UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
CURSO DE BIOMEDICINA

Eduardo Preusser de Mattos

# Análise global do metabolismo de cofatores [Fe-S] em *Azotobacter vinelandii*

Porto Alegre

2010

Eduardo Preusser de Mattos

# Análise global do metabolismo de cofatores [Fe-S] em *Azotobacter vinelandii*

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do grau de Biomédico pela Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Jeverson Frazzon

Co-orientador: Prof. Dr. João Carlos Setubal

Porto Alegre

2010

Agradecimentos e dedicatória

Primeiramente, gostaria de agradecer ao orientador e amigo Prof. Dr. Jeverson Frazzon por tudo o que aprendi nesses cinco anos de convivência e aprendizado. Sempre me incentivando e oferecendo oportunidades ímpares, boa parte da minha formação acadêmica e pessoal se devem a ele.

Ao Prof. Dr. João Carlos Setubal agradeço especialmente pelo auxílio nas análises de bioinformática e pela revisão do manuscrito. Seu apoio e sua atenção foram fundamentais para o desenvolvimento deste trabalho.

Ao Prof. Dr. Dennis R. Dean agradeço pela oportunidade única de realizar parte deste trabalho em seu laboratório na Virginia Tech. Tudo o que aprendi e vivenciei lá serão lembrados para sempre.

À Profª. Dra. Ana Paula Guedes Frazzon agradeço por todo o apoio, especialmente pela atenção e pelo carinho durante minha estada em Blacksburg. Muito obrigado ainda ao Rafa e ao Bruno, por me presentearem com o título de irmão mais velho e por tornarem esses três meses na Virginia muito mais divertidos.

Agradeço ainda ao (em breve) Dr. Gustavo Pelicioli Riboldi, o qual tive o prazer de auxiliar como aluno de iniciação científica em seu mestrado e doutorado. Devo a ele todo meu conhecimento prático de laboratório.

À Valerie L. Cash e Kerri K. Martin agradeço pela ajuda fundamental nos experimentos com o RNA não-codificante *arrF*.

Aos membros da Banca Examinadora, Prof. Dr. Rogério Margis e Prof. Dr. Tarso L. Kist, obrigado por aceitarem o convite para avaliar este trabalho.

Finalmente, agradeço e dedico este trabalho à minha família pelo apoio incondicional e por sempre acreditarem em mim. Parafraseando Sir Isaac Newton, se enxerguei mais longe foi porque estava sobre os ombros de gigantes: Marcelo, Christine, Fernando, Cris.

Índice Geral

## Resumo

Cofatores ferro-enxofre [Fe-S] são espécies químicas amplamente distribuídas na natureza, cuja versatilidade habilita-os a participar de inúmeros processos biológicos. Dada a toxicidade de Fe e S livres, os organismos vivos utilizam maquinarias especializadas na biossíntese e no transporte de cofatores [Fe-S]. Há grande interesse na identificação de novas proteínas capazes de atuar em estágios intermediários do metabolismo [Fe-S]; assim, este trabalho investiga o genoma da bactéria *Azotobacter vinelandii*, organismo modelo nesse campo de pesquisa, buscando novos membros dedicados à biologia geral de grupamentos [Fe-S]. O estudo baseia-se em algoritmos de predição de motivos conservados de aminoácidos e na determinação de sítios de ligação ao elemento regulatório [Fe-S] IscR, com o auxílio de matrizes de probabilidade. A aplicação dos recursos acima descritos gerou uma lista completa de proteínas [Fe-S] de *A. vinelandii* com unidades de transcrição estimadas e análises de função, homologia e interações proteína-proteína. Além disso, a montante de dez genes identificou-se motivos de ligação a IscR estatisticamente significativos. Esses elementos foram extensivamente investigados, sete deles codificando polipeptídeos não-caracterizados ou hipotéticos. De modo geral, os resultados obtidos agregam conhecimento ao estudo de cofatores [Fe-S] e propiciam uma nova ferramenta de referência para futuras investigações práticas.

*Cofatores [Fe-S] – Diversidade estrutural e funcional*

Cofatores ferro-enxofre [Fe-S] são espécies químicas amplamente distribuídas na natureza constituídas de ferro não-heme e sulfito inorgânico (1). Associados a mais de 200 proteínas de uma grande variedade de classes, essa substancial diversidade reside em seus aspectos funcionais e estruturais, principalmente no que diz respeito à extrema versatilidade de ambos os elementos ferro e enxofre (2, 3). Por exemplo, a característica desses átomos de perturbar densidades de elétrons (4) faz de agrupamentos [Fe-S] centros especializados no transporte biológico de elétrons, como pode ser observado nas cadeias respiratória e fotossintética assim como nos sistemas de fixação de nitrogênio. Cofatores dedicados a essa função geralmente apresentam-se nas formas [2Fe-2S], [3Fe-4S], [4Fe-4S] ou [8Fe-7S] e são capazes de captar um elétron por ciclo (Figura 1); observa-se, porém, que centros [8Fe-7S] de nitrogenases são capazes de acomodar dois elétrons por ciclo (5).

Além de papéis chave em processos de oxidorredução, cofatores [Fe-S] também formam os sítios de ligação a substratos em muitas enzimas por pelo menos três mecanismos já descritos. Primeiro, um grupamento pode facilitar a ligação e ativação de substratos em reações de hidratação/desidratação, como pode ser visto para a enzima do ciclo dos ácidos tricarboxílicos aconitase (6). A família de enzimas [Fe-S] contendo um radical SAM funciona de modo similar, catalisando a clivagem de S-adenosilmetionina (SAM) para gerar o radical 5'-deoxiadenosil (7, 8). Segundo, a ligação ou ativação do substrato podem exigir a inserção de um heterometal em um grupamento [Fe-S]. Encontra-se nesse grupo o cofator [Ni-4Fe-5S] presente na enzima monóxido de carbono desidrogenase (9). Finalmente, sítios de ligação a substratos podem ser formados por pontes de resíduos cisteinil, exemplificados nos casos dos sítios ativos das enzimas

sulfito e nitrito redutase (10), acetil coenzima A sintase (11, 12, 13) e Fe-hidrogenase (14, 15).

A partir de demonstrações de que cofatores [Fe-S] são capazes de controlar estruturas proteicas via perturbações de curto alcance em cadeias polipeptídicas (16, 17), muitas proteínas [Fe-S] vêm sendo descritas como elementos regulatórios. Esse é o caso das enzimas de reparo de DNA endonuclease III (18, 19) e MutY, do mesmo modo que as proteínas de interação ambiental SoxR and FNR. Por exemplo, quando células sofrem estresse oxidativo o cofator $[2Fe\text{-}2S]^{2+}$ de SoxR é oxidado, o que estimula a expressão de SoxS, um ativador transcricional de um grande grupo de genes de resposta a estresse oxidativo (20). A proteína FNR (fumarate and nitrate reduction), por sua vez, participa nas interações metabólicas necessárias a fim de que haja a troca entre os estados aeróbio e anaeróbio em *Escherichia coli*. Isso ocorre pela oxidação de um grupamento $[4Fe\text{-}4S]^{2+}$ dimérico, ligado ao DNA, a uma forma $[2Fe\text{-}2S]^{2+}$ monomérica (21). Dessa forma, é possível apreciar, em parte, a vasta gama de processos nos quais cofatores [Fe-S] tomam parte. A partir de estruturas químicas relativamente simples evidenciam-se cada vez mais estratégias metabólicas nas quais grupamentos formados por ferro e enxofre são capazes de atuar.

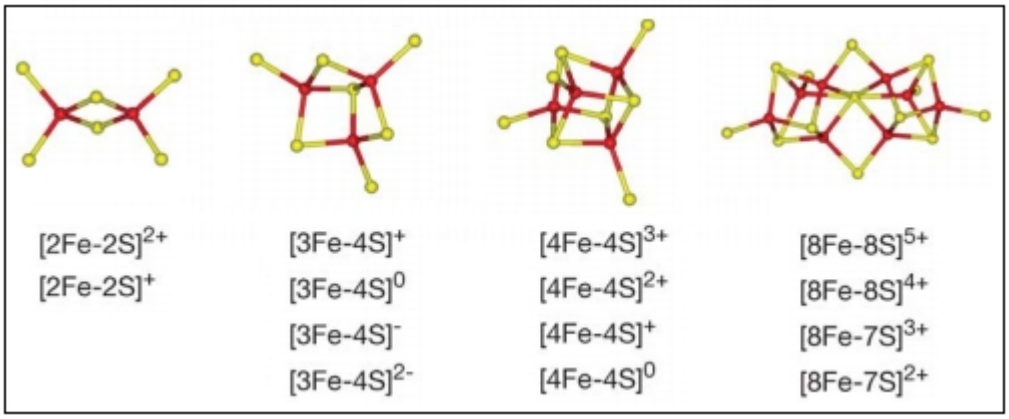



Figura 1: Principais tipos de cofatores [Fe-S] encontrados em proteínas e seus estados de oxidação mais comuns. Esferas vermelhas representam átomos de ferro; esferas amarelas denotam átomos de enxofre. Figura modificada de (1).

*Cofatores [Fe-S] – Formação e sistemas de biossíntese*

Há muitos anos sabe-se que certas apo-proteínas aceptoras de cofatores [2Fe-2S] e [4Fe-4S] podem ser ativadas *in vitro* pela simples adição de $S^{2-}$ e $Fe^{2+/3+}$ (22). Apesar de essas observações indicarem que agrupamentos de ferro e enxofre podem ser incorporados em proteínas espontaneamente, tal hipótese foi rapidamente descartada devido à toxicidade dos elementos Fe e S nas concentrações necessárias a esse processo. Uma possibilidade mais interessante foi então desenvolvida, em que Fe e S complexados seriam transportados no meio intracelular por proteínas carreadoras específicas, que promoveriam a inserção dos grupamentos nas apo-proteínas [Fe-S], contornando assim o dano tóxico.

A primeira evidência de um mecanismo de trânsito e regulação de cofatores [Fe-S] surgiu com trabalhos no campo da fixação biológica de nitrogênio, mais precisamente com a enzima nitrogenase. Nitrogenase é uma enzima [Fe-S] complexa que cataliza a redução nucleotídeo dependente de dinitrogênio (23). Suas duas subunidades – Fe proteína e MoFe proteína – não são completamente ativas assim que traduzidas, necessitando de um grande conjunto de genes acessórios para a inserção dos grupamentos ferro e ferro-molibdênio associados. Esses elementos, intitulados genes específicos à fixação de nitrogênio (*nif, nitrogen-fixation specific*) foram inicialmente identificados nas proteobactérias *Klebsiella pneumoniae* (24) e *Azotobacter vinelandii* (25, 26). Dois genes em especial lançaram as bases dos estudos de biossíntese de cofatores [Fe-S]; observou-se que a deleção de *nifS* ou *nifU* acarretava na perda de atividade de ambas as subunidades de nitrogenase (25). Como a característica comum às duas cadeias polipeptídicas é a presença de grupamentos metálicos, postulou-se que *nifS* e *nifU* deviam participar em processos iniciais específicos de montagem e inserção de cofatores [Fe-S] na maquinaria de fixação de nitrogênio. Interessantemente, observou-se na época que mesmo com duplos mutantes *nifS*/*nifU* as atividades tanto de Fe

proteína quanto de MoFe proteína não são completamente eliminadas, sugerindo que outras proteínas são capazes de complementar a função de NifS e de NifU em níveis baixos.

Estudos detalhados identificaram NifS como uma enzima homodimérica contendo um cofator piridoxal-fosfato (PLP) (27). A proteína catalisa a retirada de enxofre de L-cisteína, gerando L-alanina e é expressa somente em condições de fixação de nitrogênio. Além disso, uma O-acetilserina sintase, enzima que promove a catálise do passo limitante da biossíntese de cisteína (gene *cysE*), é co-transcrita com *nifS* (28).

NifU revelou-se como uma proteína arcabouço em que os agrupamentos [Fe-S] podem ser devidamente montados e destinados para a nitrogenase. A proteína funcional é um homodímero com um cofator [2Fe-2S]$^{2+/+}$ em cada monômero (29). Sua estrutura é composta de três domínios altamente conservados conectados por duas sequências de menor conservação (30, 31). O domínio central contém um grupamento [2Fe-2S] permanente que pode ter papel na maturação de cofatores em formação; os domínios N- e C-terminal suportam agrupamentos transientes que podem ser passados a aceptores finais (32, 33). Ainda, o domínio C-terminal é homólogo a um grande grupo de proteínas, denominado Nfu, que abrange desde bactérias até organismos superiores e que pode aceitar cofatores [2Fe-2S]2+ e [4Fe-4S]2+ lábeis (34, 35, 36).

Apesar do sistema Nif ter sido descoberto em um organismo fixador de nitrogênio, ele não é restrito a essa classe de seres vivos, já que proteínas homólogas em arranjos gênicos similares são encontradas em organismos que não fixam nitrogênio. Um exemplo é *Helicobacter pylori*, em que há evidência de um tipo de sistema Nif é necessário para a maturação das proteínas [Fe-S] da bactéria (37). O único eucarioto que parece codificar um sistema Nif é *Entamoeba histolytica* e um estudo conseguiu complementar linhagens de *E. coli* com genes *isc* e *suf* (outras maquinarias de biossíntese de agrupamentos [Fe-S], descritas a seguir) deletados

utilizando os genes ortólogos de *E. histolytica* (38).

