performed as described by Ornston and Ornston [7].

Glycolate oxidase activity was determined spectrophotometrically according to the method of Lord [5]. The 30% ammonium sulfate fraction was routinely used as source of enzyme to avoid non-specific reduction of 2,6-dichloroindophenol.

Enzyme purification and amino acid sequence analysis

MSG was purified from strain W3110 grown on 5 mM succinate in the presence of 30 mM glycolate as inducer. The procedure was carried out at 4°C and consisted of DEAE-Sephrose chromatography and heat inactivation in the presence of oxalate, followed by hydroxylapatite chromatography. The first two steps were adapted from those described by Falmagne et al. [3].

The crude extract was dialyzed against 10 mM Tris/HCl, pH 8.0, containing 1 mM MgCl₂ (buffer A), and was loaded onto a DEAE-Sephrose (Pharmacia) column equilibrated with the same buffer. The column was first washed with buffer A and eluted with a linear gradient (1–100 mM) of MgCl₂ in 10 mM Tris/HCl, pH 8.0. Fractions containing most of the MSG activity were pooled and diluted to 2.5 mg/ml in 50 mM Tris/HCl, pH 8.0, containing 10 mM MgCl₂ (buffer B). To prevent MSG inactivation during heat treatment, 10 mM oxalate was added. After incubation of the sample at 61°C for 20 min, 15 mM CaCl₂ was added to precipitate oxalate, which was removed by centrifugation at 20000 g for 15 min. MSG activity was recovered in the supernatant fraction. This sample was dialyzed against buffer A and applied to a hydroxylapatite column (Bio-Gel HTP, BioRad) equilibrated with the same buffer. The column was first washed with buffer A containing 400 mM MgCl₂, and was eluted with a linear gradient of MgCl₂ (400–1500 mM) in buffer A. Fractions containing malate-synthase activity were pooled, concentrated by ultrafiltration and the excess MgCl₂ was removed by dialysis against buffer A.

This preparation was analysed by 7.5% SDS/PAGE in the presence of sodium thioglycolate in the top reservoir to prevent amino acid destruction. An aliquot of the enzyme was treated with dansyl chloride [13, 14], run in parallel and used to identify the protein band corresponding to MSG. This was cut from the gel and the protein was electroeluted in an electrophoretic sample concentrator using 10 mM ammonium bicarbonate, pH 8.0.

The electroeluted enzyme solution was concentrated in a vacuum chamber and the N-terminal sequence was deter-

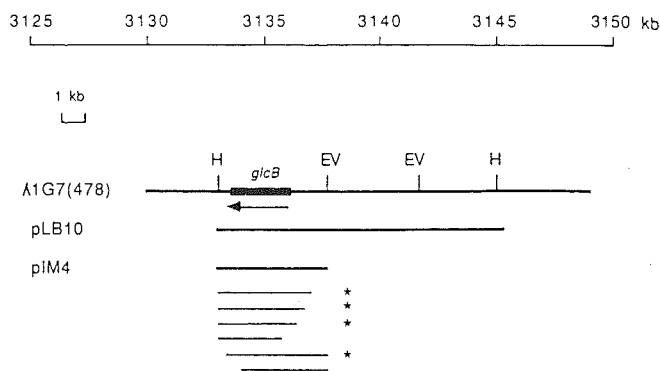


Fig. 1. Physical map of the genomic region containing the *glcB* gene. Restriction map of *E. coli* genomic region containing the *glc* locus (3125–3150 kb) carried by the phage λ1G7 clone of Kohara et al. [8] (top two lines). The shaded bar is the *glcB* gene and the arrow indicates the direction of transcription. The restriction enzymes are as follows: EV, *EcoRV*; H, *HindIII*. The restriction fragments used in the construction of plasmids in this study are shown in thick lines below the restriction map. Thin lines correspond to the fragments of the subclones constructed by deletion of the pIM4 insert. The fragments complementing MSG activity are marked (*).

mined by automated Edman degradation with an Applied Biosystems model 470A gas-phase sequenator.

Molecular-mass analysis

The molecular mass of the subunit was determined by SDS/PAGE as described by Laemmli [15], using the molecular mass markers indicated in Fig. 4.

The molecular mass of the native enzyme was determined on an Ultrogel ACA-34 calibrated column (1.5 cm × 95 cm) equilibrated and eluted with buffer A at a rate of 15 ml/h. Calibration was performed using the proteins indicated in Fig. 5. Alternatively, buffer A containing 50 mM NaCl was used to analyze possible aggregation or dissociation effects on the enzyme structure depending on buffer ionic strength.

Gene transfers, DNA manipulation and plasmid construction

Phage P1 transduction experiments were performed essentially as described by Miller [16]. Plasmid DNA was routinely prepared by the boiling method [17]. For large scale preparation, a crude DNA sample was subjected to purification through a column (Qiagen GmbH).

DNA manipulations were performed essentially as described by Sambrook et al. [18]. The DNA sequence was determined by the dideoxynucleotide-chain termination procedure of Sanger et al. [19]. Double-stranded plasmid DNA was used as the template. Plasmid purified in a Qiagen column was used for the construction of ordered deletions with Erase-a-Base (Promega). Universal or T3 primers were used except at two problem regions at approximately 320 bp and 1650 bp where oligonucleotides VT1C (5' TCAAACCA-TAACCCAGAG 3'), VT2C (5' GCCCGTGGTTTCCACCGT 3'), VT3C (5' GTTCATCAATACCGGTTT 3') and VT4C (5' ACAGAACAGACCGGAAAG 3') were chemically synthesized and used as the primers. To resolve sequencing-gel compressions, 7-deaza-dGTP and 7-deaza-dITP were used instead of dGTP.


```

          30                               60                               90
CCCAGGAGAAAGCGCGTACCGCCGCGCTGGGGCGCTCGTGAAACTAAGGCGTATGAAGAGATGGTGAACAACGGACGTACCGCGTTCGTGA
          120                               150                               180
CTGCGCCGTTATTAACGTGCGTGAAGAGCGCGGTACCGGTTGTTGGATGGGCAAATTATTGGTCCGCTGGGCGTTTCTGGTTAAACCG
          210                               240                               270
GAGCACAGGATGCGCAGGTTCGCGAAAGCGCAGCAGCGGTGTTGGCGAAATAAGCGAAAAACGAGGATATAACAATGAGTCAAACCATAAC
          300                               330                               360
CCAGAGCCGTTTACGCATTGACGCCAATTTTAAACGTTTGTGGATGAAGAAGTTTACCGGGAACAGGGCTGGACGCTGCGGCGTTCGT
Q S R L R I D A N F K R F V D E E V L P G T G L D A A A F W
          390                               420                               450
GGCCAATTTTGATGAGATCGTTCATGATCTGGCACCAGAAAATCGTCAGTTGGCTGGCAGAACCGGATCGCATTCAGGCAGCGCTTGATGA
R N F D E I V H D L A P E N R Q L L A E R D R I Q A A L D E
          480                               510                               540
GTGGCATCGCAGCAATCCGGGGCCGTAAGATAAAGCGGCGCTATAAATCTTTCCTGCGTGAACTGGGCTACCTGGTGCGCAACCGGA
W H R S N P G P V K D K A A Y K S F L R E L G Y L V P Q P E
          570                               600                               630
CGCGTGACCGTGGAAACCACCGGCATTGACAGCGAAATCACCAGCCAGGCGGGCCGACGCTGGTGGTTCGGCAATGAACCGCCGTA
R V T V E T T G I D S E I T S Q A G P Q L V V P A M N A R Y
          660                               690                               720
CGCGCTGAACCGCGGAACGCTCGCTGGGGCTCACTGTACGATGCGTTATACGGCAGCGACATCATCCCGCAGGAAGGGCGATGGTCAG
A L N A A N A R W G S L Y D A L Y G S D I I P Q E G A M V S
          750                               780                               810
CGGCTACGATCCGCAACCGGTGAGCAGGTTATCGCCTGGGTTCCGCGTTTCCTCGATGAATCTCTACCGCTGGAAAACCGCAGCTATCA
G Y D P Q R G E Q V I A W V R R F L D E S L P L E N G S Y Q
          840                               870                               900
GGATGTGGTGGCGTTAAAGTGGTTGATAAACAATTACGCATCCAGTTGAAAAATGGTAAAGAAACCAGGTTACGTACTCCAGCACAGTT
D V V A F K V V D K Q L R I Q L K N G K E T T L R T P A Q F
          930                               960                               990
TGTCGGTTACCGTGGCGATGCCCGCTGCCGACCTGCATTTTCTGTAATAAATACCGCCTGCATATTGAGCTGCAAATCGATGCCAATGG
V G Y R G D A A A P T C I L L K N N G L H I E L Q I D A N G
          1020                               1050                               1080
CGCGATTGGCAAAGACGATCCGGCGCACATCAACGATGTTATCGTCGAAGCTGTATCAGTACCATTCTCGACTGCGAAGATTCCGGTCCG
R I G K D D P A H I N D V I V E A A I S T I L D C E D S V A
          1110                               1140                               1170
GGCGTTGATGCGGAAGATAAAATCCTGCTGACCGCAACCTGCTGGGCTGATGCAGGGGACTCTGCAAGAGAAAAATGGAGAAAAACGG
A V D A E D K I L L Y R N L L G L M Q G T L Q E K M E K N G
          1200                               1230                               1260
TCGGCAATCGTGCATAAAGTAACTGACGATCGTCATTACACCGCCCGCGATGGCTCTGAAATTTCTCTGCACGGACGCTCGCTGCTGTT
R Q I V R K L N D D R H Y T A A D G S E I S L H G R S L L F
          1290                               1320                               1350
TATCCGCAACGTTGGGTCATTTGATGACCAATCCTGTGATTTGGGACAGCGAAGCAATGAAATCCCGGAAGGCATTCTTATGGCGTCAT
I R N V G H L M T I P V I W D S E G N E I P E G I L D G V M
          1380                               1410                               1440
GACTGGCGGATGCCCCTATGATTTAAAAGTGCAGAAAACTCGCGCACTGGCAGCGTCTATATTGTGAAACCGAAAATGCACGGTCC
T G A I A L Y D L K V Q K N S R T G S V Y I V K P K M H G P
          1470                               1500                               1530
CGAGGAAGTGGCGTTCGCCAACAACATGTTACCGCATTGAGACAAATGCTCGGTATGGCACCAGCAATCCCTGAAAATGGGCATTATGGA
Q E V A F A N K L F T R I E T M L G M A P N T L K M G I M D

```

Fig. 2. Nucleotide and deduced amino acid sequences of the *glcB* gene. The coding region of *glcB* has been translated and is indicated by the single-letter amino acid code. Stop codons are indicated (*). The twelve amino acid residues of the N-terminus, sequenced from the purified MSG protein, are in bold. The Shine-Dalgarno sequence is boxed. The inverted repeat in the 5' flanking region constituting a putative control element in the regulation of *glcB* expression, and that constituting a transcription terminator element are underlined.

RNA preparations and Northern-blot experiments

For preparation of total RNA, cells of a 25-ml culture grown to A_{650} 0.5, were collected by centrifugation at 5000 g for 10 min and processed as described by Belasco et al. [20]. Northern-blot hybridization was performed with each RNA sample (10 µg) following the procedure described previously [21].

RESULTS

Genetic and physical map of the *glc* locus

Cotransduction of glycolate utilization with Tn10 insertions in genes *nupG* (64.3 min) or *metC* (65 min), encom-

passing locus *glc*, indicated that both MSG and glycolate oxidase activities were transduced in the process. This was shown by experiments in which phages grown on strain Sφ1023 or NK6027 were used to infect strain JA150, a glycolate-negative strain containing a deletion *speC-glc* (64.5–65 min). Transductants were selected for tetracycline resistance and checked for glycolate utilization. Several glycolate-positive transductants were analyzed and were found to have recovered both activities. This cotransduction pattern (20% for *nupG* and 32% for *metC*) was also similar for point and insertion mutants, strains JA152 and JA153, respectively, which were unable to use glycolate.