Devido ao achado de que se identificam baixos níveis de atividade de nitrogenase em condições de fixação de nitrogênio mesmo quando *nifS* e *nifU* são deletados (25), surgiu um grande interesse em caracterizar possíveis novos genes envolvidos na biossíntese de cofatores [Fe-S]. Uma vez que a atividade remanescente de nitrogenase nas linhagens duplo mutantes *nifS*/*nifU* era baixa, postulou-se que haveria uma classe de genes com expressão basal dedicada à formação de complexos [Fe-S]. Desse modo, buscou-se isolar uma proteína com atividade cisteína desulfurase em extratos de *A. vinelandii* em condições de não fixação de nitrogênio (39). A investigação da sequência polipeptídica de uma proteína identificada de acordo com os pré-requisitos possibilitou o uso de uma estratégia de análise da região genômica em que a mesma era codificada. Esse procedimento descobriu nove genes ligados e hoje nomeados como *cysE2*, *iscR*, *iscS*, *iscU*, *iscA*, *hscB*, *hscA*, *fdx*, e *iscX* (*isc*: *iron-sulfur-cluster formation*; *hsc*: *heat-shock-cognate*).

A investigação detalhada das sequências gênicas do operon *isc* revelaram informações importantes a respeito do metabolismo de proteínas [Fe-S]. A localização de um gene parálogo a *cysE*, chamado de *cysE2* indicou que dentro desse novo sistema de biossíntese de cofatores [Fe-S], a regulação da biossíntese de L-cisteína também é um passo essencial. Contudo, sabe-se hoje que não são todos organismos que possuem essa característica de associação do metabolismo de L-cisteína com a formação de complexos [Fe-S].

O gene *iscR* é englobado em uma grande família de fatores de transcrição e funciona como um inibidor da expressão do operon isc (40, 41). De acordo com o modelo proposto, um grupamento [2Fe-2S] presente em IscR atua como sinalizador para a inibição da síntese de proteínas Isc. Assim, a versão holo-IscR indicaria à célula uma saturação de cofatores [Fe-S], não sendo necessária a expressão de novas proteínas da maquinaria *isc*

de biossíntese [Fe-S]. Contudo, a versão apo-IscR, refletindo uma carência celular de centros [Fe-S], seria incapaz de reprimir a expressão de *isc*, fazendo com que, em ultima análise, novos agrupamentos sejam formados. Essa visão é suportada pela baixa atividade de IscR em linhagens deficientes na biossíntese de cofatores [Fe-S] (40).

IscU é uma proteína homóloga à porção N-terminal de NifU, enquanto IscA é bastante similar ao produto do gene a montante (upstream) de *nifU*, denominado IscANif (42), que até o momento da descoberta de IscA não havia sido associado à função de maturação da nitrogenase. IscA e IscANif fazem parte de uma vasta família de proteínas conhecida como carreadores do tipo A (ATC, A-type carriers) (43), sendo composta por proteínas de biossíntese [Fe-S] dedicadas à montagem e ao trânsito de grupamentos.

Outro aspecto interessante do operon *isc* é a presença de uma ferrodoxina [2Fe-2S]. De fato, alguns trabalhos sugerem que a semelhança biofísica entre seu cofator associado e aquele presente no domínio central de NifU possa apontar para uma função comum (29, 44).

IscX talvez seja o membro menos conhecido do operon. Estudos *in vitro* mostraram que essa proteína é capaz de se complexar a IscS (45, 46), sugerindo um papel modulatório a esse pequeno polipeptídeo acídico. A primeira estrutura cristalográfica de IscX revelou um possível motivo hélice-volta-hélice (HTH, helix-turn-helix) típico de proteínas de ligação ao DNA (47). Em princípio, essa observação levou à hipótese de que IscX poderia ser um novo regulador de transcrição do operon *isc*. Contudo, a sua estrutura de superfície altamente negativa, determinada por ressonância magnética nuclear (48), e seu padrão de conservação sugerem que essa similaridade a um motivo HTH poderia ser uma adaptação a uma nova função. IscX ainda é capaz de ligar átomos de $Fe^{2+}$ e $Fe^{3+}$ na mesma superfície de ligação com IscS (48), o que faria de IscX um adaptador molecular que modularia a interação de IscS com parceiros ainda não-identificados.

Com relação a HscB e HscA, esses elementos já haviam sido previamente identificados em *Escherichia coli* e são homólogos aos genes codificadores das chaperonas moleculares DnaJ e DnaK, respectivamente (49, 50, 51, 52, 53). O papel dessas proteínas na biossíntese de cofatores [Fe-S] só começou a ser elucidado após observações de conservação genética de *iscS*, *iscU*, *iscA*, *hscB*, *hscA*, e *fdx*. Ao longo do tempo, abordagens bioquímicas e genéticas estabeleceram a associação direta das duas chaperonas com o sistema de biossíntese Isc (54, 55, 45, 46). HscA é capaz de interagir com IscU tanto na forma ligada a um grupamento [2Fe-2S] – holo-IscU – quanto na forma não-ligada – apo-IscU (56, 57, 58). Essa interação é auxiliada por HscB, que faz contatos com as duas outras proteínas. HscA ainda apresenta intrinsecamente um baixo nível de atividade ATPase que é estimulada pela interação com HscB e ainda mais acentuada quando complexada com HscB e IscU. Fundamental nesse processo é uma pequena sequência de aminoácidos em IscU extremamente conservada (LPPVK), que é suficiente para estimular a hidrólise de ATP de pendente de HscA (57). Apesar de grandes avanços no sentido do entendimento das interações de HscB e HscA, ainda não estão claros seus reais papéis na maturação de proteínas [Fe-S]. Mostrou-se por imunoprecipitação em leveduras, que a depleção de ambas as chaperonas associadas ao sistema Isc acarretam um acúmulo de ferro em IscU (59). Argumentou-se que frente a esse achado as chaperonas seriam necessárias para realizar a transferência de cofatores [Fe-S], mas nenhum mecanismo para tal foi evidenciado.

Outros trabalhos, avaliando a inativação de genes individuais e caracterizando o impacto sobre proteínas [Fe-S] e proteínas sem a requisição de agrupamentos (60, 45), comprovaram os drásticos efeitos no crescimento celular em linhagens inativadas para *iscS*, *iscU*, *hscB*, *hscA* ou *fdx*. Entretanto, os fenótipos observados para a inativação de *iscA* foram bem menos intensos, gerando células com uma capacidade muito maior de

reprodução do que os demais mutantes. Frente a esses resultados, a descoberta de que linhagens de *E. coli* inativadas para *iscS* ou *iscU* permaneciam viáveis foi bastante inovador. Adiciona-se à importância desse achado o fato de que as mesmas mutações em *A. vinelandii* são letais. Essas diferenças fenotípicas e experimentais começaram a ser elucidadas pela identificação em *E. coli* (e ausência em *A. vinelandii*) de um novo operon dedicado à biossíntese de cofatores [Fe-S] – o operon *sufABCDSE*.

Os genes *suf* foram inicialmente descritos pela desestabilização do grupamento [2Fe-2S] de FhuF em mutantes para *sufD* ou *sufS* (61). Acredita-se que FhuF seja uma redutase férrica que utiliza ferroxamina B ou ácido rodotorúlico como fontes de ferro (62). Essas evidências primárias também mostraram que a expressão do operon *suf* é regulada pelo repressor dependente de ferro Fur, sendo então induzida em condições de deficiência desse elemento. Novas contribuições mostraram que os genes *suf* são ainda transcritos em condições de estresse oxidativo e pertencem ao regulon do elemento responsivo a estresses oxidativos OxyR (63). Ainda, a regulação da expressão do sistema Suf se dá também pela interação com IscR (64). IscR, não sua forma não associada a um grupamento [2Fe-2S] é capaz de se ligar a um sítio de regulação a montante do sítio de início de transcrição do operon *suf* e, juntamente com OxyR, é capaz de induzir a expressão de proteínas Suf. Entende-se também que IscR é um elemento bastante importante na regulação dos operons *isc* e *suf* em organismos patogênicos, no advento da interação com hospedeiros. Por exemplo, IscR tem papel central na capacidade da bactéria patógena de plantas *Erwinia chrysanthemi* infectar diferentes tipos de vegetais (65). Interessantemente, cianobactérias codificam em seus genomas uma proteína homóloga a IscR, a montante de um operon *sufBCDS* (66). Intitulada SufR, essa proteína comporta um grupamento [4Fe-4S] e é capaz de inibir a expressão dos genes *sufB*, *sufC*, *sufD* e *sufS* (66). Visto que cianobactérias são providas apenas do sistema Suf para a biossíntese de seus cofatores [Fe-S], talvez

essa seja a evidência de uma modificação de IscR para um novo sistema, entretanto mantendo sua função básica regulatória.

A proteína SufS foi caracterizada como uma cisteína desulfurase e provou-se a evidente similaridade de sequências entre SufA e IscA, inclusive no que diz respeito às três cisteínas conservadas. Apesar de essas observações indicarem um papel claro do operon *suf* no metabolismo [Fe-S], mutantes *suf* em *E. coli* não mostraram nenhum fenótipo claro em condições normais de crescimento; somente através de estudos mutacionais de genes *suf* de *E. chrysanthemi* foi possível obter maiores informações a respeito desse sistema. Mais precisamente, disrupções dos genes individuais do operon *suf* resultaram em níveis aumentados de ferro intracelular (67). Os fenótipos mais relevantes foram obtidos com a inativação de *sufC*, incluindo sensibilidade aumentada a agentes oxidantes, habilidade reduzida de infecção de plantas, atividade diminuída de enzimas contendo cofatores [Fe-S] facilmente oxidáveis e perda da capacidade de assimilar ferro pelo sideróforo férrico crisobactina (67, 68). Notavelmente, esses fenótipos só foram observados em condições de estresse e sugeriram que o operon *suf* estaria dedicado à ativação, proteção ou reparo de proteínas [Fe-S] em condições de estresse oxidativo ou limitação de ferro.

SufS e SufE formam um complexo  em que SufE aumenta a atividade de cisteína desulfurase de SufS em até 50 vezes (69, 70). Além disso, a adição do complexo SufBCD à reação possibilita um aumento adicional na atividade de SufS, sugerindo um efeito sinergístico de SufE e SufBCD na modulação de SufS (70). O componente SufC apresenta atividade ATPase intrínseca e, como descrito anteriormente, forma um complexo solúvel bastante estável com SufB e SufD (68, 70, 71). De fato, SufBCD é similar a transportadores do tipo ABC (ATP-binding cassette), caracterizados por sua localização transmembrana. A demonstração da exclusividade citosólica de SufBCD foi a primeira evidência de um transportador tipo ABC de localização não membranar (68). A função exata do complexo SufBCD

permaneceu incerta a até pouco tempo atrás, quando se provou que SufE é capaz de interagir com SufB e promover a transferência de enxofre para o complexo SufBCD por quatro cisteínas conservadas em SufB (72). Esse trânsito de enxofre só ocorre se SufC também está presente. Esse achado lançou uma nova luz na função do complexo SufBCD, uma vez que SufB pode servir como um sítio de montagem de cofatores [Fe-S] do sistema Suf. Corroborando com essa evidência, encontram-se bactérias e arqueobactérias com proteínas SufBCD ou SufBC preditas, mas sem homólogos das proteínas arcabouço do tipo A ou tipo U (73).

De modo geral, o operon *suf* aparenta ser o sistema de biossíntese de grupamentos [Fe-S] mais heterogêneo em termos de presença ou ausência de componentes. Enquanto *E. coli* codifica os genes *sufABCDSE* (73), observa-se em cianobactérias somente a presença de *sufBCDS* e do elemento regulatório *sufR* (66). Já em diversos outros grupos bacterianos foram identificados genes *suf* homólogos a *iscU* em operons do tipo *sufBCDSU*, como é o caso dos procariotos Gram-positivos, do filo *Firmicutes*, *Bacillus subtilis* (74) e *Enterococcus faecalis* (75).

Comparativamente, entende-se que as maquinarias de biossíntese de cofatores [Fe-S] descobertas até hoje compartilham características básicas e essenciais ao processo biológico de incorporação de Fe e S em grupamentos estáveis. Primeiro, os três sistemas – Nif, Isc e Suf – codificam cisteínas desulfurases capazes de ceder enxofre elementar à montagem de centros [Fe-S]. Segundo, Nif, Isc e Suf possuem pelo menos uma proteína com propriedade de arcabouço molecular, em que grupamentos são devidamente formados os transportados sem oferecer danos à célula e/ou sem sofrer processos oxidativos (Figura 2).

Nos últimos anos, há um grande interesse em se desvendar novas proteínas que participem principalmente nos estágios intermediários do metabolismo de cofatores [Fe-S]. Tais elementos seriam dedicados ao transporte de grupamentos dos sistemas de biossíntese para proteínas

aceptoras finais, como as proteínas [Fe-S] da cadeia respiratória. Dentro dessa classe, pelo menos dois exemplos bem caracterizados podem ser citados. Dois grupos independentes relataram ao mesmo tempo, em *A. vinelandii* (76) e *E. coli* (77), a capacidade de uma proteína homóloga à região C-terminal de NifU ligar e transferir cofatores [4Fe-4S] para apo-proteínas. Essa molécula, intitulada como NfuA, é codificada por um gene isolado nos genomas de ambas as bactérias que não está diretamente ligado a nenhum operon de biossíntese de cofatores [Fe-S]. O outro exemplo diz respeito a ApbC, uma proteína cuja função foi desvendada no patógeno Salmonella enterica (78, 79, 80). O fato de que mutações em *apbC* levam a defeitos no metabolismo [Fe-S] (78) instigou os pesquisadores a postular um papel para ApbC na maturação de proteínas [Fe-S]. De fato, ApbC também é capaz de ligar um cofator [4Fe-4S] e transferi-lo *in vitro* para apo-isopropilmalato isomerase (Leu1; 80), segunda enzima do metabolismo da leucina, conhecida por conter um grupamento [4Fe-4S] em sua forma ativa.

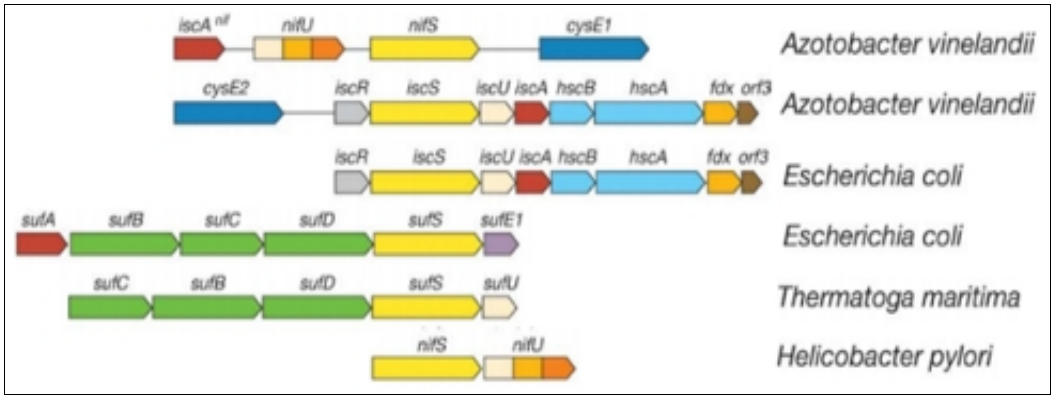



Figura 2: Representação de alguns sistemas de biossíntese de cofatores [Fe-S] já identificados. Genes com a mesma cor codificam proteínas com funções similares. *A. vinelandii* possui os sistemas nif e isc, enquanto *Escherichia coli* conta com as maquinarias isc e suf. *Thermatoga maritima* codifica apenas um sistema tipo suf; já Helicobacter pylori possui uma variação do sistema nif.

*Cofatores [Fe-S] – Azotobacter vinelandii como organismo modelo*

Por mais de um século a gamaproteobactéria *A. vinelandii* vem sendo usada por cientistas nos mais diversos desdobramentos bioquímicos e genéticos. Desde os primeiros ensaios de determinação dos parâmetros cinéticos de enzimas na década de 1930 (81), passando pelos estudos de elucidação do código genético (82), *A. vinelandii* sempre se mostrou um modelo interessante. Entretanto, o principal foco no manejo dessa bactéria nas últimas décadas foi depositado no entendimento da fixação biológica do nitrogênio.

A habilidade de expressar três diferentes nitrogenases, reguladas independentemente de acordo com a disponibilidade de metais no meio (83), em um organismo estritamente aeróbio (84), faz de *A. vinelandii* uma das poucas bactérias com um mecanismo tão especializado de fixação de nitrogênio. A partir disso, a demonstração da participação de um sistema de biossíntese de grupamentos [Fe-S] dedicado à maturação de nitrogenases (25, 26) expandiu a visão da época sobre o metabolismo do nitrogênio e lançou as bases do estudo da formação biológica de cofatores [Fe-S]. Hoje, *A. vinelandii* é fonte de profundas investigações nessa área, sendo ainda, geralmente, utilizado em ensaios de complementação quando outros organismos são estudados com o mesmo propósito.

Redação de manuscrito científico*

# Exploring [Fe-S] cluster genomic data in Azotobacter vinelandii

Eduardo P Mattos[1], João C Setubal[2], Dennis R Dean[3], Jeverson Frazzon[1,3§]

[1]Instituto de Ciência e Tecnologia de Alimentos, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

[2]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

[3]Department of Biochemistry, Virginia Polytechnic Institute and State University, Fralin Life Science Institute, Blacksburg, VA, USA

[§]Corresponding author

Email addresses:

EPM: edu_p_m@hotmail.com

JCS: setubal@vbi.vt.edu

DRD: deandr@vt.edu

JF: jfrazzon@vt.edu

*Formatação nos padrões do periódico BMC Genomics.

# Abstract

### Background

Iron-sulfur [Fe-S] clusters are widely distributed versatile chemical species that act in several biological processes. Since free intracellular Fe and S are toxic elements, living organisms have developed specialized biosynthetic machineries specific for [Fe-S] cluster formation and transport. A growing number of efforts are now being deposited on the identification of new proteins capable of performing central tasks in the [Fe-S] cofactors metabolism. Therefore, this work investigates *Azotobacter vinelandii DJ*'s genome, searching new members dedicated to [Fe-S] biology. The study is based on prediction algorithms of conserved amino acid motifs and a position-specific scoring matrix for the discovery of DNA IscR-binding sites.

### Results

The application of these techniques generated a complete list of known and predicted *A. vinelandii DJ* [Fe-S] proteins, analyzed in terms of function, homology and protein-protein interactions. Besides that, in regions upstream ten *A. vinelandii DJ* genes IscR-binding sites were significantly identified; proteins coded by these elements were also extensively investigated.

### Conclusions

Taken together, the results presented here aggregate data on experimentally characterized and *in silico* predicted *A. vinelandii DJ* [Fe-S] proteins. In this way, the compiled database, added to the proposed IscR-regulated genes, can be used by researchers to further investigate new metabolic features driven by [Fe-S] cofactors and expand the current understanding on [Fe-S] cluster biosynthesis and transfer to terminal proteins.

# Background

Iron-sulfur [Fe-S] clusters are widely distributed chemical species constituted of non-heme iron and inorganic sulfite, that act in several biological processes [1]. Considering the extended versatility of both Fe and S elements, more than 200 proteins have already been characterized in which [Fe-S] cofactors are essential for polypeptide stability and/or function [1, 2, 3]. For years it is known that [Fe-S] clusters can spontaneously assemble *in vitro* by the simple addition of $S^{2-}$ and $Fe^{2+/3+}$ [4]. This knowledge led to the hypothesis that [Fe-S] species could be directly inserted into terminal proteins; however, specially due to the toxicity of free intracellular Fe and S elements, this postulation had to be revisited. A more feasible explanation to the process of [Fe-S] clusters formation was achieved with the discovery of the nitrogen-fixation specific system (nif system) in *Klebsiella pneumoniae* [5] and *Azotobacter vinelandii* [6, 7], a biochemical machinery primarily devoted to nitrogenase maturation. Two genes encoded within the *nif* operon have established the principles of [Fe-S] cluster assembly investigation: *nifS*, coding for a cysteine desulfurase and *nifU*, coding for a molecular scaffold protein [6]. Since then, two other systems devoted to [Fe-S] cofactors biosynthesis were identified and have been extensively characterized, namely the *iron-sulfur-cluster formation* (*isc*) operon and the *sulfur mobilization system* (*suf*) gene cluster. The three biosynthesis machineries share homologous proteins that can, at least in part, complement each other functions; at the same time, the nif, isc and suf systems have specific elements thought to perform essential processes in the cells encoding them. One of such members is the *isc* encoded gene *iscR*. *iscR* gives rise to a MarA-like protein harboring a [2Fe-2S] cluster that was shown to bind to a IscR-binding site

and regulate the expression of the *isc* operon [8]. IscR acts as a cellular sensor of [Fe-S] cluster cellular pool and represses the *isc* genes when bound to a [2Fe-2S] center. This was the first evidence of a [Fe-S] cluster-driven regulatory mechanism and further experiments have established a role for IscR in the regulation of the suf system [9, 10]. Information on other biological processes affected by IscR is still limited, the best characterized examples being found in *Escherichia coli* [11]. Besides IscR, a growing number of efforts are now being deposited on the identification of new proteins capable of performing central tasks in the [Fe-S] cofactors metabolism. Therefore, the present work investigates *A. vinelandii DJ*'s genome, a model organism in the [Fe-S] clusters research field, searching new members dedicated to [Fe-S] biology. The study is based on prediction algorithms of conserved amino acid motifs and a position-specific scoring matrix for the discovery of DNA IscR-binding sites. The application of these techniques generated a complete list of known and predicted *A. vinelandii DJ* [Fe-S] proteins, analyzed in terms of function, homology and protein-protein interactions. IscR-binding sites upstream ten *A. vinelandii DJ* genes were significantly identified; proteins coded by these elements were also extensively investigated. Taken together, the results presented here aggregate data on experimentally characterized and *in silico* predicted *A. vinelandii DJ* [Fe-S] proteins. In this way, the compiled database, added to the proposed IscR-regulated genes, can be used by researchers to further investigate new metabolic features driven by [Fe-S] cofactors and expand the current understanding on [Fe-S] cluster biosynthesis and transfer to terminal proteins.

# Results and discussion

### [Fe-S] proteins organization and distribution

Extensive bioinformatics analysis have allowed us to compile a large list of known and putative [Fe-S] proteins encoded in the diazotrofic organism *A. vinelandii DJ*. Here we comment some of these results, presenting data that support the findings and hypothesizing on the function of some uncharacterized members that could be involved in [Fe-S] metabolism.

*1.1) Electron carrier systems*

*1.1.1) NADH-ubiquinone oxidoreductases*

Based on the observation that oxidation/reduction reactions are a classical example of function devoted to [Fe-S ] clusters, it is not surprise to find oxidoreductases exploring the potential of [Fe-S] centers to mobilize electrons. *A. vinelandii DJ* is not an exception to this rule and three NADH-ubiquinone oxidoreductases encoding [Fe-S] subunits can be identified upon analysis of its genome (Table 1).

*Avin28440* to *Avin28560* represents the ATP-coupled NADH-ubiquinone oxidoreductase of the Nuo type, with NuoB, NuoF and NuoI (Avin28450, Avin28480 and Avin28510, respectively) identified as [4Fe-4S] proteins, while NuoE and NuoG (Avin28470 and Avin28490, respectively) identified as [2Fe-2S] proteins. The other two oxidoreductases are ATP-uncoupled and comprise the cytosolic NADH-uibiquinone oxidoreductase Ndh (Avin12000) and the membrane-associated NADH-ubiquinone oxidoreductase involved in the transport of $Na^+$ Nqr (*Avin14590* to *Avin14640*). From the six subunits that form the Nqr enzyme, NqrF (Avin14640) is an iron-sulfur protein of the [2Fe-2S] type. On the other hand, Ndh (Avin12000)

harbors a [4Fe-4S] in its functional structure.

*1.1.2) Rnf systems*

As previously reported [29], *A. vinelandii DJ* has two copies of the electron transport complex Rnf, namely Rnf1 (*Avin50930* to *Avin50980*) and Rnf2 (*Avin19220* to *Avin19270*), an ortholog of Rfn1 that shows sequence similarity to a $Na^+$-dependent NADH-ubiquinone oxidoreductase [30]. In both cases, subunits B and C (Avin50960 and Avin50970, for Rnf1; Avin19250 and Avin19260, for Rnf2) have typical signatures for [4Fe-4S] clusters ligation (Table 2). Rnf1 complex is required for accumulation of nitrogenase Fe protein [30], while the function of the ortholog system is less clear. Evidence for a Rnf2-like complex is found in the nitrogen-fixing bacteria *Pseudomonas stutzeri A1501* [29], the closest relative to *A. vinelandii DJ*, but not in other *Pseudomonas* species.

*1.1.3) Terminal oxidases*

Downstream on the process of electron mobilization, ubiquinol-oxigen oxido-reduction can follow two distinct pathways: (a) electrons can be submitted to a cytochrome c reductase (complex III) followed by a cytochrome terminal oxidase (complex IV), or (b) electrons can be directly processed by a single-step ubiquinol-dependent cytochrome terminal oxidase [29]. *A. vinelandii DJ* carries both systems, encoding a complex III and five terminal oxidases.

The cytochrome c reductase (*Avin13060* to *Avin13080*; Table 3) exhibits a Rieske-type [2Fe-2S] subunit, PetA (Avin13060). In respect to the terminal oxidases, the only [Fe-S] cluster-containing machinery observed was the cytochrome c oxidase $cbb_3$ (Cco; *Avin19940* to *Avin20010*). The *cco* operon, conserved in all *Pseudomonas*

species analyzed to date [29], encodes four genes whose products are structural subunits of the terminal oxidase and four other genes coding for accessory and maturation proteins (see Table 3). The accessory protein CcoG (Avin19970) is the only predicted [Fe-S] member from the complex, with a signature for [4Fe-4S] cluster binding.