The transduction of the two structural genes was confirmed by complementation experiments. Several phages of

```

1560                               1590                               1620
TGAAGAACGTCGGACCTCGCTGAACTTGGGTAGCTGTATCGCTCAGGCGCGCAACCGGTCGCGTTCATCAATACCGGTTTCCTCGACCG
E E R R T S L N L R S C I A Q A R N R V A F I N T G F L D R

1650                               1680                               1710
TACCGCGCATGAAATGCATTTCGGTGTAGGAAAGCTGGCCCGATGCTGCGTAAAAATCAGATGAAATCGACGCGCTTGGATCAAAGCCTACGA
T G D E M H S V M E A G P M L R K N Q M K S T P W I K A Y E

1740                               1770                               1800
GCGTAATAACGTCGCTTTCCGGTCTGTTCTGTTGGGCTGCGCGGTAAGCGCAAAATGTAAGGCATGTGGGCAATCCCGGACCTGTGGC
R N N V L S G L F C G L R G K A Q I G K G M W A M P D L M A

1830                               1860                               1890
AGACATGTACAGCCAGAAGGGCGCACTGCGTGCCTGGGCAACACAGCCTGGGTTCCCGTCACCAACCGCTGCTACGCTCCATGCGCT
D M Y S Q K G D Q L R A G A N T A W V P S P T A A T L H A L

1920                               1950                               1980
GCCTACCACCAAAACCAAGCTACAGAGCGTACAAGCCAACTTCCCGAGCCGAGTTCAATGCTGAATTTGAACCGCTGCTGGACGATCT
H Y H Q T N V Q S V Q A N I A Q T E F N A E F E P L L D D L

2010                               2040                               2070
GCTGACTATTCCGGTGTGTAAGAAAGCTAAGTGGTGGCGCAAGAGATCCAAACAAGCTGGATAACAACGTCAGGGGATTCTGGGGTA
L T I P V A E N A N W S A Q E I Q Q E L D N N V Q G I L G Y

2100                               2130                               2160
CGTGGTGGCGCTGGGTGGAGCAGGGGATTGGTTGTTCAAAAAGTGGCCGATATTCACAATGTGGCGTTGATGGAAGACCGCGCAACGCTGGC
V V R W V E Q G I G C S K V P D I H N V A L M E D R A T L R

2190                               2220                               2250
TATCTCCAGCCAGCATATCGCCAACCTGGTTACGTCACGGTATTCTGACCAAAGAACAGGTGCAGGCGCTGCTGGAGAATATGGCGAAAGT
I S S Q H I A N W L R H G I L T K E Q V Q A S L E N M A K V

2280                               2310                               2340
GGTTGATCAGCAAAACGCTGGCGATCCGGCTTATCGTCCGATGGCGGGGAATTTCCGCTAAGCTCGTGTGCTTTTAAAGTGGCAGCGATTT
V D Q Q N A G D P A Y R P M A G N F A N S C A F K A A S D L

2370                               2400                               2430
AATCTTCTCGCGTGAAACAGCCAAACGGCTATACCGAACCGTTATTACACGCGCTGGCGTTTACCGGAAAAGAAAGTCATTAATAGCG
I F L G V K Q P N G Y T E P L L H A W R L R E K E S H * *

2460                               2490                               2520
TCAAAAAGCCCCGGCAAGTTTATTGTCGGGGCTATGATAGTAAGTGCAGTACATTATCCCGGAAATAACCGGAGCGATTATCAG

2550                               2580                               2610
GGTTTAAACAATATTAATAACCGTAGCGCATTACTAGCGGGCCTAAAACAATATGGACATCTTAACGATACTCTGCGTCTGCGTGTG
CGGAACAGA

```

Fig. 2. (Continued).

Table 1. Enzymic activities linked to the *glc* locus. Growth of strains, preparation of crude extracts and method of assay are described in the Materials and Methods section. Glycolate oxidase activity was determined in the 30% ammonium sulfate fraction and expressed in relation to the amount of protein in the crude extract. MSG activity was determined in the cell extracts of strain W3110 after heat inactivation of MSA. CAA, casein acid hydrolysate.

Strain	Growth medium	Enzyme activities for	
		glycolate oxidase	malate synthase G
		nmol · min ⁻¹ · mg ⁻¹	nmol · min ⁻¹ · mg ⁻¹
W3110	CAA	<0.001	<0.001
W3110	CAA + glycolate	50	160
JA151 (pUC18)	CAA + glycolate	<0.001	<0.001
JA151 (pLB10)	CAA + glycolate	1050	3200

the Kohara collection [8] encompassing the markers *nupG* and *metC* were digested with *HindIII*, the resulting fragments were cloned in pUC18 and the recombinant plasmids were screened for the capacity to complement strain JA151. This strain did not show detectable activities of glycolate oxidase or MSG, due to the deletion of locus *glc*. In addition, it did not contribute to malate-synthase activity from the *aceB* gene product. Of all the clones tested, only pLB10 containing the 11.5 kb *HindIII* fragment derived from phage λ 1G7 (Fig. 1)

was shown to complement both activities in transformants grown on medium containing casein acid hydrolysate, glycolate and ampicillin. Analysis of extracts of the transformed strain indicated that pLB10 encoded proteins with the two activities at much higher levels (20-fold) than those determined in glycolate-grown cells that did not bear pLB10 (Table 1). These data indicate plasmid-mediated expression of both genes and hence the location of both structural genes in the corresponding insert.

A more precise location of the MSG gene, for which we propose the notation *glcB*, was obtained by subcloning the 4.5-kb *HindIII*-*EcoRV* fragment into Bluescript (Fig. 1) and the resulting plasmid pIM4 was used to complement the MSG activity of strain JA151. Transformants were grown on casein acid hydrolysate in the presence of glycolate. Only MSG activity was recovered, this time with no glycolate oxidase present in the extracts. A series of deletions of plasmid pIM4 allowed us to determine the limits of *glcB* by performing the MSG complementation test with the plasmids containing some of the different deleted inserts (Fig. 1). In this way, the gene was located between 3133.5 kb and 3136.2 kb in the Kohara chromosomal map.

Sequence of *glcB*

A set of 30 ordered deletions was obtained and used to sequence both strands of the suggested coding region and its 5' and 3' flanking regions. 2617 base pairs were sequenced, in which a single open-reading frame of 2262 bp with a G + C content of 53.3% was found (Fig. 2). Thus, *glcB* corre-

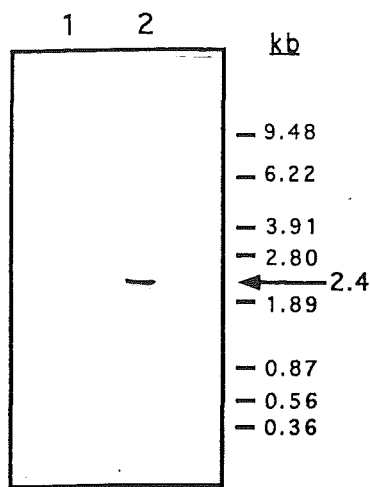


Fig. 3. Northern-blot analysis of *glcB* mRNA. Total RNAs (10 μ g) isolated from cultures of strain W3110 grown on casein acid hydrolysate (lane 1) or on the same medium but in the presence of glycolate as inducer (lane 2) were separated by electrophoresis, blotted and hybridized with the 0.4-kb *Ava*II–*Cla*I DNA probe indicated in the text. The size of standards, and the *glcB* transcript are indicated.

sponds to this open-reading frame, which can encode a 723 amino acid polypeptide with a calculated molecular mass of 80488 Da. The *glcB* open-reading frame gave a score for the frequency of optimal codon usage calculated according to Ikemura [22] of 0.71. As shown in Fig. 2, no promoter consensus box was apparent upstream from *glcB*, while a potential ribosomal recognition consensus sequence was present between positions –8 and –12 from the ATG start codon. An inverted repeat was also found upstream of *glcB* which could be involved in transcription regulation. There was an 11-bp inverted repeat separated by 6-bp after the TAA translational stop codon, which could form a stable stem-loop structure with a calculated free energy of stabilization of –22.8 kcal/mol (approximately –95.3 kJ). This structure is likely to correspond to a transcriptional termination element.

Transcription

Total RNA from cells of strain W3110 grown under non-inducing (casein acid hydrolysate) or inducing (casein acid hydrolysate plus glycolate) conditions was prepared as indicated above. Northern-blot hybridizations performed with a probe of 0.4 kb (*Ava*II–*Cla*I fragment from the *glcB* open-reading frame) showed a transcript of 2.4 kb in the RNA preparations obtained from induced cells, while no band of hybridization was detected when RNA preparations of non-induced cells were used (Fig. 3).

Purification, molecular mass and subunit composition of malate synthase G

MSG, induced by growth on glycolate, was purified following the procedure indicated in the Materials and Methods section. Although the activity is almost entirely represented by MSG in these conditions, a temperature-inactivation step was introduced to eliminate the residual MSA. The process permitted a 100-fold purification with respect to the crude extract, with a yield of 43%, giving a purified enzyme

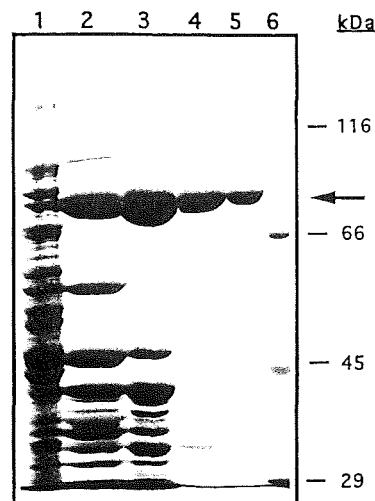


Fig. 4. SDS/PAGE analysis of steps in the purification of malate synthase G. Electrophoresis was performed on a 7.5% acrylamide gel. Lane 1, crude extract; lane 2, pooled DEAE–Sephacel column fractions; lane 3, heat treated supernatant; lanes 4 and 5, 40 μ g and 20 μ g, respectively, of pooled hydroxylapatite column fractions; lane 6, molecular-mass standards. The position of MSG is indicated (\leftarrow). Staining was with Coomassie brilliant blue.

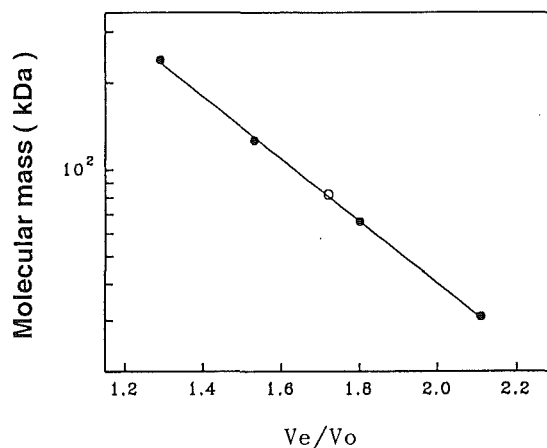


Fig. 5. Determination of the molecular mass of native malate synthase G by gel-filtration chromatography with an Ultrogel AcA-34 column. The molecular masses of various standard proteins (●) are as follows: carbonic anhydrase (31 kDa), bovine serum albumin (66 kDa), yeast alcohol dehydrogenase (126 kDa) and catalase (240 kDa). Values have been plotted against their V_e/V_o (relative elution volume) for the determination of the molecular mass of the native MSG enzyme. The corresponding interpolated value is indicated (○).

with a specific activity of 7.0 U/mg protein. Purification of the enzyme was monitored not only by measuring the enzyme activity but also by PAGE (Fig. 4).

The purity of this sample was determined by analyzing 20 μ g or 40 μ g purified protein by PAGE under denaturing conditions. As shown in Fig. 4, the lanes overloaded with the specific malate synthase presented an insignificant amount of contaminants. The mobility of the protein under these denaturing conditions, as compared to the markers indicated in Fig. 4, gave a molecular mass for this protein of 82 kDa.