*1.1.4) Carbon monoxide dehydrogenase*

*A. vinelandii DJ* contains a set of genes predicted to encode a membrane bound Ni-dependent carbon monoxide dehydrogenase and accessory proteins (CODH; *Avin04450* to *Avin04500* and *Avin47010*). The type of CO metabolism seen in homologous Ni-dependent CODHs is typical of anaerobic organisms and is in contrast to *A. vinelandii DJ* obligatory aerobic life style [29]. The enzyme is characterized by an α-homodimer of CooS (Avin04490; Table 4) that catalyzes the reversible oxidation of CO to $CO_2$ [31]. Electrons acquired through this process are transferred to CooF (Avin04500), a hydrophobic [4Fe-4S] protein similar to oxidoreductases. The products of the other genes of the predicted *coo* operon are likely to code for maturation and accessory proteins. This might be the case of Avin04480 (see Table 4), a possible flavin adenine dinucleotide oxidoreductase that could activate the C cluster [$NiFe_4S_4$] of CODH through a [4Fe-4S] center [29].

*1.1.5) Formate dehydrogenase*

Formate dehydrogenases (FDHs) are a diverse group of enzymes from prokaryotes and eukaryotes that promote the oxidation of formate to $CO_2$ and $H^+$. *A. vinelandii DJ* encodes a $NAD^+$-independent FDH (*Avin03800* to *Avin03840*) [29] with three genes whose products assemble in the active enzyme and two genes for

accessory proteins (Table 5). The β subunit of FDH, FdhH (Avin03820), contains a [4Fe-4S] cluster and interacts with FdhG (Avin03810), the alpha subunit. FdhG has a molybdopterin-guanine dinucleotide cofactor in its active site to extract electrons from formate. The other [4Fe-4S] protein encoded in the *fdhDGHIE* gene cluster is FdhE (Avin03840; see Table 5), an accessory component necessary to assembly of FDH.

*1.2) Thioredoxins*

Thioredoxins comprehend a large family of small proteins of about 12 kDa with a conserved active site Cys-Gly-Pro-Cys dedicated to the catalysis of many redox reactions [32]. Widespread among all living organisms, thioredoxin-like proteins, in the reduced state, promote the reduction of exposed disulfides [33] and can be of great advantage to the cell. Some of the roles that have been attributed to thioredoxins are (a) facilitating agents on disulfide-containing protein refolding [34]; (b) antioxidants capable of reducing hydrogen peroxide [35], scavenge free radicals [36] and protect cells against oxidative stress [37].

*A. vinelandii DJ* seems to encode three thioredoxin proteins distantly located from each other (Table 6). Avin47420 corresponds to the well characterized TrxA protein from *E. coli* – 65% identity [38, 39, 40] – while Avin45140 is 38% identical to *E. coli* thioredoxin 2, TrxC [33]. Indeed, it has the conserved CGPC motif and the four N-terminal conserved residues known to be important to protein function and stability [33] (Figure 1). The third thioredoxin-like gene, *Avin06670*, gives rise to an uncharacterized product similar to the E. coli TrxA protein, although the CGPC motif is substituted for a CAPC motif. Despite this difference, proteins with high degrees of similarity to *Avin06670* are found in Pseudomonads, suggesting that these species

might have a third functional thioredoxin protein.

The prediction of [2Fe-2S] clusters in TrxA (Avin47420) and Avin06670 was based on previous observations that a slight variant from E. coli TrxA is able to form dimers due to cluster ligation [41, 33]. These TrxA proteins with CACA motifs were shown to harbor one [2Fe-2S] per dimer and catalyze the formation of disulfide bonds [41]. After interpreting the data from crystallization of TrxA(CACA) thioredoxin [33], the authors indeed suggest that dimerization by cluster ligation could happen to the native protein, although this was no yet observed. Following this line of thought, we expect [2Fe-2S] centers in both *A. vinelandii DJ* TrxA and Avin06670, an uncharacterized thioredoxin-like protein.

*1.3) Glutaredoxins*

Glutaredoxins (Grxs) are small proteins traditionally associated to glutathione (GSH)-disulfide oxidoreductions. The first evidence of such functions was identified in *Escherichia coli*, in which Grx was shown to donate electrons for ribonucleotide reductase on a GSH-dependent fashion [42]. Since then, diverse new types of Grxs were identified and sorted in specific classes according to their structure and active site motifs [43]. Regarding these informations, there is now compelling information to associate specific Grx isoforms with a role in the [Fe-S] cluster metabolism.

Despite not generally being encoded in the [Fe-S] cluster biosynthesis operons – *nif*, *isc* and *suf* – various works have characterized class II glutaredoxins that are able to ligate [Fe-S] clusters. Classes I and II are the dominant Grx isoforms present in prokaryotic organisms [43] and class II, specifically, comprises all Grxs with a CGFS active site, as it is seen in *Saccharomyces cerevisiae* Grx5 and *E. coli* Grx4 proteins. Thorough experimentation with *S. cerevisiae* Grx5 have established a

clear association between class II Grxs and [Fe-S] cluster assembly [44, 45, 46, 47, 48]. Decisive on this issue were the demonstrations that class II Grxs from both prokaryotes and eukaryotes, when targeted to the mitochondrial matrix, can effectively complement a yeast *grx5* deletion strain [49]. Besides that, class II Grxs from different organisms were identified as containing a labile [2Fe-2S] cluster [49, 50, 51]. On the other hand, the role of class I Grxs in [Fe-S] cluster biosynthesis is less clear. Isoforms grouped in this class show a relative diversity on the structure of the active site, such as CPY[C/S], CGYC, CPFC and CSY[C/S] [43]. Characterization of [Fe-S] cluster-bound class I Grxs is so far observed in eukaryotes, as happens to human GRX2 – CSYC active site – [50, 52, 53, 54]. In these cases, the conjugated [2Fe-2S] clusters are thought to act as redox sensors in normal or oxidative stress conditions [43].

*A. vinelandii DJ* encodes three glutaredoxin genes in its genome. From the products of these genes, Avin14040 shares 72% identity to *E. coli* Grx4 and it is likely to form holodimers via a [2Fe-2S] cluster interaction (Table 7). The second Grx identified, Avin04690, although it is 53% identical to *E. coli* Grx4, contain a CPYC motif typical of class I dithiol-glutaredoxins. It is possible that this protein can bind [2Fe-2S] clusters to some extend, since class I Grxs with this type of [Fe-S] center have been characterized. Avin04690 is predicted to be encoded in an operon with a rhodanese-like protein and with the protein-export chaperone SecB (Avin04680 and Avin04700, respectively; see Table 7). There is some evidence to discard a possible interaction between Avin04690 and SecB [55], however the association of a glutaredoxin homolog (Avin04690) with a rhodanese-like protein (Avin04680) is an interesting feature of this operon that should be further evaluated. Finally, the *nif*-associated glutaredoxin (Avin51060) is the best understood Grx of *A.*

*vinelandii DJ*. Inactivation of the *Avin51060* gene showed striking negative effects on both MoFe protein and Fe protein, the two components of MoFe nitrogenase [29]. In fact, Fe protein exhibited a 50% loss in its activity and these effects were thought to occur due to oxygen sensitivity and/or impair on [Fe-S] cluster reconstitution once *Avin51060* was disrupted. The fact that a glutaredoxin capable of binding [Fe-S] clusters, added to the observed phenotypes upon its inactivation, is encountered in the *nif* operon suggests that this protein can bind [4Fe-4S] cofactors instead of [2Fe-2S]. Indeed, a recent review on glutaredoxins [43] discussed possible new Grxs involved in iron or molybdenum metabolism, as well as putative *nif*-associated glutaredoxins from nitrogen-fixing bacteria encoded in systems very similar to that observed in *A. vinelandii DJ*. Whether this type of glutaredoxin is capable of binding cofactors other than [2Fe-2S] is still a matter of debate; however the observation that class II Grxs in a yeast *grx5* deletion strain are capable of rescuing the activity of the [4Fe-4S] enzyme aconitase [49] supports this idea.

*1.4) Molybdenum cofactor biosynthesis protein A homologs*

The gene directly downstream of the *fdh* operon, *moaA (*Avin03850), codes for a S-adenosylmethionine radical (SAM) enzyme associated to two [4Fe-4S] clusters and involved in the first step of the biosynthesis of the molybdenum cofactor. As previously described [29], Avin30330 is another MoaA homolog found in *A. vinelandii DJ*. Here, by analyzing proteins with motifs predicted for radical SAM interaction, we have found three new possible [4Fe-4S] radical SAM enzymes dedicated to the molybdenum metabolism (Table 8).

The most prominent candidate is Avin35520, a protein that shares similarity to both MoaA homologs (Avin03850 and Avin30330) as well as containing the same

predicted domains – radical SAM interaction and molybdenum cofactor synthesis C. Avin35520 seems to be shorter in its N-terminal region than Avin03850 and Avin30330, however it has four conserved cysteines that could coordinate an [Fe-S] cluster (Figure 2). The other two identified proteins are transcribed from genes near each other but in opposite strands (Avin20630 and Avin20660; see Table 8). These candidates show a lower degree of similarity to the MoaA homologs (Avin03850 and Avin30330) and contain only the radical SAM domain. Nonetheless, they are both described as proteins from the MoaA/ NifB/ PqqE family and, since MoaA, NifB and PqqE are all [Fe-S] proteins, Avin20630 and Avin20660 are predicted to bind clusters as well. Interestingly, both proteins, Avin20630 and Avin20660, seem to be part of operons where a glycosil transferase 2-like enzyme is also encoded. This family of proteins is known to catalyze steps in the biosynthesis of sugars and many members are implicated on formation of glycosidic bonds. Taking this fact in consideration, it is possible that Avin20630 and Avin20660 are not related to the molybdenum metabolism, but could be [Fe-S] proteins designated to other functions.

*1.5) Xanthine dehydrogenase-like complexes*

Xanthine dehydrogenase (XDH) and xanthine oxidase (XO) comprise the two forms of the complex metalloflavoprotein central to purine degradation xanthine oxidoreductase. The enzyme catalyzes the last steps in urate formation, hydroxylating hypoxanthine to xanthine and xanthine to uric acid [56, 57]. Among bacteria, the best studied XDH is that from *Rhodobacter capsulatus*, which was shown to be a cytoplasmic heterodimeric enzyme encoded by the *xdhA* and *xdhB* genes [58]. Detailed examination on XDH structure have shown that the XdhA binds two [2Fe-2S] clusters and a FAD cofactor, while XdhB binds the molybdopterin

(MTP) form of molybdenum cofactor (Moco) [59, 58]. Also, a third gene co-transcribed with *xdhAB* was found and named *xdhC*, after observations that XdhC is essential for XDH activity, since Moco is not inserted into the enzyme [60]. *R. capsulatus* XdhC binds Moco/MTP and transfers the cofactor to Moco-free apo-XDH [61].

In *A. vinelandii DJ* the *xdhCBA* operon is conserved (*Avin22270 to Avin22290*) and protein domain analysis, associated to operon predictions, allowed us to identify three new putative transcription units where xanthine dehydrogenase-like elements seem to be encoded (Table 9). Similar to all three predicted operons is a resemblance to isoquinoline 1-oxidoreductase (IOR) genes from *Brevundimonas diminuta* (later *Pseudomonas diminuta 7*) [62]. IOR catalyzes the hydroxylation of isoquinoline, a heterocyclic aromatic compound, to 1-oxo-1,2-dihydroisoquinoline with subsequent reduction of a suitable electron acceptor utilizing $H_2O$ as the source of the oxygen atom incorporated into substrate [62].

Each operon identified has a gene coding for a putative [2Fe-2S] protein with similarity to the small subunit of *B. diminuta* isoquinoline IOR, IorA – Avin25850, 55% identity to IorA; Avin26060 and Avin35540, both 53% identical to IorA. Also, genes coding for *B. diminuta* IorB-like proteins, the IOR large subunit, are found within these operons – Avin26070 and Avin35550, both 31% identical to IorB; Avin25840, 30% identical to IorB). Interestingly, two of these predicted operons encode conserved hypothetical proteins that could help assemble Moco/MTP into the putative enzymes formed by *iorAB*-like genes [63]. In fact, Avin25830 is a conserved XdhC-domain protein encoded within the *Avin25820 to Avin25850* operon and is 24% identical to *A. vinelandii DJ* XdhC protein. Figure 3 shows the domain organization of *A. vinelandii DJ* XdhA and IorA-like proteins, compared to *B.*

*diminuta* IorA, as well as *A. vinelandii DJ* XdhB and IorB-like proteins, compared to *B. diminuta* IorB.

*1.6) Rhodanese-like proteins*

Rhodanese-domain proteins comprise a large family of molecules dedicated to the transfer of sulfur from thiosulfate or 3-mercaptopyruvate to cyanide, with the formation of thiocyanate [64]. Organisms usually encode several rhodanese paralogues as rhodanese tandem repeats, single-domain proteins or associated to other functional domains [65]. The transfer reaction is commonly catalyzed by an active cysteine residue, although the active site loop in rhodanese proteins shows wide variability [65].

Various research groups have reported roles for rhodanese-like proteins in oxidative stress prevention and maintenance of cellular homeostasis [66, 67, 68, 69]. Besides that, it was observed that disruption of the *A. vinelandii DJ* rhodanese gene *rhdA* (*Avin07450*) led to impaired [Fe-S] proteins activation [70, 71], suggesting a role for RhdA on [Fe-S] clusters protection against oxidative damage. Together, these facts have led use to better investigate the *A. vinelandii DJ* rhodanese homologues.

Through Pfam analysis we were able to identity 11 rhodanese-domain proteins in *A. vinelandii DJ* and predict their respective genomic neighborhoods (Table 10). Five members of this group are single-domain rhodaneses (Avin04680, Avin20970, Avin46860, Avin46930 and Avin51050), four are tandem repeats of the rhodanese domain (Avin07450, Avin21500, Avin30960 and Avin32980) and two are rhodaneses associated to specific domains (Avin02700 and Avin48690). With the exception of Avin04680, all other *A. vinelandii DJ* rhodanese-domain proteins

exhibit at least one cysteine residue that could be part of an active site structure (Figure 4).

A more detailed investigation about each rhodanese-like protein was then assessed by STRING and operon analysis. In this way, MoeB2 (Avin02700) has a rhodanese domain associated to the molybdopterin biosynthesis protein domains ThiF and MoeZ_MoeB. In fact, the upstream gene to *moeb2*, *Avin02710*, encodes a structurally similar element containing the ThiF and MoeZ_MoeB domains, but lacking the rhodanese domain. The best STRING homologue to MoeB2 is the *Burkholderia thailandensis E264* protein MoeZ (BTH_II0296), sharing 65.43% identity to MoeB2 (Figure 5). STRING strongly supports the role of MoeB2-like proteins in molybdenum metabolism, as it can be observed on the interactions between MoeZ and BTH_I2202, BTH_I1706, BTH_I1704 and MoaC (see Figure 5). Also from relevant interest is the connection between MoeZ and the cysteine desulfurase IscS. In fact this prediction is very prominent, since there is practical evidence suggesting it. By studying the sulfur source for molybdenum cofactor biosynthesis in eukaryotes, Marelja and colleagues [72] showed that the human cysteine desulfurase Nfs1 interacts and transfers sulfur to the C-terminal rhodanese domain of MOCS3. MOCS3 has essentially the same domain organization that is seen in *A. vinelandii DJ* MoeB2, and is able to sulfurate the C terminus of MOCS2A, the smaller subunit of the molybdopterin synthase [73]. Another study demonstrated that Nfs1 and ubiquitin-activating enzyme-like protein Uba4, associated with other proteins, are vital to tRNA(Lys2)(UUU) and tRNA(Glu3)(UUC) thio-modification at the second position [74]. Uba4 has also the same domain structure observed in MOCS3 and MoeB2. Taken together, these results suggest that MoeB2 is part of the molybdenum cofactor biosynthesis machinery, besides it could interact with cysteine

desulfurases to transfer sulfur to yet uncharacterized proteins.

Avin04680 is the only A. vinelandii DJ rhodanese-like protein where no putative active site cysteine residues were found (see Figure 4). Even though, it appears to be encoded in an operon with Avin04690 – a class I glutaredoxin containing a CPYC motif – and SecB (Avin04700), a protein-export chaperone. STRING analysis revealed a conservation of this operon structure among gammaproteobacteria and coexpression of SecB and the rhodanese protein in E coli K12. Although the interaction between SecB and the other products of this operon is still dubious [55], the occurrence of a possible sulfurtransferase and a putative [Fe-S] cluster-containing glutaredoxin on the same transcription unit deserves further investigation.

The next rhodanese protein evaluated was RhdA (Avin07450), the best characterized member of this family in *A. vinelandii DJ*. RhdA is a two-domain rhodanese repeat protein with an uncommon active site motif (HCQTHHR) [71]. Its 3D structure is known [75] and there is compelling biochemical evidence to suggest a role for RhdA in protecting cells against oxidative damage, specially preventing inactivation of enzymes containing labile [Fe-S] cofactors [71].

Avin20970 is a single-domain rhodanese protein that based on orientation and proximity appears to be in an operon with a serine/ threonine phosphatase (Avin20950), an hypothetical protein (Avin20960) and a serine *O*-acetyltransferase (CysE, Avin20980), the enzyme that carries out the first and limiting step in the pathway of cysteine biosynthesis. The closest homologue to Avin20970 and best STRING hit is *Azoarcus sp. BH72* protein azo0365, which analysis have also indicated association with a CysE-like protein. Other organisms harbor the same interaction, suggesting that the sulfurtransferase activity of this rhodanese could be

associated to the conversion of L-serine into *O*-acetyl-L-serine by CysE.

At least two other rhodaneses in *A. vinelandii DJ* seem to be directly involved to cysteine metabolism: Avin21500 and Avin30960. The first one, Avin21500, is a tandem repeat rhodanese-domain protein predicted to be in an operon with MetC (Avin21510), a cystathionine β-lyase. MetC is part of the transsulfuration pathway to the biosynthesis of methionine which utilizes cysteine as the sulfur donor [76]. In *E. coli*, MetC catalyzes the third step of the route and does not belong to an operon. However, some organisms, including *A. vinelandii DJ* and some *Pseudomonas* species, contain a *metC* gene in an operon with a downstream gene coding for a rhodanese-like protein. In fact, STRING and Microbes Online database indicated gene fusion events between these two elements in *Pseudomonas syringae DC300*, *P. putida GB-1* and *P. fluorescens SBW25* (*PSPTO_2632, PputGB1_3161* and *PFLU2239* genes, respectively). The operon organization and the fusion events of these two genes, added to the putative sulfurtransferase activity of the rhodanese protein (Avin21500), might indicate an interaction between this protein and MetC to scavenge sulfur from cysteine to further use in methionine biosynthesis.

The second rhodanese-like protein that might be involved in cysteine metabolism in *A. vinelandii DJ* is the four-domain repeat rhodanese Avin30960. Two genes coding for hypothetical proteins (*Avin30940* and *Avin30950*) and one coding for a cysteine dioxygenase type I (*Avin30970*) are predicted to be in an operon with *Avin30960*. STRING supports the association between the rhodanese Avin30960 and the cysteine dioxygenase type I, as the same organization is observed in other organisms – Bradyrhizobium sp. BTAi1, Klebsiella pneumoniae subsp. pneumoniae MGH 78578, Pseudomonas aeruginosa 2192, P. fluorescens Pf-5, P. putida GB-1, P. syringae pv. Phaseolicola 1448A and Ralstonia pickettii 12D. Until recently, $Fe^{2+}$-

dependent cysteine dioxygenases were believed to be enzymes strictly related to eukaryotes, catalyzing the the degradation of cysteine to sulfinic acid [77]. However, phylogenetic and biochemical data have demonstrated the ability of some eubacteria to perform this cysteine catabolic pathway [78]. Interestingly, at least two works have identified cysteine desulfurase (NifS-like) proteins capable of using sulfinic acid as a substrate [79, 80] for sulfur generation. Nevertheless, as Dominy and colleagues argue [78], the product of NifS-like proteins reactions performing with sulfinic acid would not generate reduced sulfur, which is required for sulfur incorporation into [Fe-S] proteins. On this matter, the observed conservation of the rhodanese/ cysteine dioxygenase operon could suggest an yet uncharacterized function to the sulfurtransferase – rhodanese – component on the chemistry of cysteine degradation to sulfinic acid.

Three additional rhodanese-like proteins identified in *A. vinelandii DJ* share consistent homology to characterized sulfurtransferases. Avin32980 is a two-domain rhodanese similar to *E. coli* 3-mercaptopyruvate sulfurtransferase (MST) [81]. The first determined MST structure was that from *Leishmania major* [82] and different from strict rhodaneses, MSTs show specificity for 3-mercaptopyruvate as a sulfur donor, transfering it to cyanide [64]. Some years ago a research team partially purified *A. vinelandii* MST [83], however no activity assays are known. The two other *A. vinelandii DJ* proteins sharing homology to rhodaneses on which there is some experimental evidence are Avin46860 and Avin46930. Both are single-domain rhodaneses homologous to *E. coli* proteins GlpE (56% identity to Avin46860) and PspE (32% identity to Avin46930). A recent work [84] has characterized the *E. coli* proteins – where GlpE is cytosolic and PspE is periplasmic – and has observed that *glpE*/*pspE* mutants do not display apparent phenotypes. This finding possibly

indicates that neither gene is essential for sulfur metabolism [84]. From the neighborhood prediction in *A. vinelandii DJ* it is noteworthy to mention the probable occurrence of Avin46930 in an operon with a sulphate transporter (Avin46920) and a beta-lactamase-like protein (Avin46910). STRING reveals the occurrence of gene fusions between rhodaneses homologous to Avin46930 and proteins similar to Avin46910 in *Polaromonas sp.* (Bpro_4842) and *Nitrosococcus oceani* (Noc_2007).

One of *A. vinelandii DJ* rhodaneses, RhdK (Avin48690), has a singular architecture comprising a sulfurtransferase domain and an ankirin repeat domain. Gene neighborhood prediction localizes this element in an operon with genes coding for a mcbC-like oxidorreductase (Avin48680), a putative metallocluster binding protein (Avin48700), a conserved hypothetical protein (Avin48710) and a glutathione S-transferase (Avin48720). Comparison to other organisms through Microbes Online database revealed that rhodaneses with ankirin repeats and mcbC-like oxidorreductases are likely to be encoded in the same transcription units, at least in some eubacteria (Figure 6). NCBI's Conserved Domains repository describes the mcbC-like oxidorreductases (cd02142) as a subset of FMN-dependet proteins capable of modifying polypeptides by cyclizing a thioester to form a ring. Also, this family comprises oxydase domains of non-ribosomal peptide synthetases, as it can be exemplified by the epothilone biosynthesis pathway protein EpoB [85]. In this way, it could be speculated that the ankirin-repeat rhodaneses and mcbC-like oxidoreductases may have some biochemical interaction relevant to the organisms where both proteins are encoded by genes in the same operon. The thioester reaction catalyzed by mcbC-like oxidoreductases and the putative sulfurtransferase activity of Avin48690 – it is predicted to have an active cysteine residue (see Figure 3) – also contribute to a possible interaction.

The last rhodanese-like protein identified in *A. vinelandii DJ* is the *nif*-encoded RhdN (Avin51050). The only close homologue observed is PST_1303 (67.27% identity) from the also nitrogen-fixing bacterium *Pseudomonas stutzeri A1501* and is found in a *nif* operon equivalent to that of *A. vinelandii DJ* (Avin50990 to Avin51060; see Table 9). Possibly, these rhodaneses could be involved in some sulfur-transfer reaction that would ultimately lead to nitrogenase maturation. Although this issue has to be better investigated, there is some evidence that support the theory. Pagani and colleagues [86] have demonstrated that the activity of the Fe protein component of *Klebsiella pneumoniae* nitrogenase can be restored upon addition of bovine liver rhodanese in the presence of thiosulfate ferric citrate and reduced lipoate. Even though a mammalian sulfurtransferase was used, the authors in fact discussed the possibility of a *nif*-specific rhodanese leading to the activation of Fe protein [86].

*1.7) General remarks*

The results presented here summarize some of our findings concerning characterized and possible new [Fe-S] proteins in the nitrogen-fixing bacterium *A. vinelandii DJ*. Although only some relevant examples were discussed here, the complete set of proteins identified is much larger (Figure 7). This group comprises members from different functional classes and, possibly, capable of binding various [Fe-S] cofactors species, as it was recently demonstrated that the vanadium nitrogenase regulatory protein VnfA (Avin02780) binds a [4Fe-3S] cluster [87].

When the set of known and possible [Fe-S] proteins is analyzed in terms of members per functional category – assigned during *A. vinelandii DJ* genome's annotation – it is possible to observe that the majority of the identified proteins is

predicted to be involved in energy metabolism pathways (Figure 8). This result agrees with the well established roles of [Fe-S] clusters on oxido-reduction and electron transfer processes [1], even though a consistent number of candidates falls into at least two other prominent categories: biosynthesis of cofactors, prosthetic groups and carriers (group 4) and central intermediary metabolism (group 5; see Figure 8). Besides that, almost 20 predicted [Fe-S] proteins are clustered in the 'Unidentified' category (group 18.99, see Figure 8), stressing out the lack of knowledge on the biochemistry of many [Fe-S] proteins.

### Estimating IscR regulatory sequence motifs

On an attempt to better characterize the extension of IscR transcriptional regulation in *A. vinelandii DJ*, the occurrence of IscR-binding motifs upstream to coding regions was evaluated. The first step in this process was the generation of a position-specific score matrix (PSSM) based on the IscR-binding region upstream the *iscR* genes of *A. vinelandii DJ* and sequenced *Pseudomonas* species. This PSSM was compared to the IscR-binding matrix for *Beta* and *Gammaproteobacteria* from RegTransBase's curated database of transcription factor binding sites. The matrices showed a high degree of similarity, while some local differences were obvious (Figure 9). These slight discrepancies between profiles were accounted to the genomic background of the *Pseudomonadaceae* family – especially high GC content – introduced in our PSSM by using only *Azotobacter* and *Pseudomonas* sequences. However, this genomic background is desirable to a more rigorous and realistic prediction of DNA motifs.