The molecular mass of the native enzyme was estimated by gel filtration. 2 mg purified MSG was applied to the Ultrogel AcA34 column calibrated with the indicated marker

BRNMS	MELETSVYRPNVAVDSPDGEVVRGRYDQVFAKILTRDALGF---VAELQREFRG	52
ECMSA	MTEQAT-----TTDELAFTRPYGEQEKQILTAEAVEF---LTELVTHTFP	42
SCMS1	M---VKVSLDNVLLVDVDKEPFFKPSSTTVGDILTKDALEF---IVLLHRFTNN	49
ECMSG	MSQTTIQSRLRI-----DANFKRFVDEEVLPGTGLDAAAFAWRNFDEIVHDLAP	48
	* * *	
BRNMS	HVRYAMECRREVKRRYN---SGAVP-----GFDPSTKFIR-DGEWVCASVP	94
ECMSA	QRNKLLAARIQQQQDID---NGTLP-----DFISETASIR-DADWKIRGIP	84
SCMS1	KRKQLLENRQVVQKKLD---SGSYHL-----DFLPETANIRNDPTWQGPILA	93
ECMSG	ENRQLLAERDRIQAALDEWHRSNPGPVKDKAAYSFLRELGYLVPQPER---VT	99
 * *	
	1 2	
BRNMS	PAVADRRVEITGFVERKMIINALNSG-----AKVFMADFEDALSPSWENL	139
ECMSA	ADLEDRRVEITGFVERKMIINALNSN-----VKVFMADFEDSLAPDWNKV	129
SCMS1	PGLINRSTVEITGFPLRMLINALNAP-----VNTYMTDFEDSASPTWNNM	138
ECMSG	VETTGIDSEITTSQAGPQLVVPAMNARYALNANARWGSGLYDALYGSDIIPQEGAM	154
 * *	
	3	
BRNMS	MRG-----DQVNLKDAVDGS	153
ECMSA	IDG-----DINLRDAVNGT	143
SCMS1	VYG-----DQVNLDAIRNQ	152
ECMSG	VSGYDQPRGEQVIAWVRRFLDESPLENGSYQDVVAFKVVDKQLRIQLKNGKETT	209
	. * *	
BRNMS	-----ITFNDKARNKVYKL-----NDQVAKL---FV---	176
ECMSA	-----ISYTNEA-GKIYQL-----KPNPAVL---IC---	165
SCMS1	-----IDF--DTPRKSQYK-----NGNVANLPTIIV---	176
ECMSG	LRTPAQFVGYRGDAAPTCIILLKNNGLHIELQIDANGRIGKDDPAHINDVIVEAA	264
 *	
BRNMS	-----RPRGWHLPEAHIL	189
ECMSA	-----RVRGLHLPEKHVT	178
SCMS1	-----RPRGWHMVEKHLY	189
ECMSG	ISTILDCEDSVAAVDAEDKILLYRNLLGLMQGTLQEKMEKNGRQIVRKLNDDRHY	319
 *	
BRNMS	IDGEPATGCLVDFGLYFFHNYAKFR-----	214
ECMSA	WRGEAIPGSLFDFALYFFHNYQAL-----	202
SCMS1	VDDEPISASIFDFGLYFYHNAKELI-----	214
ECMSG	TAADGSEISLHGRSLLFIRNVGHLMTPVIWDSEGNEIPEGILDGVMGTGATALYD	374
 *	
	4 5	
BRNMS	---QTQSGFGPFYLPKMEHSREAKIWNVSF ERAEKMAGIERGSRATVLIETL	266
ECMSA	----LAKGSGPYFYLPKIQSWQEAAWSEVFSYAEDRFNLPRTGIKATLLIETL	252
SCMS1	---KL---GKGPYFYLPKMEHHLKAKLWVDVFCVAQDYIGIPRGTIRATVLIETL	263
ECMSG	LKVQKNSRTGSVYLVKPKMHGFQEVAFANKLFTRIETMLGMAPNTLKMGMIDDEER	429
 *	

Fig. 6. Multiple alignment of the amino acid sequences of different malate synthases. BRNMS, *Brassica napus* malate synthase; SCMS1, *S. cerevisiae* malate synthase 1; ECMSA, *E. coli* isoenzyme malate synthase A; ECMSG, *E. coli* isoenzyme malate synthase G. Identical residues among these proteins are indicated (*) and well-conserved positions are also shown (·). Boxes (1–9) correspond to the highly conserved regions found in different malate synthase proteins reported by Hikida et al. [23].

proteins and eluted with the two buffer systems described in the Materials and Methods section. The elution pattern was independent of the presence of NaCl in the buffer and gave a molecular mass for the native enzyme of 82 kDa, indicating that the enzyme is a monomer (Fig. 5).

To ascertain that the structure was not artifactually modified in the purification process, a crude extract sample was applied to the same gel filtration column and the elution pattern was monitored by measuring the MSG enzymic activity. The enzyme displayed the same relative elution volume (V_e/V_0) under these conditions, again indicating a monomeric structure for the crude extract sample.

N-terminal sequencing of purified malate synthase G

MSG, purified from *E. coli* K12 strain W3110 or from strain JA151 transformed with pIM4, was electrophoresed and electroeluted as indicated in the Materials and Methods section and the amino acid of the N-terminus was sequenced

up to residue 12 by automated Edman degradation. The enzyme of both origins gave the same sequence: S-Q-T-I-T-Q-S-R-L-R-I-D-, which matched exactly the deduced N-terminal sequence of the open-reading frame, and showed unambiguously that the cloned and sequenced gene encoded the MSG protein.

Amino acid sequence comparison with other malate synthases

The alignment of MSG with other malate synthases gave no significant sequence similarity but displayed conservation of several fragments. A comparison with MSA of *E. coli* gave a similarity of 30%, somewhat below the similarity of approximately 55% described by Hikida et al. [23] among several malate synthases of plants, fungi and bacteria. Alignment of *E. coli* MSG with *E. coli* MSA, yeast malate synthase-1 (*mIs1* gene product) and *Brassica napus* malate synthase was performed with the CLUSTAL V program [24]. As

		6	7	
BRNMS	PAVFQMNELLYELRDHSVGLNCGRWYIFSYVKTFAQHPDRLLPDRVVLVGMGQHF			321
ECMSA	PAVFQMDIILHALRDHIVGLNCGRWYIFSYIKTKLKNYPDRVLPDRQAVTMDKPF			307
SCMS1	PAAFQMEI IYQLRQHSSGLNCGRWYIFSTIKRLRNDPNHILPNINQVMTMSPF			318
ECMSG	RTSLNLRSCIAQARNRVAFINTG---FLDEITGDEMHSVMEAGPMLRKNQMKSTPW			481
	..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..*			
		8		
BRNMS	MRSYSDLLIRT---CHKRGVHAMGGMAADIPIRDDPKANEMALDLVKKDKLREVR			373
ECMSA	LNAYSRLLIK---CHKRGAFAMGGMAAFIPSKDEHNNQV-LNKVKADKLEAN			358
SCMS1	MDAYVKRLINT---CHRRGVHAMGGMAADIPIKDDPAANEKAMTKVRNDKIRELT			370
ECMSG	IKAYERNNVLSGLFCGLRGKAQIG-----KGMWAMPDLMADMSQKGDQLR			527
	..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..*			
BRNMS	AGHDGTWAAHPGLIPICMDAFSH--MGNNPNQIKSMKRDDASAITTEEDLLQ--IP			424
ECMSA	NGHDGTWIAHPGLADTAMAVFNDI-LGSRKNQLEVM-REQDAPITADQLLA--PC			409
SCMS1	NGHDGSWVAHPALAPICNEVFINM---GTPNQIYFI---PENVTAAANLLETKIP			419
ECMSG	AGANTAWVPSPTAATLHALHYHQTNVQSVQANIAQTEFNAEFELLDLLTI PVA			582
	..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..*			
		9		
BRNMS	R-----GVRTLEGLRLNTRVGIQYLAAWLT-GSGSVPLYN-----LMEDAATAEI			468
ECMSA	D-----GERTEEGMRANIRVAVQYIEAWIS-GNGCVPIYG-----LMEDAATAEI			453
SCMS1	N-----GEITTEGIVQNLDIGLQYMEAWLR-GSGCVPIINN-----LMEDAATAEV			463
ECMSG	ENANWSAQEIQQELDNNVQGLIGYVVRWVEQIGCSKVPDIHNVALMEDAATLRI			637
	..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..*			
BRNMS	SRVQNWQWIRYGVELDGDGLGVRVSKELFGRVVEEMERIEKEVKGDKFKRGMYS			523
ECMSA	SRTSIWQWIHHQKTLN---GKPVTKALFRQMLGEMKVIASELGEERFSQGRFD			505
SCMS1	SRCOLYQWVKHGVTLKDTG--EKVTPELTEKILKEQVERLSKASPLGD--KNKFA			514
ECMSG	SSQHIANWLRHGI-----LTKEQVQASLENMAKVVDDQONAGDPAYR			678
	..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..*			
BRNMS	EACKMFTKQCTAAELDDFLTLAV-----YDHIVAH--YPINASRL			561
ECMSA	DAARLMEQITTSDELIDFLTLPG-----YRLLA-----			533
SCMS1	LAAKYFLPEIRGEKFSFLTLTL-----YDEIVSTKATPTDLSKL			554
ECMSG	PMAGNFANSCAFKAASDLIFLGVKQPNGYTEPLLHAWRLREKESH			723
	..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..* ..*			

Fig. 6. (Continued).

shown in Fig. 6, this alignment displayed higher similarity at the N-terminal and C-terminal domains joined by a much less similar central region, most likely formed by the extra sequences in the longer polypeptide of MSG. Comparison of these sequences showed that, of the nine well-conserved boxes found in most malate synthases [23], three were also highly conserved (boxes 1, 2 and 9), four were partially conserved, and two were lost (boxes 7 and 8) in MSG. Multiple alignment analysis of most of the reported malate synthases performed with the PRETTY program (University of Wisconsin Genetics Computer Group package) gave the following consensus motives: EIT (in box 1), ALNA (in box 2) and LMDXAT (in box 9). Interestingly, Cys496 is also found in equivalent positions in all malate synthases.

DISCUSSION

Locus *glc* at 64.5 min of the bacterial chromosome had been described previously as giving mutations that abolished the phenotype of glycolate utilization [2]. Our experiments indicate that at least two genes encoding MSG and glycolate oxidase, enzymes responsible for glycolate utilization, lie in this locus.

The DNA fragment encoding MSG contains a single open-reading frame encoding a protein that is larger than usual (82 kDa) for most malate synthases so far described (62–64 kDa), including not only the MSA of *E. coli* but also the enzymes of other organisms such as yeast and plants. A larger malate synthase may also be found in *Saccharomyces cerevisiae* according to a transcript revealed under certain

specific conditions [25]. Several observations exclude the possibility that the assignment of the larger size for MSG was due to an artifact or sequencing error: (a) the enzyme protein purified from wild-type cells displayed a molecular mass in accordance with that calculated from the corresponding deduced amino acid sequence; (b) the peptide sequence of the N-terminal fragment indicates that the ATG initiation codon is correctly assigned. However, it should be emphasized that the starting methionine residue is removed in the cell processing of the polypeptide; (c) there is a highly conserved box at the C-terminal region of the polypeptide (amino acids 628–638), indicating that this final part of the sequence belongs to a malate-synthase structure.

The score for the frequency of optimal codon usage of 0.71 is consistent with the high expression of *glcB* under inducing conditions. In addition, this open-reading frame displays a highly conserved ribosome recognition sequence, which probably allows a good translation of the transcript. Furthermore, Northern-blot experiments indicate that the gene is expressed apparently as a single 2.4-kb transcript and that this expression occurs only in the presence of glycolate, which acts as an inducer. Nevertheless, additional information on the gene organization of locus *glc* and on the stability of its transcripts will be required to know if *glcB* is a single transcriptional unit. Finally, it may be pointed out that the inverted repeat downstream of the stop codons seems to fit better in the ρ -independent transcription terminator features, since it is mostly formed by G and C nucleotides and is followed by several U and A residues as described by d'Aubenton Caraffa [26]. Besides, it seems not to fulfill most of the requirements proposed for the ρ -dependent terminators

[27]. More information will be required to define the transcription organization of locus *glc*.

Determination of the relative molecular mass either by SDS/PAGE or gel filtration shows that the enzyme is a monomeric protein. This is in contrast to what has been described for other malate synthases such as the synthases from germinating soybean cotyledons (*Glycine max.* L.), which may be dimeric or decameric aggregates depending on the ionic strength [28], or the malate synthase from *S. cerevisiae*, which has been described as a homotrimeric protein [29]. Our determination of molecular mass of the crude extract enzyme shows that aggregation of subunits, which could be dissociated in the purification process (heat treatment, salt concentration) does not occur in our case.

Although MSG has a low similarity with other malate synthases, its sequence shows high conservation of three motifs found in most of the plant, bacteria and yeast enzymes, suggesting involvement of these motifs in the active center or other important regulatory features. In addition, a cysteine residue (position 496) is also found in all malate synthases, in agreement with the importance of sulfhydryl groups for the catalytic mechanism reported in several studies [23, 30]. A second cysteine residue (position 617) closer to the C-terminus is also conserved in *E. coli* and yeast malate synthases (Fig. 6), and is a conserved serine residue in plants [23].

This study was supported by grant PB91-437 from the *Dirección General de Investigación Científica y Técnica* of Spain. We thank Robin Rycroft for editorial assistance and *Servicio de secuenciación de proteínas, Universidad de Barcelona, Facultad de Farmacia*, for assistance in amino acid sequencing.