Following PSSM generation, the DNA profile obtained was used to scan *Azotobacter*'s genome for putative IscR-binding regions. The MEME Suite Motif-

based sequence analysis tools were used to this purpose and a total of 36 binding sites, including the one upstream *iscR*, were identified. A selection of the most concordant hits (e.g. the best conserved motifs) scoring below a $1x10^{-5}$ threshold was then made and can be observed on Table 11A. The same 36 initial motifs were also submitted to MREC algorithm for cross-validation of data generated by the MEME package (Table 11B). While some high-scoring hits were predicted by only one of the two methods used, there was a group of sequences that both motif-prediction softwares were able to identify (depicted in gray in Tables 11A and 11B). Members predicted by MEME Suite and MREC were designated as 'best hits' and they were chosen for a closer investigation, attempting to establish a link to a possible IscR transcriptional regulation.

In *Escherichia coli*, IscR was observed to bind to DNA regions from about 40 to 80 nucleotides upstream the *iscR* gene, downregulating transcription of the *isc* operon [8]. Therefore, the first parameter evaluated on the set of best hits was the distance of predicted IscR-binding motifs relative to the predicted start codons. As Figure 10 shows, some IscR-regulated candidates have putative IscR-binding motifs similarly distanced from their translation start sites as the *A. vinelandii DJ iscR* gene (*Avin40410*) has. Composing this group are *rplL* (*Avin06160*), *nifB* (*Avin51010*), *gor* (*Avin25010*), *Avin49130*, and *Avin41790*. Only the last two genes code for uncharacterized proteins and, although little is known of global IscR regulation even for well-studied portions of bacterial genomes, the present work prioritize the study of hypothetical and not described candidates. Motifs more distantly located from start codons (see Figure 10) relate to five uncharacterized candidates – two genes coding for hypothetical proteins (*Avin16910* and *Avin26260*) and three genes that share homology to known molecules (*Avin12430*, *Avin17910*, and *Avin51780*).

Next, the genomic neighborhood of the 10 genes having the best scored IscR-binding motifs was evaluated (Table 12). The majority of candidates appear to be single encoded molecules, while only three are seen in operons. *nifB* (*Avin51010*) is part of the well described *nif* (*ni*trogen *f*ixation) system, a highly regulated machinery devoted to nitrogenase expression. *rplL* (*Avin06160*) corresponds to the 50S ribosomal protein L7/L12 and is transcribed with several other members of the translation apparatus. Finally, *Avin17910* is an uncharacterized gene predicted to encode a group 1 glycosyl transferase and to be co-transcribed with two conserved hypothetical partners (see Table 12). Although nitrogen fixation and translation are thorough investigated processes, an evident IscR regulation of these features was not yet experimentally observed. However, as Figures 11 and 12 show, an association between IscR and regulation of nitrogen fixing and ribosomal proteins cannot be over ruled, due to the great complexity of both biological pathways, considering the predictions reported here.

The putative glycosyl transferase gene *Avin17910* and its possible co-transcribed partners *Avin17920* and *Avin17930* are encountered in an operon structure common to various *Pseudomonas* species (Figure 13), suggesting a conserved role for these elements in the *Pseudomonadaceae* family. Usually, proteins with glycosyl transferase motifs like Avin17910 are involved in the biosynthesis of disaccharides, oligosaccharides and polysaccharides, catalyzing the formation of glycosidic bonds. Avin17910 has a predicted molecular weight of 40405.8 Da and a theoretical pI equivalent to 9.50 , while its instability index, as computed by Protparam, classifies the protein as unstable. Avin17920 harbors the conserved uncharacterized domain DUF2334, found in many hypothetical bacterial proteins. This polypeptide is estimated to have a pI of 9.73, a molecular weight of 29613.0 Da

and is also hypothesized as an unstable molecule. Finally, Avin17930 has the UPF0104 domain related to uncharacterized proteins and to lysylphosphatidylglycerol (L-PG) synthetase, the enzyme responsible for the addition of lysine to membrane's phosphatidylglycerol. L-PG biosynthesis is central to bacterial virulence mechanisms and evasion from the host immune system and is best understood in Gram-positive organisms like *Staphylococcus aureus* [88]. Avin17930 is also classified as an unstable molecule, with a theoretical pI of 11.69 and a molecular weight of 35714.0 Da.

Since protein 3-D structures are better conserved than sequences, BLAST searches against the PDB (Protein Data Bank) dataset were retrieved in an attempt to identify characterized homologues of the proteins encoded in the *Avin17920* to *Avin17930* operon. For Avin17910, the most similar structure found was *Bacillus anthracis'* ORF BA1558 (PDB: 2JJM_A). Although the function of BA1558 remains unknown, crystallization experiments have identified it as a group 4 glycosyl transferase [89]. Members of this group are involved a variety of processes, as lipopolysaccharide biosynthesis, phosphatidylinostol mannoside synthesis and mycothiol biosynthesis; however, BA1558 shares no more than 23% sequence identity to any protein of these pathways [89]. No sequences within the PDB database are similar to Avin17920 and Avin17930.

Considering the single-encoded genes predicted to have IscR-binding motifs on their 5'-untranslated regions, *gor* (*Avin25010*) is the only known one. This gene is 46% identical to its *E. coli* ortholog, a well characterized element that codes for glutathione reductase [90, 91]. Glutathione reductase is a flavoprotein that reduces oxidized glutathione to form reduced glutathione (GSH), the major nonprotein sulfhydryl in living organisms. GSH takes part in many different intra-cellular

processes, including maintenance of reduced thiol groups, protection from oxygen-induced cell damage, and generation of deoxyribonucleotide precursors for DNA synthesis. An IscR-regulation of this protein could be understood as an integration of [Fe-S] cluster metabolism to cell survival processes.

The three following candidates, *Avin16910*, *Avin26260* and *Avin49130*, are hypothetical genes to which there is no information available. *Avin26260* is thought to encode a basic protein of 15934.9 Da. Estimations of its stability index classify this polypeptide as stable in solution. Both Avin49130 and Avin16910 are computed as unstable proteins, the first having a molecular weight of 8730.5 Da and a theoretical pI equivalent to 6.02 ; the second, 19403.4 Da and a pI of 9.99.

*Avin12430* is a conserved gene in the *Pseudomonadaceae* family predicted to encode a TonB-dependent siderophore receptor . Proteins of this kind are clustered in a large family, exemplified by *E. coli* TonB. TonB protein interacts with outer membrane receptor proteins that carry out high-affinity binding and energy-dependent uptake of specific substrates into the periplasmic space [92]. TonB-dependent regulatory systems usually consist of six components: a specialized outer membrane TonB-dependent receptor, an energizing TonB-ExbBD protein complex, a cytoplasmic membrane-localized anti-sigma factor and an extra cytoplasmic function (ECF)-subfamily sigma factor [93]. The TonB complex acts as a molecular sensor from the environment and transmits signals into the cytoplasm, leading to transcriptional activation of target genes. The best PDB database homologue of Avin12430 is the alcaligin outer membrane receptor FauA from *Bordetella pertussis* (PDB: 3EFM_A). Alcaligin is a siderophore produced under iron-limiting conditions that is released to the extracellular compartment to chelate iron. Ferric-alcaligin complexes are then taken up by *B. pertussis* through the TonB-dependent family

member FauA [94]. Considering these data, Avin12430 could also be involved in iron uptake and IscR could regulate its transcription according to cellular iron levels, since the [Fe-S] cluster in IscR acts as an iron sensor [8].

The last two genes predicted to be regulated by IscR are thought to act on carbohydrate metabolism. *Avin41790* is annotated as a gene coding for a carbohydrate transport membrane protein, while *Avin51780* codes for a putative alpha-glucosidase. Pfam computes a MSF_1 (Major Facilitator Superfamily, type 1) amino acid domain for Avin41790. MFS proteins act on the transport of many substrates across cytoplasmic or internal membranes, including ions, sugar phosphates, drugs, neurotransmitters, nucleosides, amino acids, and peptides. Bacterial members are primarily dedicated to the uptake of nutrients and to the efflux of drugs. NCBI's Conserved Domains Database also classifies Avin41790 as a AraJ-like arabinose efflux permease (COG2814). *E. coli* AraJ shares sequence similarity with drug efflux proteins of the MFS family, suggesting that it may function as a sugar efflux system [95]. On the other hand, Avin51780 has a motif corresponding to the 97th family of glycosidases, called Glyco_hydro_97 (Pfam). Glycosidases catalyze the cleavage of glycosidic bonds, generating either anomeric inverted or retained products [96]. Proteins within this family have a typical and complete (beta/alpha)8-barrel or catalytic TIM-barrel type domain, which can be involved in oligomerisation and carbohydrate binding [97]. BLAST searches on the PDB database have found a family 97 alpha-glycosidase from *Bacteroides thetaiotaomicron* (*Bt*GH97a, PDB: 2JKA_A) 40% identical to Avin51780. *Bt*GH97a was showed to act by an inversion mechanism dependent on calcium upon analysis by $^{1}$H nuclear magnetic resonance [96]. Therefore, the observation of IscR-binding sites upstream both *Avin41790* and *Avin51780* indicates a putative function for [Fe-S]

cluster-mediated regulation on carbohydrate metabolism.

# Conclusions

Despite the theoretical tone of this work and the need for future laboratory validation, the ideas proposed here are supported by a strong bioinformatics background. Taken together, the results presented here aggregate data on experimentally characterized and *in silico* predicted *A. vinelandii DJ* [Fe-S] proteins. In this way, the compiled database, added to the proposed IscR-regulated genes, can be used by researchers to further investigate new metabolic features driven by [Fe-S] cofactors and expand the current understanding on [Fe-S] cluster biosynthesis and transfer to terminal proteins.

# Materials and methods

### [Fe-S] protein sequences analysis

*A. vinelandii DJ* genomic information was primarily retrieved from the organism's genome project home page [12]. Complementary data were also found at the National Center for Biotechnology Information [13].

STRING 8 (Search Tool for the Retrieval of Interacting Genes/Proteins [14]) was used for the discovery of protein-protein interactions, both physical and functional ones [15]. The current software release is supported by data from 630 organisms accounting for about 2.5 million proteins deposited. The prediction method utilized relies on different information sources, that together greatly improve overall accuracy. STRING consults repositories of experimentally determined biological processes, as well as it seeks relevant data about the user's query – or proteins homologous to it – on scientific journals libraries, like Pubmed. Besides that, STRING's algorithm predicts interaction between proteins based on the genomic neighborhood of the protein of interest, gene fusions, co-expression patterns and evolutionary relationships. Utilizing all described criteria, a final score between 0 and 1 is attributed to each protein-protein interaction prediction to denote its probability of occurrence.

Protein domains were by Pfam [16], a database of protein domain families generated through seed alignments of amino acid sequences and profile hidden Markov models (HMMs) [17]. Query sequences are scanned against these profiles to predict domains based on internal thresholds. The current release, 24.0, contains 11912 Pfam-A families – high quality, curated sets of data – and additional lower quality Pfam-B families generated using the ADDA (Automatic Domain Decomposition Algorithm) database [18].

Estimations of adjacent genes within the same operon or transcription unit were assessed through MicrobesOnline Operon Predictions Suite [19, 20]. Based entirely on genomic sequences, the predictions retrieved by the software show good agreement with microarray data, since genes encoded in the same operon tend to have similar expression patterns. In order to decide if two adjacent genes belong to the same operon, MicrobesOnline considers (i) the intergenic distance between the genes, (ii) orthologs in the same disposition in other genomes, (iii) functional similarity by the Clusters of Orthologous Groups (COG) standard [21] and (iv) similarity of codon usage by codon adaptation index [22].

### Regulatory IscR sequence motifs estimation

The MEME Suite Motif-based sequence analysis tools [23], version 4.3.0, were used for generation of an IscR-binding domain position-specific scoring matrix (PSSM). Specifically, MEME (Multiple Em for Motif Elicitation) tool was applied to this task using the expectation maximization algorithm [24], a mathematical approach suited to find the best possible scores in an iterative process without testing all possibilities [25]. *A. vinelandii DJ*'s genome was scanned with the constructed PSSM by FIMO (Find Individual Motif Occurrences). FIMO searches for hits on a specified genome based on motifs provided by the user. The algorithm reports probabilities of positive occurrences, smaller than a specified threshold, as p-values, assuming a zero-order background model. P-values are then used to calculate q-values – a report of the minimal false discovery rate at which motifs are considered significant. Tomtom Motif Comparison Tool was used to compare DNA motifs to databases of known motifs. Scores are also reported as p- and q-values, based on the number of overlapping positions between the query motif and a motif from the

database [26].

MREC (Motif REcognition Computer program [27] identifies *cis*-regulatory motifs in sets of promoter sequences, scoring groups of candidate motifs with similar sequences using p-values [28]. This software was applied to cross-validate putative IscR-regulated candidates discovered by the MEME Suite. Scores are independent of the length of sequences, which was observed to greatly improve predictions. The software was tested on both *in silico* and biological data from prokaryotes and was proved to outperform Cosmo and MEME motif-finding tools.