REFERENCES

- Chung, T., Klumpp, D. J. & LaPorte, D. C. (1988) Glyoxylate bypass operon of *Escherichia coli*. Cloning and determination of the functional map, *J. Bacteriol.* **170**, 386–392.
- Vanderwinkel, E. & De Vlieghere, M. (1968) Physiologie et génétique de l'isocitritase et des malate synthases chez *Escherichia coli*, *Eur. J. Biochem.* **5**, 81–90.
- Falmagne, P., Vanderwinkel, E. & Wiame, J. M. (1965) Mise en évidence de deux malate synthases chez *Escherichia coli*, *Biochim. Biophys. Acta* **99**, 246–258.
- Kornberg, H. L. & Sadler, J. R. (1961) The metabolism of C2 compounds in micro-organisms. 8. A dicarboxylic acid cycle as a route for the oxidation of glycolate by *Escherichia coli*, *Biochem. J.* **81**, 503–513.
- Lord, J. M. (1972) Glycolate oxidoreductase in *Escherichia coli*, *Biochim. Biophys. Acta* **267**, 227–237.
- Krakow, G., Barkulis, S. S. & Hayashi, J. A. (1961) Glyoxylic acid carboxylase. An enzyme present in glycolate-grown *Escherichia coli*, *J. Bacteriol.* **81**, 509–518.
- Ornston, L. N. & Ornston, M. K. (1969) Regulation of glyoxylate metabolism in *Escherichia coli* K12, *J. Bacteriol.* **98**, 1098–1108.
- Kohara, Y., Akiyama, K. & Isono, K. (1987) The physical map of the whole *Escherichia coli* chromosome. Application of a new strategy for rapid analysis and sorting of a large genomic library, *Cell* **50**, 495–501.
- Baldomá, L., Badía, J., Sweet, G. & Aguilar, J. (1990) Cloning, mapping and gene product identification of *rhaT* from *Escherichia coli* K12, *FEMS Microbiol. Lett.* **72**, 103–108.
- Singer, M., Baker, T. A., Schnitzler, G., Deischel, S. M., Goel, M., Dove, W., Jaacks, J. W. & Grossman, C. A. (1989) A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*, *Microbiol. Rev.* **53**, 1–24.
- Boronat, A. & Aguilar, J. (1979) Rhamnose induced propanediol oxidoreductase in *Escherichia coli*. Purification properties and comparison with the fucose-induced enzyme, *J. Bacteriol.* **140**, 320–326.
- Boronat, A. & Aguilar, J. (1981) Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: Differences in induction of propanediol oxidoreductase, *J. Bacteriol.* **147**, 181–185.
- Talbot, D. N. & Yphantis, D. A. (1971) Fluorescent monitoring of SDS gel electrophoresis, *Anal. Biochem.* **44**, 246–253.
- Tijssen, P. & Kurstak, E. (1979) A simple and sensitive method for the purification and peptide mapping of proteins solubilized from denonucleosis virus with SDS, *Anal. Biochem.* **99**, 97–104.
- Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head bacteriophage T4, *Nature* **224**, 680–685.
- Miller, J. H. (1972) *Experiments in molecular genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Holmes, D. S. & Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids, *Anal. Biochem.* **114**, 193–197.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd edn, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl Acad. Sci. USA* **74**, 5463–5467.
- Belasco, J. G., Beatty, J. T., Adams, C. W., von Gabain, A. & Cohen, S. N. (1985) Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic *rxvA* transcript, *Cell* **40**, 171–181.
- Moralejo, P., Egan, S. M., Hidalgo, E. & Aguilar, J. (1993) Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*, *J. Bacteriol.* **175**, 5585–5594.
- Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol.* **151**, 389–409.
- Hikida, M., Atomi, H., Fukuda, Y., Aoki, A., Hishida, T., Teranishi, Y., Ueda, M. & Tanaka, A. (1991) Presence of two transcribed malate synthase genes in an *n*-alkane utilizing yeast *Candida tropicalis*, *J. Biochem.* **110**, 909–914.
- Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) CLUSTAL V. Improved software for multiple sequence alignment, *Comput. Appl. Biosci.* **8**, 189–191.
- Hartig, A., Simon, M. M., Schuster, T., Daugherty, J. R., Yoo, H. S. & Cooper, T. G. (1992) Differentially regulated malate synthase genes participate in carbon and nitrogen metabolism of *Saccharomyces cerevisiae*, *Nucleic Acids Res.* **20**, 5677–5686.
- d'Aubenton Carafa, Y., Brody, E. & Thermes, C. (1990) Prediction of Rho-independent *Escherichia coli* transcription terminators, *J. Mol. Biol.* **216**, 835–858.
- Morgan, W. D., Bear, D. G., Litchman, B. L. & von Hippel, P. H. (1985) A sequence and secondary structure requirements for Rho-dependent transcription termination, *Nucleic Acids Res.* **13**, 3739–3754.
- Henry, H., Escher, C.-L. & Widmer, F. (1992) Salt-mediated interconversions and purification of malate synthase from germinating soybean cotyledons (*Glycine max.* L.), *Plant Sci. (Limerick)* **82**, 21–27.
- Durchschlag, H., Biedermann, G. & Eggerer, H. (1981) Large-scale purification and some properties of malate synthase from baker's yeast, *Eur. J. Biochem.* **114**, 255–262.
- Durchschlag, H. & Zipper, P. (1988) Primary and post-irradiation inactivation of the sulfhydryl enzyme malate synthase. Correlation of protective effects of additives, *FEBS Lett.* **237**, 208–212.

glc Locus of *Escherichia coli*: Characterization of Genes Encoding the Subunits of Glycolate Oxidase and the *glc* Regulator Protein

MARIA-TERESA PELLICER, JOSEFA BADÍA, JUAN AGUILAR,* AND LAURA BALDOMÀ

Department of Biochemistry, Faculty of Pharmacy, University of Barcelona, 08028 Barcelona, Spain

Received 18 October 1995/Accepted 29 January 1996

The locus *glc* (min 64.5), associated with the glycolate utilization trait in *Escherichia coli*, is known to contain *glcB*, encoding malate synthase G, and the gene(s) needed for glycolate oxidase activity. Subcloning, sequencing, insertion mutagenesis, and expression studies showed five additional genes: *glcC* and in the other direction *glcD*, *glcE*, *glcF*, and *glcG* followed by *glcB*. The gene *glcC* may encode the *glc* regulator protein. Consistently a chloramphenicol acetyltransferase insertion mutation abolished both glycolate oxidase and malate synthase G activities. The proteins encoded from *glcD* and *glcE* displayed similarity to several flavoenzymes, the one from *glcF* was found to be similar to iron-sulfur proteins, and that from *glcG* had no significant similarity to any group of proteins. The insertional mutation by a chloramphenicol acetyltransferase cassette in either *glcD*, *glcE*, or *glcF* abolished glycolate oxidase activity, indicating that presumably these proteins are subunits of this enzyme. No effect on glycolate metabolism was detected by insertional mutation in *glcG*. Northern (RNA) blot experiments showed constitutive expression of *glcC* but induced expression for the structural genes and provided no evidence for a single polycistronic transcript.

Glycolate is metabolized in *Escherichia coli* through oxidation to glyoxylate (15, 18) in a reaction catalyzed by the enzyme glycolate oxidase (21, 27). Glyoxylate is a branching point in the metabolic pathway since it is metabolized by two divergent condensation reactions. One reaction condenses glyoxylate with acetyl coenzyme A and is catalyzed by malate synthase G (36), while the other reaction condenses two molecules of glyoxylate in a process catalyzed by glyoxylate carboligase, which simultaneously decarboxylates the condensation product to tartronic semialdehyde (7). This latter compound is reduced to glycerate and subsequently phosphorylated to glycerate-3-phosphate. These three enzyme-catalyzed reactions constitute what is known as the glycerate pathway.

Locus *glc* at min 64.5 of the *E. coli* chromosome has been associated with the glycolate utilization trait (25). This locus contains the *glcB* gene, encoding malate synthase G, and also the gene(s) encoding glycolate oxidase. However, the *gcl* gene, encoding glyoxylate carboligase, is not linked to the *glc* locus; it has been located in a different position, at min 12 (7). The *glcB* gene and enzyme have been characterized previously (25). In this report we describe and characterize the rest of the genes that complete the *glc* locus and identify those encoding subunits of glycolate oxidase and the regulatory gene of the *glc* operon.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages. All the strains used were *E. coli* K-12 derivatives. The genotypes and sources of the bacterial strains and plasmids are given in Table 1. Phage P1vir was used in transduction experiments.

Growth conditions. Cells were grown aerobically on Luria broth or minimal medium as described previously (3). For growth on minimal medium, the following compounds were added at the following concentrations unless otherwise specified: D-xylose, 10 mM; glycolate, 30 mM; glyoxylate, 30 mM; and casein acid hydrolysate (CAA), 1%. When necessary, thiamine at 1.65 µg/ml and the fol-

lowing antibiotics at the concentrations indicated were added to the medium: chloramphenicol, 30 µg/ml; ampicillin, 100 µg/ml; and tetracycline, 12.5 µg/ml. 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and isopropyl-β-D-thiogalactoside (IPTG) were used at 30 and 10 µg/ml, respectively.

Preparation of cell extracts and enzyme assays. For enzyme assays, cells were harvested at the end of the exponential phase and cell extracts were prepared as described previously (4) in 10 mM Tris-HCl, pH 8.0, containing 1 mM MgCl₂. Determinations of total malate synthase activity and the relative concentration of the two forms of malate synthase, A and G, were performed as described by Ornston and Ornston (27).

Glycolate oxidase activity was determined spectrophotometrically according to the method of Lord (21). In this case, to avoid nonspecific reduction of 2,6-dichloroindophenol, the 30% ammonium sulfate fraction resuspended in 10 mM phosphate buffer, as indicated by this author, was routinely used as a source of enzyme.

Protein concentration was determined by the method of Lowry et al. (22) using bovine serum albumin as the standard.

In vivo expression of plasmid-carried genes in maxicells. The expression of plasmid-encoded proteins was performed in maxicells as described by Stoker et al. (34) with strain JA120 as a host. Proteins were labeled with L-[³⁵S]methionine (50 µCi/ml) and detected by fluorography after separation by sodium dodecyl sulfate-10 or 13.5% polyacrylamide gel electrophoresis (SDS-PAGE).

DNA manipulation. Plasmid DNA was routinely prepared by the boiling method (17). For large-scale preparation, a crude DNA sample was purified through a column (Qiagen GmbH, Düsseldorf, Germany). DNA manipulations were performed essentially as described by Sambrook et al. (29). The DNA sequence was determined by the dideoxy-chain termination procedure of Sanger et al. (30) by using the T7 sequencing kit from Pharmacia LKB Biotechnology. Double-stranded plasmid DNA was used as the template. Ordered deletions were obtained with the Erase-a-Base system (Promega Biotec, Madison, Wis.). To avoid sequencing errors, both strands were sequenced by regular dGTP as well as dITP or 7-deaza-dGTP reactions.

DNA sequence analysis and protein alignments were done with the PC/GENE software package (Intelligenetics Inc.) and with the University of Wisconsin Genetics Computer Group package programs.

Genetic techniques. Phage P1 transduction experiments were performed as described by Miller (24). The chloramphenicol resistance gene cassette CAT19 was used in the gene inactivation experiments (14, 38) by inserting it into the restriction sites indicated in Fig. 4. This cassette had no terminator designed into the downstream region of the chloramphenicol acetyltransferase (CAT) gene and thus did not always cause polarity when inserted in the same orientation as the interrupted gene. Plasmids carrying inactivated genes (Table 1) were linearized and used to transform strain JC7623 to chloramphenicol resistance (Cm^r). This strain efficiently recombines linear DNA into its chromosome (38). P1vir lysates obtained from the selected Cm^r recombinants were used to transduce the CAT insertions into the parental strain MC4100. The locations of these mutations in the *glc* locus were verified by cotransductional analysis with the

* Corresponding author. Mailing address: Department of Biochemistry, Faculty of Pharmacy, University of Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain. Phone: 34-3-402 4521. Fax: 34-3-402 1896. Electronic mail address: Jaguilar@far.ub.es.

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Source or reference
Strains		
XL1Bblue	<i>recA1 lac endA1 gyrA96 thi hsdR17 supE44 relA1</i> (F' <i>proAB lacI^a lacZ ΔM15 Tn10</i>)	Stratagene
MC4100	<i>araD Δlac rpsL ffbB deoC ptsF rbsR</i>	6
SØ1023	<i>relA1 spoT1 thi-1 nupG511::Tn10</i>	32
NK6027	<i>relA1 spoT1 thi-1 Δ(gpt-lac)5 metC162::Tn10</i>	32
JC7623	<i>arg thi thr leu pro his strA recB21 recC22 sbcB15</i>	37
JA120	<i>thi-1 thr-1 leuB6 lacY1 tonA21 supE44 recA1 srlA::Tn10</i>	1
JA151	<i>Δ(speC-glc) aceB recA::cat</i>	25
JA154	MC4100 <i>glcC::cat</i>	This study
JA155	MC4100 <i>glcD::cat</i>	This study
JA156	MC4100 <i>glcE::cat</i>	This study
JA157	MC4100 <i>glcF::cat</i>	This study
JA158	MC4100 <i>glcG::cat</i>	This study
Plasmids^a		
Bluescript	Ap ^r	Stratagene
pBR322	Ap ^r Tc ^r	Boehringer Mannheim
pCAT19	Ap ^r Cm ^r ; source of <i>cat</i> gene	14
pLB10	<i>glcBGFEDC</i> (11.5-kb <i>Hind</i> III fragment) in pUC18	25
pIM4	<i>glcBGF'</i> (4.5-kb <i>Hind</i> III- <i>Eco</i> RV fragment) in Bluescript	25
pTP21	<i>glcBGFEDC</i> (9-kb <i>Hind</i> III- <i>Eco</i> RI fragment)	This study
pTP22	<i>glcBGFED'</i> (6.5-kb <i>Hind</i> III- <i>Sal</i> I fragment)	This study
pTP23	<i>glcBGF'E'</i> (5.2-kb <i>Hind</i> III- <i>Bam</i> HI fragment)	This study
pTP24	<i>glcB'GFEDC</i> (6.8-kb <i>Clal</i> - <i>Eco</i> RI fragment)	This study
pTP25	<i>glcF'EDC'</i> (3.8-kb <i>Eco</i> RV fragment)	This study
pTP26	<i>glcE'DC</i> (3.2-kb <i>Bam</i> HI fragment)	This study
pTP27	<i>glcE'D'</i> (3.3-kb <i>Kpn</i> I- <i>Sal</i> I fragment)	This study
pTP28	<i>glcB'GFE'</i> (2.8-kb <i>Clal</i> - <i>Kpn</i> I fragment)	This study
pTP29	<i>glcF'E'</i> (0.8-kb <i>Eco</i> RV- <i>Bam</i> HI fragment)	This study
pTP30	<i>glcD'C</i> (1.3-kb <i>Mlu</i> I- <i>Bam</i> HI fragment)	This study
pTP51	As pTP26 insert but <i>glcC::cat</i> , <i>Sma</i> I-Cm ^r cassette into <i>Nru</i> I	This study
pTP52	As pTP26 insert but <i>glcD::cat</i> , <i>Sma</i> I-Cm ^r cassette into <i>Mlu</i> I	This study
pTP53	As pTP22 insert but <i>glcE::cat</i> , <i>Bam</i> HI-Cm ^r cassette into <i>Bam</i> HI	This study
pTP54	As pTP22 insert but <i>glcF::cat</i> , <i>Sma</i> I-Cm ^r cassette into <i>Eco</i> RV	This study
pTP55	As pTP28 insert but <i>glcG::cat</i> , <i>Sma</i> I-Cm ^r cassette into <i>Nru</i> I	This study

^a Plasmids pTP21 to pTP30 are in Bluescript, and plasmids pTP51 to pTP55 are in pBR322. A prime indicates that a particular gene is truncated.

merC::Tn10 (strain NK6027) and *nupG::Tn10* (strain SØ1023) markers (data not shown).

Isolation of RNA and Northern blot hybridization. For preparation of total RNA, cells of a 25-ml culture grown to an A_{650} of 0.5 were collected by centrifugation at $5,000 \times g$ for 10 min and processed as described by Belasco et al. (2). Northern (RNA) blot hybridization was performed with each RNA sample (10 μ g) by following the procedure described previously (26).

Nucleotide sequence accession number. The sequence reported here is deposited in the EMBL/GenBank database under accession number L43490.

RESULTS AND DISCUSSION

Location of glycolate oxidase genes in the *glc* locus. Previously we have reported that the recombinant plasmid pLB10, containing an 11.5-kb *Hind*III fragment, expressed malate synthase G and glycolate oxidase activities and that the pLB10-derived recombinant plasmid pIM4 was able to express only malate synthase G (25). In order to locate the gene encoding glycolate oxidase more precisely (Fig. 1), a set of subclones was derived from pLB10 and used to restore glycolate oxidase activity in strain JA151. Only plasmids containing the 6-kb region between *glcB* and the *Eco*RI restriction site (plasmids pTP21 and pTP24) were able to supply this activity. Enzyme activities were similar in the presence and in the absence of inducer, indicating that expression was from vector promoters.

Nucleotide sequence of the 5.3-kb fragment upstream of *glcB*. To identify the genes carried on the genomic fragment upstream of *glcB*, several subclones, covering the region between the *Clal*I (internal to *glcB*) and the second upstream *Bam*HI restriction sites, were constructed (plasmids pTP26,

pTP27, pTP28, and pTP29). Serial deletions of these plasmids were obtained and sequenced at least twice on each strand. Figure 2 shows the 5,335 bp of DNA sequenced between the already mentioned *Bam*HI site and the ATG start codon of *glcB* gene previously reported by Molina et al. (25). Five new open reading frames were observed, and they are named *glcD*, *glcE*, *glcF*, and *glcG*, on the same strand, and immediately followed by *glcB*. Diverging from *glcD* on the other strand is *glcC*.

Expression studies of the sequenced *glc* genes. The maxicell system (34) was used to identify the gene products of the *glc* locus described above. Strain JA120 was transformed with several plasmids containing different parts of this locus and also with the Bluescript vector. Transformed cells were grown on CAA and CAA plus glycolate. Consistent with the expression from a vector promoter, the same pattern was observed in both conditions. Plasmid pTP21, containing all the genes of this locus, expressed protein bands of 82, 56, and 40 kDa (Fig. 3A, lane 1). Plasmid pTP22, lacking *glcC* and *glcD*, did not express the 56-kDa protein (lane 2), whereas plasmid pTP23, lacking also *glcE*, expressed only the 82-kDa protein (lane 3). Plasmid pIM4, which lacked *glcF*, gave the same expression pattern as pTP23 (lane 4). Consistently, plasmid pTP25 expressed the two proteins of 56 and 40 kDa but not the 82-kDa protein corresponding to malate synthase G (lane 6), while plasmid pTP26 expressed only a protein of 56 kDa (lane 7).

In order to resolve the low-molecular-weight protein corresponding to *glcG*, the expressed products were also separated

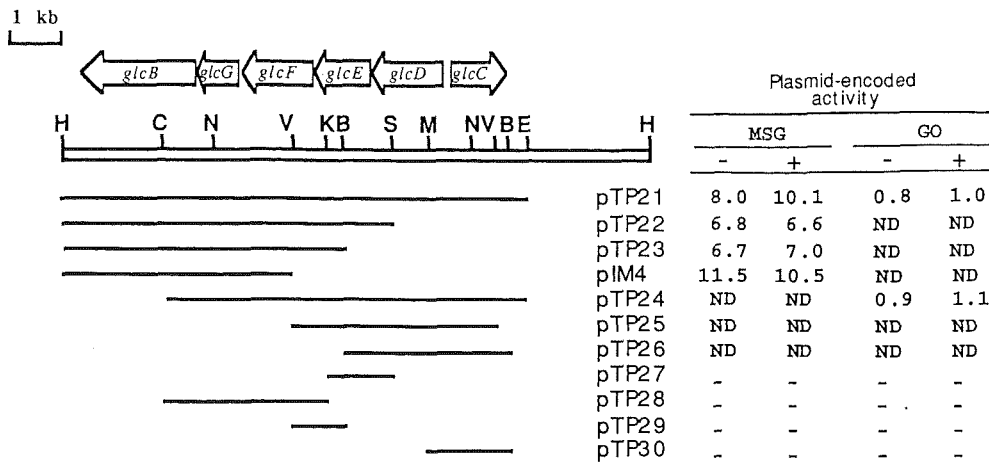


FIG. 1. Restriction map of the *glc* locus. The open bar represents the genomic fragment present in the recombinant plasmid pLB10 (25). The inserts of the plasmids used in this study are indicated by thin lines below the restriction map. These inserts were cloned in the Bluescript vector with the structural *glc* genes in the same orientation as *lacZ*. For each plasmid, malate synthase G (MSG) and glycolate oxidase (GO) activities are expressed in units per milligram of protein. Enzyme activities were determined in cell extracts of transformed cells of strain JA151 grown on CAA either in the absence (-) or in the presence (+) of glycolate. ND, not detectable. Arrows indicate the extension and direction of transcription of the identified genes. The restriction sites for several cleaving enzymes are indicated as follows: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; M, *Mlu*I; N, *Nru*I; S, *Sal*I; and V, *Eco*RV.

by SDS-13.5% PAGE. A 13-kDa protein was clearly identified in this way. To check that this protein was the product of a *glc* gene and not a product of a truncated gene between *glcB* and the *Hind*III restriction site, plasmid pTP28 was also analyzed and proved to express this polypeptide (Fig. 3B).

This expression pattern allowed us to assign the 56-kDa protein to the *glcD* gene, the 40-kDa protein to *glcE*, and the 13-kDa protein to *glcG*. The molecular weights of these expressed proteins were in accordance with those presented below and deduced from the corresponding nucleotide sequence. This system failed to identify *glcF* and *glcC* gene products.

Identification of the *glcF* gene product was achieved when the location of some of these proteins in the cell membrane was determined. To this end, membranes were obtained (28) from *glc*-deleted strain JA151 transformed with plasmid pTP21, using as a control the same strain transformed with the vector or the wild-type strain MC4100 (Fig. 3C). Cultures were grown on CAA plus glycolate. The electrophoretic development of membrane proteins stained by Coomassie blue showed the 56- and 40-kDa proteins encoded by the plasmid and an additional 45-kDa protein, not detected in the maxicell expression experiments, which may correspond to the *glcF* gene product (Fig. 3C). This protein is likely to have a low expression rate and was visualized only after its partial purification through the isolation of membranes.

Mutational analysis of the sequenced *glc* genes. The assignment of functions to the sequenced genes was determined by inactivating each of them and studying the effects on glycolate utilization. Inactivation was performed by insertional mutation of the CAT19 cassette in the locus *glc* of the wild-type strain as indicated in Materials and Methods. Insertions in all genes except in *glcG* rendered the cells glycolate negative (Fig. 4).

Determination of enzyme activities in wild-type cells grown under inducing conditions yielded values of 200 mU/mg for malate synthase G and of 60 mU/mg for glycolate oxidase. The inactivation of *glcC* (strain JA154) abolished both malate synthase G and glycolate oxidase, yielding undetectable enzyme activities. Independent inactivation of *glcD*, *glcE*, or *glcF* (strains JA155, JA156, and JA157, respectively) yielded undetectable levels for glycolate oxidase activity but not for malate synthase. These observations strongly suggested the participa-

tion of the products of these genes as subunits of the glycolate oxidase structure. Disruption of *glcG* (strain JA158) had no effect on either of the two activities (Fig. 4).

Each inactivation was complemented by a plasmid carrying the fragment in which the insertional mutation was located (Fig. 4), indicating that the inactivation was due to this mutation and that no other cell functions were impaired. The complemented cells displayed glycolate oxidase activities of 40 mU/mg. As expected, in the complementation by *glcC*, both enzyme activities were restored simultaneously (240 mU/mg for malate synthase G and 75 mU/mg for glycolate oxidase).

Features and predicted properties of *glcC*. The gene *glcC* starts at position 866 and ends at position 102 (Fig. 2). There is a Shine-Dalgarno sequence (GAGG) 11 bp upstream from the start codon. This gene encodes a 254-amino-acid protein with a calculated molecular mass of 28,901 Da. Its deduced amino acid sequence was compared to protein sequences in the GenBank database and shown to be similar to several prokaryotic regulatory proteins. A helix-turn-helix motif which was homologous to the pyruvate dehydrogenase complex repressor (16) and L-lactate dehydrogenase regulatory protein (13) of *E. coli* and to the *hutC* repressor protein of *Klebsiella aerogenes* (31), among others, was found.

The high similarity of the *glcC* deduced amino acid sequence to other regulatory proteins and the fact that its inactivation simultaneously affected both the glycolate oxidase and the malate synthase G activities indicated that this gene encodes the *glc* regulatory protein. This protein would act as an activator since mutations in *glcC* did not render the expression of the *glc* structural genes constitutive but rather abolished it.