## Authors' contributions

EPM performed the analyses and wrote the paper. JCS, DRD and JF designed the computational experiments and revised the manuscript. All authors contributed deeply to the final version of the text.

# References


1. Johnson, DC, Dean, DR, Smith, AD, Johnson, MK: **Structure, function, and formation of biological iron-sulfur clusters**. *Annu. Rev. Biochem.* 2005, **74**:247-81.

2. Balk, J, Lobréaux, S: **Biogenesis of iron-sulfur proteins in plants**. *Trends Plant Sci.* 2005, **10**:324-31.

3. Lill, R, Mühlenhoff, U: **Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases**. *Annu. Rev. Biochem.* 2008, **77**:669-700.

4. Malkin, R, Rabinowitz, JC: **The reconstitution of clostridial ferredoxin**. *Biochem. Biophys. Res. Commun.* 1966, **23**:822-7.

5. Arnold, W, Rump, A, Klipp, W, Priefer, UB, Pühler, A: **Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae**. *J. Mol. Biol.* 1988, **203**:715-38.

6. Jacobson, MR, Cash, VL, Weiss, MC, Laird, NF, Newton, WE, Dean, DR: **Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii**. *Mol. Gen. Genet.* 1989, **219**:49-57.

7. Jacobson, MR, Brigle, KE, Bennett, LT, Setterquist, RA, Wilson, MS, Cash, VL, Beynon, J, Newton, WE, Dean, DR: **Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii**. *J. Bacteriol.* 1989, **171**:1017-27.

8. Schwartz, CJ, Giel, JL, Patschkowski, T, Luther, C, Ruzicka, FJ, Beinert, H, Kiley, PJ: **IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins**. *Proc. Natl. Acad. Sci. U.S.A.* 2001, **98**:14895-900.

9. Rincon-Enriquez, G, Crété, P, Barras, F, Py, B: **Biogenesis of Fe/S proteins**

**and pathogenicity: IscR plays a key role in allowing Erwinia chrysanthemi to adapt to hostile conditions**. *Mol. Microbiol.* 2008, **67**:1257-73.

10. Wang, T, Shen, G, Balasubramanian, R, McIntosh, L, Bryant, DA, Golbeck, JH: **The sufR gene (sll0088 in Synechocystis sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria**. *J. Bacteriol.* 2004, **186**:956-67.

11. Giel, JL, Rodionov, D, Liu, M, Blattner, FR, Kiley, PJ: **IscR-dependent gene expression links iron-sulphur cluster assembly to the control of O2-regulated genes in Escherichia coli**. *Mol. Microbiol.* 2006, **60**:1058-75.

12. *Azotobacter vinelandii* Genome Project [http://www.azotobacter.org/]

13. National Center for Biotechnology Information [http://www.ncbi.nlm.nih.gov/]

14. STRING: functional protein association networks [http://string.embl.de/]

15. Jensen, LJ, Kuhn, M, Stark, M, Chaffron, S, Creevey, C, Muller, J, Doerks, T, Julien, P, Roth, A, Simonovic, M, Bork, P, von Mering, C: **STRING 8--a global view on proteins and their functional interactions in 630 organisms**. *Nucleic Acids Res.* 2009, **37**:D412-6.

16. Pfam [http://pfam.sanger.ac.uk/]

17. Finn, RD, Mistry, J, Tate, J, Coggill, P, Heger, A, Pollington, JE, Gavin, OL, Gunasekaran, P, Ceric, G, Forslund, K, Holm, L, Sonnhammer, ELL, Eddy, SR, Bateman, A: **The Pfam protein families database**. *Nucleic Acids Res.* 2010, **38**:D211-22.

18. Heger, A, Holm, L: **Exhaustive enumeration of protein domain families**. *J. Mol. Biol.* 2003, **328**:749-67.

19. MicrobesOnline Operon Predictions [http://www.microbesonline.org/operons/]

20. Price, MN, Huang, KH, Alm, EJ, Arkin, AP: **A novel method for accurate operon predictions in all sequenced prokaryotes**. *Nucleic Acids Res.* 2005, **33**:880-92.

21. Tatusov, RL, Natale, DA, Garkavtsev, IV, Tatusova, TA, Shankavaram, UT, Rao, BS, Kiryutin, B, Galperin, MY, Fedorova, ND, Koonin, EV: **The COG database: new developments in phylogenetic classification of proteins from complete genomes**. *Nucleic Acids Res.* 2001, **29**:22-8.

22. Sharp, PM, Li, WH: **Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons**. *Nucleic Acids Res.* 1986, **14**:7737-49.

23. The MEME Suite Motif-based sequence analysis tools [http://meme.sdsc.edu/meme4_3_0/intro.html]

24. Bailey, TL, Elkan, C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers**. *Proc Int Conf Intell Syst Mol Biol* 1994, **2**:28-36.

25. Mrázek, J: **Finding sequence motifs in prokaryotic genomes--a brief practical guide for a microbiologist**. *Brief. Bioinformatics* 2009, **10**:525-36.

26. Gupta, S, Stamatoyannopoulos, JA, Bailey, TL, Noble, WS: **Quantifying similarity between motifs**. *Genome Biol.* 2007, **8**:R24.

27. CSBL Motif finding: MREC [http://csbl.bmb.uga.edu/~bingqiang/MREC1/]

28. Li, G, Liu, B, Xu, Y: **Accurate recognition of cis-regulatory motifs with the correct lengths in prokaryotic genomes**. *Nucleic Acids Res.* 2010, **38**:e12.

29. Setubal, JC, dos Santos, P, Goldman, BS, Ertesvåg, H, Espin, G, Rubio, LM, Valla, S, Almeida, NF, Balasubramanian, D, Cromes, L, Curatti, L, Du, Z, Godsy, E, Goodner, B, Hellner-Burris, K, Hernandez, JA, Houmiel, K, Imperial, J, Kennedy, C, Larson, TJ et al.: **Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes**. *J. Bacteriol.* 2009,

**191**:4534-45.

30. Curatti, L, Brown, CS, Ludden, PW, Rubio, LM: **Genes required for rapid expression of nitrogenase activity in Azotobacter vinelandii**. *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**:6291-6.

31. Ragsdale, SW: **Life with carbon monoxide**. *Crit. Rev. Biochem. Mol. Biol.* 2004, **39**:165-95.

32. Holmgren, A: **Thioredoxin**. *Annu. Rev. Biochem.* 1985, **54**:237-71.

33. Collet, J, Peisach, D, Bardwell, JCA, Xu, Z: **The crystal structure of TrxA(CACA): Insights into the formation of a [2Fe-2S] iron-sulfur cluster in an Escherichia coli thioredoxin mutant**. *Protein Sci.* 2005, **14**:1863-9.

34. Lundström, J, Holmgren, A: **Protein disulfide-isomerase is a substrate for thioredoxin reductase and has thioredoxin-like activity**. *J. Biol. Chem.* 1990, **265**:9114-20.

35. Spector, A, Yan, GZ, Huang, RR, McDermott, MJ, Gascoyne, PR, Pigiet, V: **The effect of H2O2 upon thioredoxin-enriched lens epithelial cells**. *J. Biol. Chem.* 1988, **263**:4984-90.

36. Schallreuter, KU, Wood, JM: **The role of thioredoxin reductase in the reduction of free radicals at the surface of the epidermis**. *Biochem. Biophys. Res. Commun.* 1986, **136**:630-7.

37. Nakamura, H, Matsuda, M, Furuke, K, Kitaoka, Y, Iwata, S, Toda, K, Inamoto, T, Yamaoka, Y, Ozawa, K, Yodoi, J: **Adult T cell leukemia-derived factor/human thioredoxin protects endothelial F-2 cell injury caused by activated neutrophils or hydrogen peroxide**. *Immunol. Lett.* 1994, **42**:75-80.

38. Mark, DF, Chase, JW, Richardson, CC: **Genetic mapping of trxA, a gene affecting thioredoxin in Escherichia coli K12**. *Mol. Gen. Genet.* 1977, **155**:145-52.

39. Gleason, FK, Holmgren, A: **Thioredoxin and related proteins in procaryotes**. *FEMS Microbiol. Rev.* 1988, **4**:271-97.

40. Lim, CJ, Geraghty, D, Fuchs, JA: **Cloning and nucleotide sequence of the trxA gene of Escherichia coli K-12**. *J. Bacteriol.* 1985, **163**:311-6.

41. Masip, L, Pan, JL, Haldar, S, Penner-Hahn, JE, DeLisa, MP, Georgiou, G, Bardwell, JCA, Collet, J: **An engineered pathway for the formation of protein disulfide bonds**. *Science* 2004, **303**:1185-9.

42. Holmgren, A: **Hydrogen donor system for Escherichia coli ribonucleoside-diphosphate reductase dependent upon glutathione**. *Proc. Natl. Acad. Sci. U.S.A.* 1976, **73**:2275-9.

43. Rouhier, N, Couturier, J, Johnson, MK, Jacquot, J: **Glutaredoxins: roles in iron homeostasis**. *Trends Biochem. Sci.* 2010, **35**:43-52.

44. Rodríguez-Manzaneque, MT, Ros, J, Cabiscol, E, Sorribas, A, Herrero, E: **Grx5 glutaredoxin plays a central role in protection against protein oxidative damage in Saccharomyces cerevisiae**. *Mol. Cell. Biol.* 1999, **19**:8180-90.

45. Rodríguez-Manzaneque, MT, Tamarit, J, Bellí, G, Ros, J, Herrero, E: **Grx5 is a mitochondrial glutaredoxin required for the activity of iron/sulfur enzymes**. *Mol. Biol. Cell* 2002, **13**:1109-21.

46. Mühlenhoff, U, Gerber, J, Richhardt, N, Lill, R: **Components involved in assembly and dislocation of iron-sulfur clusters on the scaffold protein Isu1p**. *EMBO J.* 2003, **22**:4815-25.

47. Alves, R, Herrero, E, Sorribas, A: **Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism. II. Role of glutaredoxin Grx5**. *Proteins* 2004, **57**:481-92.

48. Vilella, F, Alves, R, Rodríguez-Manzaneque, MT, Bellí, G, Swaminathan, S,

Sunnerhagen, P, Herrero, E: **Evolution and cellular function of monothiol glutaredoxins: involvement in iron-sulphur cluster assembly**. *Comp. Funct. Genomics* 2004, **5**:328-41.

49. Molina-Navarro, MM, Casas, C, Piedrafita, L, Bellí, G, Herrero, E: **Prokaryotic and eukaryotic monothiol glutaredoxins are able to perform the functions of Grx5 in the biogenesis of Fe/S clusters in yeast mitochondria**. *FEBS Lett.* 2006, **580**:2273-80.

50. Rouhier, N, Unno, H, Bandyopadhyay, S, Masip, L, Kim, S, Hirasawa, M, Gualberto, JM, Lattard, V, Kusunoki, M, Knaff, DB, Georgiou, G, Hase, T, Johnson, MK, Jacquot, J: **Functional, structural, and spectroscopic characterization of a glutathione-ligated [2Fe-2S] cluster in poplar glutaredoxin C1**. *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**:7379-84.

51. Picciocchi, A, Saguez, C, Boussac, A, Cassier-Chauvat, C, Chauvat, F: **CGFS-type monothiol glutaredoxins from the cyanobacterium Synechocystis PCC6803 and other evolutionary distant model organisms possess a glutathione-ligated [2Fe-2S] cluster**. *Biochemistry* 2007, **46**:15018-26.

52. Comini, MA, Rettig, J, Dirdjaja, N, Hanschmann, E, Berndt, C, Krauth-Siegel, RL: **Monothiol glutaredoxin-1 is an essential iron-sulfur protein in the mitochondrion of African trypanosomes**. *J. Biol. Chem.* 2008, **283**:27785-98.

53. Lillig, CH, Berndt, C, Vergnolle, O, Lönn, ME, Hudemann, C, Bill, E, Holmgren, A: **Characterization of human glutaredoxin 2 as iron-sulfur protein: a possible role as redox sensor**. *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**:8168-73.

54. Mesecke, N, Mittler, S, Eckers, E, Herrmann, JM, Deponte, M: **Two novel monothiol glutaredoxins from Saccharomyces cerevisiae provide further insight into iron-sulfur cluster binding, oligomerization, and enzymatic activity of**

**glutaredoxins**. *Biochemistry* 2008, **47**:1452-63.

55. Prinz, WA, Aslund, F, Holmgren, A, Beckwith, J: **The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm**. *J. Biol. Chem.* 1997, **272**:15661-7.

56. Hille, R, Sprecher, H: **On the mechanism of action of xanthine oxidase. Evidence in support of an oxo transfer mechanism in the molybdenum-containing hydroxylases**. *J. Biol. Chem.* 1987, **262**:10914-7.

57. Mendel, RR, Bittner, F: **Cell biology of molybdenum**. *Biochim. Biophys. Acta* 2006, **1763**:621-35.

58. Leimkühler, S, Kern, M, Solomon, PS, McEwan, AG, Schwarz, G, Mendel, RR, Klipp, W: **Xanthine dehydrogenase from the phototrophic purple bacterium Rhodobacter capsulatus is more similar to its eukaryotic counterparts than to prokaryotic molybdenum enzymes**. *Mol. Microbiol.* 1998, **27**:853-69.