Features and predicted properties of *glcD* and *glcE*. The gene *glcD* (positions 1117 to 2616 of Fig. 2) is separated from *glcC* by a 251-bp region. A good ribosome-binding site (GAAG) was identified 11 bp upstream of the ATG codon. This gene encodes a 499-amino-acid polypeptide with a calculated molecular mass of 53,749 Da. The comparison with protein sequences in the GenBank database showed a significant similarity to the sequences of D-lactate dehydrogenases from *Kluyveromyces lactis* (20) and from *Saccharomyces cerevisiae* (19) and also to that from *E. coli*, although the last showed less similarity. Since it has been reported that the glycolate oxidase

1 ggatcctcgaaccaatacagggcatgaataaacatgctggaacgcattggggcggagccgtacaagcgttaccgattgtatgaaagcaga 90
 91 ttaataaccagctaactcaggttcacatctccagcggccacggcagcgaatcaaattggtggcctccagctcgatatacgtggagattcttt 180
 * S L N M E L P V A S R I L H H G E L E I D H L N K
 181 ttcacggtccgcacatgatcgctgctgctgctggcgacgtgcccagccgctgcaaacaccgctttagatccgcgatgctggcga 270
 K V T R V H D R A A R Q A V H P L R Q L V A N Y I R A H Q R
 271 tcgatctgcttttttctgctggtgtagatgtagagattatctacgctggcaaacactgaattaacatcagatcgggtcaatgattgcagc 360
 D I Q K K Q Q P R H Y L N N V S A F V S N F M L D T L S Q L
 361 gtaaacaccagcacctgattgtgagaagcctgacaaatggcgagatggaagcgtgatccaactcggcatggttcgatcagcgaatctct 450
 T F V L V Q N H S A Q C I A L H F A H D L Q A H E I L S I E
 451 ttgttgttctcactggcggcgagcattttttcataacagcgggttatcacaacaaaatcagcctgcttccagcgttgcggccagcctt 540
 K N N E S A A L M K E Y C R T I V V F D A Q T G L T A A L R
 541 gccgattcgcctccagtaaatgcccgaacgtcagcagatcgtacagcgttcgcccgtcgtactgaacagatggatcagcgggctgggtg 630
 A S E G E L L A R V D L L D Y L T R P Q T S F L H I L P S T
 631 tctcgcacccgattaagcctgcccagcagagaatcggcaccctgcccgtttcaataatcccgcgcccgccagcagcggctcagcccttca 720
 D Q V R N L R A V R S D R G Q A T E I I G R G R L V T L G E
 721 cgcagtgccggagcgtgagaagccgagcctttcacacagtcagcgttcggaggggaagcggctgaccgaccttcagtacgccctcgataatt 810
 R L A S R S F G L K E C L R R E S P L P Q G V K L V G D I I
 811 aaccgttcgatactctctgcaaccacttcgcaaatggggcagcgttcattcttccaggacctcgtgacaggttaggaccaattttt 900
 L R E Y S E A V V E C I P R R E D K M
 ← *glcC*
 901 ctgttcttaactcgcaaaacacgcacatcagcgaagtattatgttatcatcaatttaacattgagttaaccaagacaaggtcacagagct 990
 991 ggaaaaaaaaatggtctgaccggtagctaaagagatagacgaaaacgaaaagcccgttaataactgttcacagaagcagcggcgaaaaat 1080
 1081 cagctgccacacacacacacaaaagcgaagcctactcatgagcatcttgtacgaagagcgtctttagtgccgctttaccggatgtcgaccgc 1170
 M S I L Y E E R L D G A L P D V D R
glcD →
 1171 acatcgggtactgatggcactgcccgtgagcagcgttcctcctggacttgagatcctgcataccgatgaggagatcattccttacaggttgacggg 1260
 T S V L M A L R E H V P G L E I L H T D E E I I P Y E C D G
 1261 ttgagcggatcgcacgcgtccattactggttcttctgcttaagcaaatggaacaggtgacagcgattctggctgtctgccaatcgccctg 1350
 L S A Y R T R P L L V V L P K Q M E Q V T A I L A V C H R L
 1351 cgtgtaccgggtggtgaccctggtgacggcaccgggctttctggtggcgcgtgcccgtggaanaaggtgtgttgttgggtgatggcggc 1440
 R V P V V T R G A G T G L S G G A L P L E K G V L L V M A R
 1441 ttaaaagagatcctcgacattaaccccgttggctgcggcggcggcgtgcagccagcggcgtggttaacctggcgatctccagcggcgttgc 1530
 F K E I L D I N P V G R R A R V Q P G V R N L A I S Q A V A
 1531 ccgcataatctctactacgcaccggacccttctcacaatgcctgttccattggcggcaatgtggctgaaaatgccggcggcgtccac 1620
 P H N L Y Y A P D P S S Q I A C S I G G N V A E N A G G V H
 1621 tgctgaaatattggtctgaccgtacataacctgctgaaaattgaagtgcaaacgctggacggcggcagcactgacacttggatcggacggc 1710
 C L K Y G L T V H N L L K I E V Q T L D G E A L T L G S D A

FIG. 2. Nucleotide sequence of the 5.3-kb genomic fragment upstream of *glcB* in the *glc* locus. Only one strand is shown. The genes *glcC* (transcribed clockwise), and *glcD*, *glcE*, *glcF*, and *glcG* (transcribed counterclockwise) have been translated in the one-letter amino acid code; amino acid symbols are written below the first nucleotide of the corresponding codon. Stop codons are indicated by asterisks. The start codon of *glcB* (accession number X74547) is also indicated. Putative ribosome binding sites are underlined. The GenBank accession number of this sequence is L43490.

1711 ctggattcaactggttttgaacctgctggcgctgttcaccggatcgggaaggtatgctcggcgtgaccaccggaagtacggtaaaactgctg 1800
 L D S P G F D L L A L F T G S E G M L G V T T E V T V K L L

1801 ccgaagccgcccgtggcggggttctgttagccagctttgactcggtagaaaaagccggacttggcgttggtgacatcatcgccaatggc 1890
 P K P P V A R V L L A S F D S V E K A G L A V G D I I A N G

1891 attatccccggggctggagatgatggataacctgtcgatccgcggcggaagattttattcatgcccgttatcccgtcgacgcccga 1980
 I I P G G L E M M D N L S I R A A E D F I H A G Y P V D A E

1981 gcgattttgttatgcgagctggacggcgtggagtctgacgtacaggaagactgcgagcgggttaacgacatcttgttgaagcggggcg 2070
 A I L L C E L D G V E S D V Q E D C E R V N D I L L K A G A

2071 actgacgtccgtctggcacaggacgaagcagagcgcgtacgtttctggcggctggcaaaaatgcgttcccggcggttaggacgtatctcc 2160
 T D V R L A Q D E A E R V R F W A G R K N A F P A V G R I S

2161 ccgattactactgcatggatggcaccatcccgcgtcggccctgctggcgtactggaagcattgcccgtttatcgacgaatgat 2250
 P D Y Y C M D G T I P R R A L P G V L E G I A R L S Q Q Y D

2251 ttacgtgttgccaactctttcatgcccggagatggcaacatgcaccggtaaatccttttcgatgccaacgaaccgggtgaattggcccgc 2340
 L R V A N V F H A G D G N M H P L I L F D A N E P G E F A R

2341 gcggaagagctggcggggaagatcctcgaactctgcgttgagttggcggcagcatcagtgccgaacatggcatcggggcagaaaaaac 2430
 A E E L G G K I L E L C V E V G G S I S G E H G I G R E K I

2431 aatcaaatgtggcccagttcaacagcgatgaaatcacgaccttccatgcgggtcaaggcggcgtttgaccccgatggttggctgaacct 2520
 N Q M C A Q F N S D E I T T F H A V K A A F D P D G L L N P

2521 gggaaaaacattcccacgctacaccgctgtgctgaatttggcgcctgcatgtgcatcacggtcatttaccttccctgaactggagcgt 2610
 G K N I P T L H R C A E F G A M H V H H G H L P F P E L E R

2611 ttctgatgctacgcgagtggtgattacagccagcgcgtgctggagcaggtaaatcaggcgttagcagataaaacggcgtggtgattcagg 2700
 F * M L R E C D Y S Q A L L E Q V N Q A I S D K T P L V I Q G
glcE →

2701 gcagcaatagcaaagccttttagtgcgccctgtcacggggcaaacgctggatgttctgtgtcatcgggcattgttaattacgaccgga 2790
 S N S K A F L G R P V T G Q T L D V R C H R G I V N Y D P T

2791 ccgagctggtgataaccgcccgtgtcggaaacggcgtggtgacaattgaagcggcgtgaaagcgggggcaaatgctcccctgtgagc 2880
 E L V I T A R V G T P L V T I E A A L E S A G Q M L P C E P

2881 cggcgcattatggtgaagaagccacctggggcgggatggtgcctcggcgtggcggggccgcgtcggcgtggagcgttccgtccgcg 2970
 P H Y G E E A T W G G M V A C G L A G P R R P W S G S V R D

2971 attttgcctcggcacgcgcatcattaccggcgtgaaaaacatctgcgttttggcgggaagtgatgaaaaacgttggcggatacgtc 3060
 F V L G T R I I T G A G K H L R F G G E V M K N V A G Y D L

3061 tctcacggttaattggtcggaaagctacggttgtcttggcgtgctcactgaaatctcaatgaaagtgttaccgacccgcgcctccctga 3150
 S R L M V G S Y G C L G V L T E I S M K V L P R P R A S L S

3151 gcctcgtcgggaaatcagcctgcaagaagccatgagtgaatcggcagtggaactccagccattaccattagtggttatgttact 3240
 L R R E I S L Q E A M S E I A E W Q L Q P L P I S G L C Y F

3241 togacaatgcgttggatccgcttgaggcggcgaaggatcggtaaaagcagcgcgtgaactgctgggtggcgaagaggttggcggtc 3330
 D N A L W I R L E G G E G S V K A A R E L L G G E E V A G Q

3331 agttctggcagcaattgcgtgaacaacaactgcggttctctcgttaccaggtacottatggcgcatttcattaccaggtgatgcggca 3420
 F W Q Q L R E Q Q L P F F S L P G T L W R I S L P S D A P M

FIG. 2—Continued.

3421 tgatggattaccogggagcaactgatcgactggggggggggttacgctggctgaaatcgacagccgaggacaatcaaatccatcgca 3510
 M D L P G E Q L I D W G G A L R W L K S T A E D N Q I H R I

3511 tcgcccgaacgctggcggtcatgacccgcttagtgccggagatgggtgctttgcccgcctatcggtccctttattccgctatcacc 3600
 A R N A G G H A T R F S A G D G G F A P L S A P L F R Y H Q

3601 agcagcttaaacagcagctcgacccttgccggctgtttaacccccgctgcgatgtacggggaactttgaggagcaggctatgcaaacccaa 3690
 Q L K Q Q L D P C G V F N P G R M Y A E L * M Q T Q
glcF →

3691 ttaactgaagagatgcccagacaacggcgccgctggaagccgacagcatcctcgccgctgtgttcactggcgattttgtacggcaacc 3780
 L T E E M R Q N A R A L E A D S I L R A C V H C G F C T A T

3781 tgcccaacctatcagcttctggcgatgaactggacggggcgccgcatctatctgattaacaggtgctggaaggcaacgaagt 3870
 C P T Y Q L L G D E L D G P R G R I Y L I K Q V L E G N E V

3871 acgcttaaacacaggagcatctcgatcgctgcctcacttgccgtaattgtgaaaccactgtccttctggtgctgctatcacatttg 3960
 T L K T Q E H L D R C L T C R N C E T T C P S G V R Y H N L

3961 ctgatatcgccgctgatattgtcgagcagaaagtgaacgccactgcccggagcgaatactgcccgaaggattgcccaggtagtgcg 4050
 L D I G R D I V E Q K V K R P L P E R I L R E G L R Q V V P

4051 cgtccgggggttctccgtgctgacgcaggtagggtgctgctgacccgtttttaccggacaggtcagagcaaaactgctgctgaa 4140
 R P A V F R A L T Q V G L V L R P F L P E Q V R A K L P A E

4141 acggtgaaagctaaccggctccggcctgcccataagcgtccgggttttaagtgtggaaggctgcccagcctacgcttttgcgcaac 4230
 T V K A K P R P P L R H K R R V L M L E G C A Q P T L S P N

4231 accaacggcgcaactgcccagtgctggatcgtctgggatcagcgtcatgccagcstaacgaagcaggctgttggcgccggtggactat 4320
 T N A A T A R V L D R L G I S V M P A N E A G C C G A V D Y

4321 catcttaatgcccagagaaaggctggcacggggcgcaataatattgatgctggtggcccggcattgaagcaggtgcccaggaatt 4410
 H L N A Q E K G L A R A R N N I D A W W P A I E A G A E A I

4411 ttgcaaacggccagggctgcccggcgtttgtcaaagatggggcagatgctgaaaaacgatgcgttatatgcccataaagcagctcag 4500
 L Q T A S G C G A F V K E Y G Q M L K N D A L Y A D K A R Q

4501 gtcagtgaaactggcggtcgatttagtgaacttctgcccggaggaaccgctggaaaaactggcaattcggggcataaaaagctggccttc 4590
 V S E L A V D L V E L L R E E P L E K L A I R G D K K L A F

4591 cactgtccgtgtaccctacaacatgcgcaaaagctgaacggcgaagtggaaaaagtgttgccttcttggattaccctaacggacgtt 4680
 H C P C T L Q H A Q K L N G E V E K V L L R L G F T L T D V

4681 cccagacccatctgtgctgcccgttcagcgggaacatagcgttaacgcacccgatctggcacgccagctgcccggataacaaaatgat 4770
 P D S H L C C G S A G T Y A L T H P D L A R Q L R D N K M N

4771 gcgctggaagcggcaaacggaaatgatcgtcaccgccaacattggttgcagacgcacatctggcgagcggcgtcgtacctctgtgct 4860
 A L E S G K P E M I V T A N I G C Q T H L A S A G R T S V R

4861 cactggattgaaattgtagaacaagcccttgaagggaataacaaaatgaaaactaaagtccattcttagccagcaaatggcgagtgcaat 4950
 H W I E I V E Q A L E K E * M K T K V I L S Q Q M A S A I
glcG →

4951 tattgcccaggtcaggaagaggcgcagaaaaataactggtctggttccattgctggttgcgatgacggcggctcatctgctggcgttaag 5040
 I A A G Q E E A Q K N N W S V S I A V A D D G G H L L A L S

5041 tcgcatggagattgcccggcattgcccgttatatctcccaggagaaagcgcgtaccgcccggctggggcgtgtaaaactaaggccta 5130
 R M D D C A P I A A Y I S Q E K A R T A A L G R R E T K G Y

5131 tgaagagatgggtgaacaaaggacgtaccgcttctgactgcccgttatatacgtcgtggaaggcggcgtaccggttgggtggatgg 5220
 E E M V N N G R T A F V T A P L L T S L E G G V P V V V D G

5221 gcaattattggtgcccgtggcgtttctggttaaccggagcacaggatgcccaggtcgcgaaaggcggcagcagcgggttggcgaata 5310
 Q I I G A V G V S G L T G A Q D A Q V A K A A A A V L A K *

5311 agcgaacagaggagataaacaatg 5335
 M
glcB →

FIG. 2—Continued.

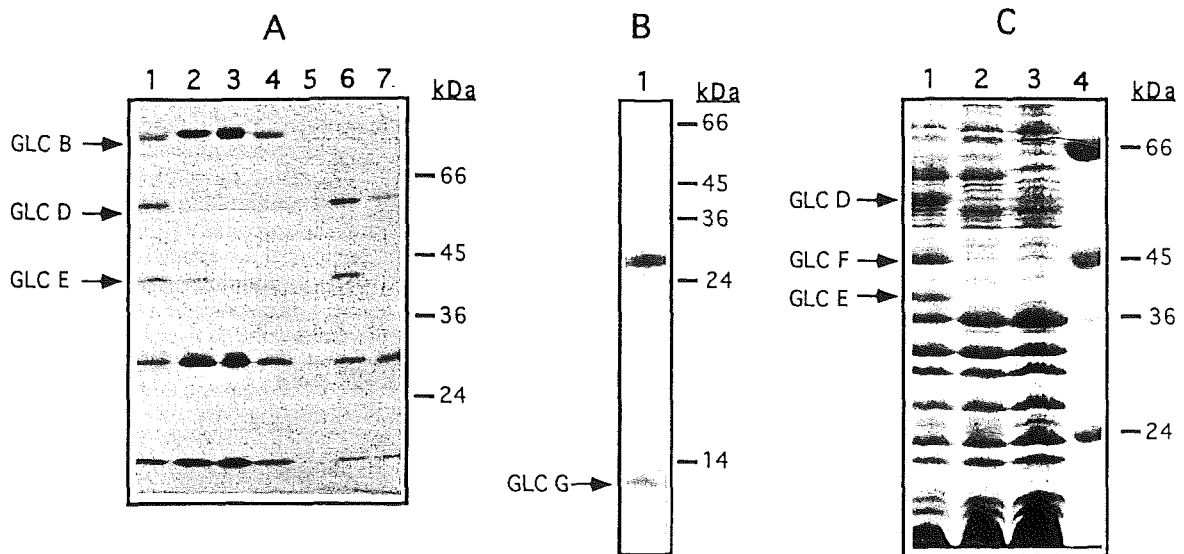


FIG. 3. (A) Autoradiograph of the gene products encoded by recombinant plasmids containing different parts of the *glc* locus and expressed in the maxicell system using strain JA120 as a host. Proteins were separated by SDS-10% PAGE. Lane 1, plasmid pTP21; lane 2, plasmid pTP22; lane 3, plasmid pTP23; lane 4, plasmid pIM4; lane 5, BlueScript vector; lane 6, plasmid pTP25; lane 7, plasmid pTP26. (B) Autoradiograph of the gene products encoded by plasmid pTP28 separated by SDS-13.5% PAGE. (C) SDS-PAGE of crude membrane proteins. Samples of 150 µg of membrane proteins obtained as indicated in the text were applied to a 10% polyacrylamide gel, electrophoresed, and stained with Coomassie blue. Lane 1, strain JA151 transformed with plasmid pTP21; lane 2, strain JA151 transformed with vector Bluescript; lane 3, wild-type strain MC4100. Gene products are indicated on the left of each panel. Molecular mass markers indicated on the right of the three panels were bovine serum albumin (66 kDa), ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), trypsinogen (24 kDa), and lactalbumin (14 kDa).

of *E. coli* can use D-lactate as a substrate (21), this gene could encode a protein involved in the oxidation of glycolate.

The initiation codon of *glcE* overlaps with the termination codon of *glcD*. Its putative Shine-Dalgarno sequence (GGAG) has been located 12 bp upstream of the ATG codon. The *glcE* gene encodes a 350-amino-acid protein with a calculated molecular mass of 38,352 Da. When compared with the protein sequences in the GenBank database, the highest similarity was again found with the D-lactate dehydrogenase from *S. cerevisiae*. On the basis of this similarity, this protein could be assigned to another subunit of glycolate oxidase.

Upon examination of the primary structures of *glcD* and *glcE* gene products, possible flavin-binding domains were easily detected. In the case of *glcD* we have identified a partially conserved consensus sequence like that described by Branden and Tooze (5). This consensus sequence contained the motif Gly-X-Gly-X-X-Gly separated from a conserved glutamate by 17

residues and included several invariant hydrophobic positions (Fig. 5A). In addition, a sequence showing significant homology to the active site of several oxidoreductases was found for both the *glcD* and *glcE* proteins. This sequence contained a histidine residue (His-140 for the *glcD* protein and His-91 for the *glcE* protein) separated from a cysteine residue by 13 and 12 amino acids, respectively, the same distance as in anaerobic glycerol-3-phosphate dehydrogenase (9), fumarate reductase (8), or succinate dehydrogenase (39).

It has been described for some flavin adenine dinucleotide or flavin mononucleotide enzymes like anaerobic glycerol-3-phosphate dehydrogenase that the enzyme activity is higher when measured in the presence of these nucleotides. To test the possible activation of the glycolate oxidase activity of our preparations by these cofactors, we used concentrations of flavin adenine dinucleotide ranging from 2 to 50 µM and flavin mononucleotide from 0.2 to 5 mM in the assay mixture. The

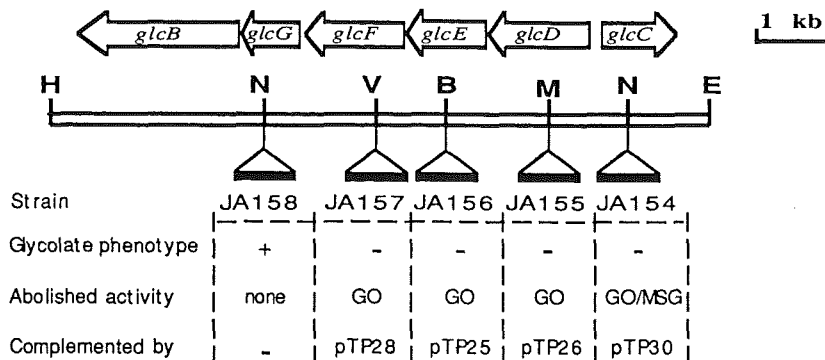


FIG. 4. Physical map of the *glc* locus insertional mutagenesis by CAT19 cassette. The open bar represents the genomic DNA of *E. coli* wild-type strain MC4100, and the arrows represent the extension and direction of transcribed genes. Insertions of the CAT cassette are indicated by black bars below the map. The restriction sites used for each insertion are indicated as follows: B, *Bam*HI; M, *Mlu*I; N, *Nru*I; and V, *Eco*RV. Relevant characterization of each insertion mutant strain is presented. GO, glycolate oxidase; MSG, malate synthase G.

A				
84	TRGAGTGLSGGALPLEKGVLLVMARFKE	111	GLC D	
5	TVIMGGGLAGLLCGLQLQKHGLRCAIVT	32	GLP B	
11	VIIIGGGATGAGIARDICALRGLRVILVE	38	GLP A	
9	AVVIGAGGAGIARLAQISQSGQTCALLS	36	SDH A	
8	VVVLGAGPAGYSAAFRCADLGLLETVIVE	35	LPD	
B				
22	LRACVHCGFCTATCPTYQLLG	42	GLC F	
72	LDRCLTCRNCETTCPSGVRYH	92	GLC F	
6	FENCIKCTVCTTACPVSRVNP	26	GLP C	
53	LKYCINCKRCEVACPSDVKIG	73	GLP C	
145	LYECILCACCSSTSCPSFWWNP	165	SDH B	
145	FSGCINCGLCYAACPFQGLNP	165	FRD B	
5	TDECISCGACAAECPVEAIEH	25	FDX (CT)	
34	ADTCIDCGACEAVCPTGAVKA	54	FDX (CT)	
5	EETCISCGACAAECPVNAIEQ	25	FDX (RR)	
34	ADTCIDCGNCANVCPVGPAPVA	54	FDX (RR)	
+--C+--C--C---CP-----•				

FIG. 5. (A) Consensus sequence of the flavin cofactor binding sites from the following *E. coli* enzymes: GlcD, glycolate oxidase; GlpA and GlpB, anaerobic glycerol-3-phosphate dehydrogenase (9); SdhA, succinate dehydrogenase (39); and Lpd, lipoamide dehydrogenase (33). The invariant positions for glycines and the conserved glutamate are indicated in boldface. The dots mark the rather conserved hydrophobic positions which complete the consensus for the flavin binding site as indicated by Branden and Tooze (5). The numbering corresponds to the amino acid positions in each protein. (B) Amino acid sequences of iron-sulfur centers from the following *E. coli* proteins: GlcF, glycolate oxidase; GlpC, anaerobic glycerol-3-phosphate dehydrogenase (9); SdhB, succinate dehydrogenase (12); and FrdB, fumarate reductase (10). Alignment also includes ferredoxins from *Clostridium thermosaccharolyticum* (35) and *Rhodospirillum rubrum* (23) labeled as Fdx (CT) and Fdx (RR), respectively. The different cysteine clusters are aligned, and the highly conserved residues are shown in boldface. Numbering indicates the position of amino acids in each protein. The consensus is presented at the bottom, where cysteines are indicated by C, proline is indicated by P, conserved hydrophobic amino acids are indicated by plus signs, and an aliphatic amino acid is indicated by a dot.

lack of sensitivity of glycolate oxidase activity in our conditions is noteworthy.

Furthermore, the *glcD*- and *glcE*-encoded subunits of the glycolate oxidase were found to be absolutely dependent on the *glcF* subunit for activity. Preparations obtained from plasmid pTP25 expressing only *glcDE* did not transfer electrons directly to phenazine methosulfate, an acceptor used without the iron-sulfur intermediate by other flavoenzymes such as *glpAB*-encoded subunits of anaerobic glycerol-3-phosphate dehydrogenase (9). This observation reinforced the idea of a different mechanism for glycolate oxidase.

Features and predicted properties of *glcF*. Three possible start codons at positions 3532, 3556, and 3679 (Fig. 2) were identified for the gene *glcF*. The first two ATG codons overlapped with the *glcE* gene and had no well-conserved Shine-Dalgarno sequences upstream. The open reading frame, starting at position 3679 and ending at position 4902, displayed a

putative ribosome-binding site (GAGGAG) 12 bp upstream of the corresponding ATG codon. The use of this translation initiation signal predicts the expression of a protein of 407 amino acid residues with a molecular mass of 45,083 Da.

The *glcF* subunit, required for the function of the glycolate oxidase enzyme, was a cysteine-rich protein containing 17 residues, most of them organized in two clusters typical of iron-sulfur proteins. Upon alignment of these clusters with those of other respiratory enzymes it was clear that besides the characteristic spatial arrangement of the cysteine residues (Cys-X-X-Cys-X-X-Cys-X-X-X-Cys), additional positions were conserved in these sequences. For instance, a hydrophobic residue located two positions before the first cysteine, which was followed by another hydrophobic residue. The last cysteine residue was followed by a conserved proline and, six residues later, by an aliphatic amino acid (Fig. 5B).

Among the proteins giving high-scoring segment pairs with *glcF* gene product, as indicated by the BLAST program, there were the GlpC subunit of anaerobic glycerol-3-phosphate dehydrogenase from *E. coli* (9) and the iron-sulfur subunit of succinate dehydrogenase from different species. These enzymes are multimeric proteins with similar catalytic mechanisms, oxidizing molecules and transferring electrons through iron-sulfur subunits.

The *glcF* low expression, probably giving a lower cellular concentration, may indicate that several *glcD*-*glcE* products are associated with one *glcF* product. Alternatively, the subunits encoded by *glcD*-*glcE* might act catalytically on the *glcF* protein, transferring their electrons to this terminal subunit of the enzyme (9).

Features and predicted properties of *glcG*. The initiation codon of *glcG* is located 5 bp downstream of the *glcF* termination codon and is preceded by a Shine-Dalgarno sequence (GGAA) 11 bp upstream. This gene encodes a 134-amino-acid protein with a calculated molecular mass of 13,737 Da. There was no significant similarity to any protein or group of proteins, although some similarity to a 142-amino-acid protein of unknown function was observed (ORFY next to the *dha* regulon from *Citrobacter freundii*, accession number CFU09771) (11).

At present we have been unable to assign any function to the *glcG* gene product, as its inactivation by the CAT cassette insertion did not affect glycolate utilization under the conditions tested. Furthermore, strain JA151 transformed with a plasmid expressing a truncated GlcG protein displayed normal levels of glycolate oxidase activity (not shown). This plasmid was obtained by deletion from the *Cla*I restriction site of plasmid pTP24 (Fig. 1) and the extent of the deletion, which removed up to amino acid 30 of the C-terminal end of GlcG, was ascertained by sequencing.

Transcription. Total RNA was prepared as indicated above from cells of strain MC4100 grown either on D-xylose as a noninducing carbon source or on D-xylose plus glyoxylate, D-xylose plus glycolate, or glycolate alone as an inducing carbon source. Northern blot hybridizations of these RNA preparations were performed with a probe of each of the *glc* genes.

For the *glcD* structural gene, a transcript of 2.6 kb was detected only under inducing conditions in the presence of glycolate or glyoxylate, indicating the specificity of the transcription (Fig. 6). Similar results were obtained for the other structural genes, showing transcripts ranging from 2.1 to 2.6 kb (not shown). Thus, none of the bands detected in the Northern experiment was in accordance with the expected size of the possible transcripts. No direct evidence for a polycistronic operon was obtained from Northern blot experiments or polarity effects. Alternatively, the discrete size of the transcripts observed could be indicative of transcript degradation.

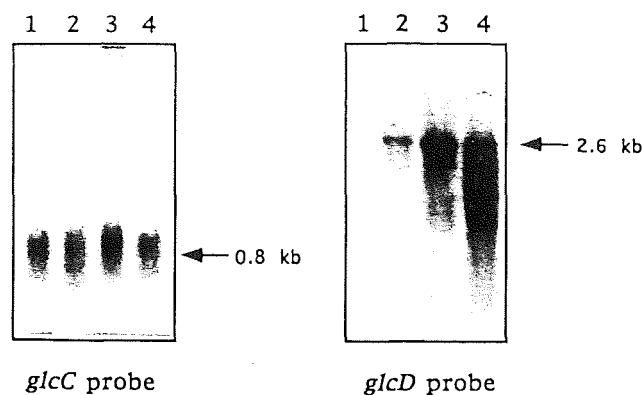


FIG. 6. Northern blots of total RNA from strain MC4100 grown on D-xylose (lanes 1), D-xylose plus glyoxylate (lanes 2), D-xylose plus glycolate (lanes 3), and glycolate (lanes 4). Hybridization was performed with a 500-bp *glcC* specific probe (*NruI-EcoRV* internal fragment) (left) or a 810-bp *glcD* specific probe (*SalI* internal fragment) (right).

A constitutive transcript of 0.8 kb was detected with the *glcC* probe, well in agreement with the size of this regulatory gene's expected transcript (Fig. 6). Finally, hybridization with probes containing sequences beyond the *glcB* and *glcC* termini displayed no bands in Northern blots of RNAs from either induced or noninduced cells. This seems to indicate that no *glc* genes are found over these limits.

ACKNOWLEDGMENTS

This work was supported by grant PB-94-0829 from the Dirección General de Investigación Científica y Técnica, Madrid, Spain. We thank Robin Rycroft for editorial assistance.

REFERENCES

- Badía, J., L. Baldomà, J. Aguilar, and A. Boronat. 1989. Identification of the *rhaA*, *rhaB* and *rhaD* gene products from *Escherichia coli* K12. *FEMS Microbiol Lett.* 65:253-258.
- Belasco, J. G., T. Beatty, C. W. Adams, A. von Gabain, and S. N. Cohen. 1985. Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic *psaA* transcript. *Cell* 40:171-181.
- Boronat, A., and J. Aguilar. 1979. Rhamnose-induced propanediol oxidoreductase in *Escherichia coli*: purification, properties, and comparison with the fucose-induced enzyme. *J. Bacteriol.* 140:320-326.
- Boronat, A., and J. Aguilar. 1981. Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: differences in induction of propanediol oxidoreductase. *J. Bacteriol.* 147:181-185.
- Branden, C., and J. Tooze. 1991. Introduction to protein structure, p. 148-150. Garland Publishing, New York.
- Casadaban, M. J. 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J. Mol. Biol.* 104:541-555.
- Chang, Y. Y., A. Y. Wang, and J. E. Cronan, Jr. 1993. Molecular cloning, DNA sequencing, and biochemical analyses of *Escherichia coli* glyoxylate carboxylase. *J. Biol. Chem.* 268:3911-3919.
- Cole, S. T. 1982. Nucleotide sequence coding for the flavoprotein subunit of the fumarate reductase of *Escherichia coli*. *Eur. J. Biochem.* 122:479-484.
- Cole, S. T., K. Eigelmeir, S. Ahmed, N. Honore, L. Elmes, W. F. Anderson, and J. H. Weiner. 1988. Nucleotide sequence and gene-polypeptide relationship of the *glpABC* operon encoding the anaerobic sn-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* 170:2448-2456.
- Cole, S. T., T. Grundtrom, B. Jaurin, J. J. Robinson, and J. H. Weiner. 1982. Location and nucleotide sequence of *frdB*, the gene encoding for the iron-sulphur protein subunit of the fumarate reductase of *Escherichia coli*. *Eur. J. Biochem.* 126:211-216.
- Daniel, R., K. Stuert, and G. Gottschalk. 1995. Biochemical and molecular characterization of the oxidative branch of glycerol utilization by *Citrobacter freundii*. *J. Bacteriol.* 177:4392-4401.
- Darlison, M. G., and J. R. Guest. 1984. Nucleotide sequence encoding the iron-sulphur protein subunit of the succinate dehydrogenase of *Escherichia coli*. *Biochem. J.* 223:507-517.
- Dong, J. M., J. S. Taylor, D. J. Latour, S. Iuchi, and E. C. C. Lin. 1993. Three overlapping *lct* genes involved in L-lactate utilization by *Escherichia coli*. *J. Bacteriol.* 175:6671-6678.
- Fuqua, W. C. 1992. An improved chloramphenicol resistance gene cassette for site-directed marker replacement mutagenesis. *BioTechniques* 12:223-225.
- Hansen, R. W., and J. A. Hayashi. 1962. Glycolate metabolism in *Escherichia coli*. *J. Bacteriol.* 83:679-687.
- Haydon, D. J., M. A. Quail, and J. R. Guest. 1993. A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex in *Escherichia coli* is located within the *pdhR* gene. *FEBS Lett.* 336:43-47.
- Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114:193-197.
- Kornberg, H. L., and J. R. Sadler. 1961. The metabolism of C₂ compounds in microorganisms. 8. A dicarboxylic acid cycle as a route for the oxidation of glycolate by *Escherichia coli*. *Biochem. J.* 81:503-513.
- Lodi, T., and I. Ferrero. 1993. Isolation of the DLD gene of *Saccharomyces cerevisiae* encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase. *Mol. Gen. Genet.* 238:315-324.
- Lodi, T., D. O'Connor, P. Goffrini, and I. Ferrero. 1994. Carbon catabolite repression in *Kluyveromyces fragilis*: isolation and characterization of the KIDL gene encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase. *Mol. Gen. Genet.* 244:622-629.
- Lord, J. M. 1972. Glycolate oxidoreductase in *Escherichia coli*. *Biochim. Biophys. Acta* 267:227-237.
- Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall. 1951. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193:265-273.
- Matsubara, H., K. Inoue, T. Hase, H. Hiura, T. Kakuno, and J. Yamashita. 1983. Structure of the extracellular ferredoxin from *Rhodospirillum rubrum*: close similarity to clostridial ferredoxins. *J. Biochem.* 93:1385-1390.
- Miller, J. H. 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Molina, I., M. T. Pellicer, J. Badía, J. Aguilar, and L. Baldomà. 1994. Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme. *Eur. J. Biochem.* 224:541-548.
- Moralejo, P., S. M. Egan, E. Hidalgo, and J. Aguilar. 1993. Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. *J. Bacteriol.* 175:5585-5594.
- Ornston, L. N., and M. K. Ornston. 1969. Regulation of glyoxylate metabolism in *Escherichia coli* K-12. *J. Bacteriol.* 98:1098-1108.
- Sallal, A.-K. J., and N. A. Nimer. 1989. The intracellular localization of glyoxylate oxidoreductase in *Escherichia coli*. *FEBS Lett.* 258:277-280.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Schwacha, A., and R. A. Bender. 1990. Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Klebsiella aerogenes*. *J. Bacteriol.* 172:5477-5481.
- Singer, M., T. A. Baker, G. Schnitzler, S. M. Deischel, M. Goel, W. Dove, K. J. Jaacks, A. D. Grossman, J. W. Erickson, and C. A. Gross. 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. *Microbiol. Rev.* 53:1-24.
- Stephens, P. E., H. M. Lewis, M. G. Darlison, and J. R. Guest. 1983. Nucleotide sequence of the lipamide dehydrogenase gene of *Escherichia coli* K-12. *Eur. J. Biochem.* 135:519-527.
- Stoker, N. G., J. M. Pratt, and H. B. Holland. 1984. "In vivo" gene expression systems in prokaryotes, p. 154-172. In B. D. Hames and S. H. Higgins (ed.), *Transcription and translation: a practical approach*. IRL Press Ltd., Oxford.
- Tanaka, M., M. Haniu, K. T. Yasunobu, R. H. Himes, and J. M. Akagi. 1973. The primary structure of the *Clostridium thermosaccharolyticum* ferredoxin, a heat stable ferredoxin. *J. Biol. Chem.* 248:2215-2217.
- Vandervinkel, E., and M. De Vlieghere. 1968. Physiologie et génétique de l'isocitrate et des malate synthases chez *Escherichia coli*. *Eur. J. Biochem.* 5:81-90.
- Wackernagel, W. 1973. Genetic transformation in *E. coli*: the inhibitory role of the *recBC* DNase. *Biochem. Biophys. Res. Commun.* 51:306-311.
- Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker. 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. *J. Bacteriol.* 161:1219-1221.
- Woods, D., M. G. Darlison, R. J. Wilde, and J. R. Guest. 1984. Nucleotide sequence encoding the flavoprotein and hydrophobic subunits of the succinate dehydrogenase of *Escherichia coli*. *Biochem. J.* 222:519-534.