59. Truglio, JJ, Theis, K, Leimkühler, S, Rappa, R, Rajagopalan, KV, Kisker, C: **Crystal structures of the active and alloxanthine-inhibited forms of xanthine dehydrogenase from Rhodobacter capsulatus**. *Structure* 2002, **10**:115-25.

60. Leimkühler, S, Klipp, W: **Role of XDHC in Molybdenum cofactor insertion into xanthine dehydrogenase of Rhodobacter capsulatus**. *J. Bacteriol.* 1999, **181**:2745-51.

61. Neumann, M, Schulte, M, Jünemann, N, Stöcklein, W, Leimkühler, S: **Rhodobacter capsulatus XdhC is involved in molybdenum cofactor binding and insertion into xanthine dehydrogenase**. *J. Biol. Chem.* 2006, **281**:15701-8.

62. Lehmann, M, Tshisuaka, B, Fetzner, S, Lingens, F: **Molecular cloning of the isoquinoline 1-oxidoreductase genes from Pseudomonas diminuta 7, structural analysis of iorA and iorB, and sequence comparisons with other molybdenum-**

**containing hydroxylases**. *J. Biol. Chem.* 1995, **270**:14420-9.

63. Canne, C, Stephan, I, Finsterbusch, J, Lingens, F, Kappl, R, Fetzner, S, Hüttermann, J: **Comparative EPR and redox studies of three prokaryotic enzymes of the xanthine oxidase family: quinoline 2-oxidoreductase, quinaldine 4-oxidase, and isoquinoline 1-oxidoreductase**. *Biochemistry* 1997, **36**:9780-90.

64. Westley, J, Adler, H, Westley, L, Nishida, C: **The sulfurtransferases**. *Fundam Appl Toxicol* 1983, **3**:377-82.

65. Bordo, D, Bork, P: **The rhodanese/Cdc25 phosphatase superfamily. Sequence-structure-function relations**. *EMBO Rep.* 2002, **3**:741-6.

66. Nagahara, N, Katayama, A: **Post-translational regulation of mercaptopyruvate sulfurtransferase via a low redox potential cysteine-sulfenate in the maintenance of redox homeostasis**. *J. Biol. Chem.* 2005, **280**:34569-76.

67. Nagahara, N, Yoshii, T, Abe, Y, Matsumura, T: **Thioredoxin-dependent enzymatic activation of mercaptopyruvate sulfurtransferase. An intersubunit disulfide bond serves as a redox switch for activation**. *J. Biol. Chem.* 2007, **282**:1561-9.

68. Santos, PM, Benndorf, D, Sá-Correia, I: **Insights into Pseudomonas putida KT2440 response to phenol-induced stress by quantitative proteomics**. *Proteomics* 2004, **4**:2640-52.

69. Krivobok, S, Kuony, S, Meyer, C, Louwagie, M, Willison, JC, Jouanneau, Y: **Identification of pyrene-induced proteins in Mycobacterium sp. strain 6PY1: evidence for two ring-hydroxylating dioxygenases**. *J. Bacteriol.* 2003, **185**:3828-41.

70. Cereda, A, Carpen, A, Picariello, G, Iriti, M, Faoro, F, Ferranti, P, Pagani, S: **Effects of the deficiency of the rhodanese-like protein RhdA in Azotobacter**

**vinelandii**. *FEBS Lett.* 2007, **581**:1625-30.

71. Cereda, A, Carpen, A, Picariello, G, Tedeschi, G, Pagani, S: **The lack of rhodanese RhdA affects the sensitivity of Azotobacter vinelandii to oxidative events**. *Biochem. J.* 2009, **418**:135-43.

72. Marelja, Z, Stöcklein, W, Nimtz, M, Leimkühler, S: **A novel role for human Nfs1 in the cytoplasm: Nfs1 acts as a sulfur donor for MOCS3, a protein involved in molybdenum cofactor biosynthesis**. *J. Biol. Chem.* 2008, **283**:25178-85.

73. Matthies, A, Rajagopalan, KV, Mendel, RR, Leimkühler, S: **Evidence for the physiological role of a rhodanese-like protein for the biosynthesis of the molybdenum cofactor in humans**. *Proc. Natl. Acad. Sci. U.S.A.* 2004, **101**:5946-51.

74. Nakai, Y, Nakai, M, Hayashi, H: **Thio-modification of yeast cytosolic tRNA requires a ubiquitin-related system that resembles bacterial sulfur transfer systems**. *J. Biol. Chem.* 2008, **283**:27469-76.

75. Bordo, D, Deriu, D, Colnaghi, R, Carpen, A, Pagani, S, Bolognesi, M: **The crystal structure of a sulfurtransferase from Azotobacter vinelandii highlights the evolutionary relationship between the rhodanese and phosphatase enzyme families**. *J. Mol. Biol.* 2000, **298**:691-704.

76. Lee, H, Hwang, B: **Methionine biosynthesis and its regulation in Corynebacterium glutamicum: parallel pathways of transsulfuration and direct sulfhydrylation**. *Appl. Microbiol. Biotechnol.* 2003, **62**:459-67.

77. Sekowska, A, Kung, HF, Danchin, A: **Sulfur metabolism in Escherichia coli and related bacteria: facts and fiction**. *J. Mol. Microbiol. Biotechnol.* 2000, **2**:145-77.

78. Dominy, JEJ, Simmons, CR, Karplus, PA, Gehring, AM, Stipanuk, MH:

**Identification and characterization of bacterial cysteine dioxygenases: a new route of cysteine degradation for eubacteria**. *J. Bacteriol.* 2006, **188**:5561-9.

79. Jaschkowitz, K, Seidler, A: **Role of a NifS-like protein from the cyanobacterium Synechocystis PCC 6803 in the maturation of FeS proteins**. *Biochemistry* 2000, **39**:3416-23.

80. Mihara, H, Kurihara, T, Yoshimura, T, Soda, K, Esaki, N: **Cysteine sulfinate desulfinase, a NIFS-like protein of Escherichia coli with selenocysteine lyase and cysteine desulfurase activities. Gene cloning, purification, and characterization of a novel pyridoxal enzyme**. *J. Biol. Chem.* 1997, **272**:22417-24.

81. Spallarossa, A, Forlani, F, Carpen, A, Armirotti, A, Pagani, S, Bolognesi, M, Bordo, D: **The "rhodanese" fold and catalytic mechanism of 3-mercaptopyruvate sulfurtransferases: crystal structure of SseA from Escherichia coli**. *J. Mol. Biol.* 2004, **335**:583-93.

82. Alphey, MS, Williams, RAM, Mottram, JC, Coombs, GH, Hunter, WN: **The crystal structure of Leishmania major 3-mercaptopyruvate sulfurtransferase. A three-domain architecture with a serine protease-like triad at the active site**. *J. Biol. Chem.* 2003, **278**:48219-27.

83. Pagani, S, Franchi, E, Colnaghi, R, Kennedy, C: **Identification of sulfurtransferase enzymes in Azotobacter vinelandii**. *FEBS Lett.* 1991, **278**:151-4.

84. Cheng, H, Donahue, JL, Battle, SE, Ray, WK, Larson, TJ: **Biochemical and Genetic Characterization of PspE and GlpE, Two Single-domain Sulfurtransferases of Escherichia coli**. *Open Microbiol J* 2008, **2**:18-28.

85. Schneider, TL, Shen, B, Walsh, CT: **Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation**. *Biochemistry* 2003, **42**:9722-30.

86. Pagani, S, Eldridge, M, Eady, RR: **Nitrogenase of Klebsiella pneumoniae. Rhodanese-catalysed restoration of activity of the inactive 2Fe species of the Fe protein**. *Biochem. J.* 1987, **244**:485-8.

87. Nakajima, H, Takatani, N, Yoshimitsu, K, Itoh, M, Aono, S, Takahashi, Y, Watanabe, Y: **The role of the Fe-S cluster in the sensory domain of nitrogenase transcriptional activator VnfA from Azotobacter vinelandii**. *FEBS J.* 2010, **277**:817-32.

88. Fedtke, I, Götz, F, Peschel, A: **Bacterial evasion of innate host defenses--the Staphylococcus aureus lesson**. *Int. J. Med. Microbiol.* 2004, **294**:189-94.

89. Ruane, KM, Davies, GJ, Martinez-Fleites, C: **Crystal structure of a family GT4 glycosyltransferase from Bacillus anthracis ORF BA1558**. *Proteins* 2008, **73**:784-7.

90. Greer, S, Perham, RN: **Glutathione reductase from Escherichia coli: cloning and sequence analysis of the gene and relationship to other flavoprotein disulfide oxidoreductases**. *Biochemistry* 1986, **25**:2736-42.

91. Scrutton, NS, Berry, A, Perham, RN: **Purification and characterization of glutathione reductase encoded by a cloned and over-expressed gene in Escherichia coli**. *Biochem. J.* 1987, **245**:875-80.

92. Chimento, DP, Kadner, RJ, Wiener, MC: **The Escherichia coli outer membrane cobalamin transporter BtuB: structural analysis of calcium and substrate binding, and identification of orthologous transporters by sequence/structure conservation**. *J. Mol. Biol.* 2003, **332**:999-1014.

93. Koebnik, R: **TonB-dependent trans-envelope signalling: the exception or the rule?**. *Trends Microbiol.* 2005, **13**:343-7.

94. Brillet, K, Meksem, A, Lauber, E, Reimmann, C, Cobessi, D: **Use of an in-**

**house approach to study the three-dimensional structures of various outer membrane proteins: structure of the alcaligin outer membrane transporter FauA from Bordetella pertussis**. *Acta Crystallogr. D Biol. Crystallogr.* 2009, **65**:326-31.

95. Reeder, T, Schleif, R: **Mapping, sequence, and apparent lack of function of araJ, a gene of the Escherichia coli arabinose regulon**. *J. Bacteriol.* 1991, **173**:7765-71.

96. Gloster, TM, Turkenburg, JP, Potts, JR, Henrissat, B, Davies, GJ: **Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora**. *Chem. Biol.* 2008, **15**:1058-67.

97. Naumoff, DG: **GH97 is a new family of glycoside hydrolases, which is related to the alpha-galactosidase superfamily**. *BMC Genomics* 2005, **6**:112.

# Figures

**Figure 1: Alignment of the N-terminal portion of the three thioredoxin-like proteins from *Azotobacter vinelandii DJ.* – *Avin06670*, *Avin45140* and *Avin47420*.**

Sequences of the three thioredoxin-like proteins Avin06670, Avin45140 and Avin47420 are shown. The figure highlights the conserved underlined CGPC motif (CAPC for *Avin06670*) with its two cysteines depicted in gray and the four additional cysteines from *Avin45140* forming two CXXC motifs.

**Figure 2: Alignment of two MoaA homologous proteins (Avin03850 and Avin30330) with Avin35520.**

Avin35520 is a possible new member dedicated to the biosynthesis of molybdenum cofactor. Avin35520 has a shortened N-terminal portion and contains four conserved cysteine residues compared to the MoaA homologues.

**Figure 3: Domain organization of *Azotobacter vinelandii DJ xdh*-like and *ior*-like gene products.**

Architectures of *A. vinelandii DJ* XdhA and IorA-like proteins, compared to *B. diminuta* IorA; domain organization of *A. vinelandii DJ* XdhB and IorB-like proteins, compared to *B. diminuta* IorB. XdhA and IorA-like proteins share the two [2Fe-2S] clusters binding region, while IorB-like proteins seem to be tandem repeats of XdhB domain Ald_Xan_dh_C2 (Pfam).

**Figure 4: Schematic representation of domains comprehended in *Azotobacter***

***vinelandii DJ* rhodanese-like proteins.**

Domains drawn to scale using ExPASy Proteomics Server's tool My Domains according to Pfam coordinates. Red diamonds: predicted active site cysteine residues (Pfam).

**Figure 5: STRING analysis of MoeB2 (Avin02700) a rhodanese domain protein.**

MoeB2 is associated to the molybdopterin biosynthesis protein domains ThiF and MoeZ_MoeB and its closest STRING hit – showed here – is Burkholderia thailandensis E264 protein MoeZ (BTH_II0296), sharing 65.43% identity to MoeB2. The software infers strong association of the query sequence to molybdenum metabolism (MoeZ interactions with BTH_I2202, BTH_I1706, BTH_I1704 and MoaC), as well as a relevant connection with IscS (see text).

**Figure 6: MicrobesOnline database genome alignment with genes coding for rhodaneses with ankirin repeats.**

Some eubacteria appear to have a tendency to encode this particular kind of rhodanese (highlighted in light red) together with mcbC-like oxidorreductases (light blue).

**Figure 7: Panel on *Azotobacter vinelandii DJ* characterized and predicted [Fe-S] proteins.**

Proteins are color-coded according to the nature of [Fe-S] cofactors they are known or predicted to ligate. Proteins are also organized in the functional categories assigned by *A.*

*vinelandii DJ* genome project or clustered in larger groups for convenience.

**Figure 8: Distribution of *Azotobacter vinelandii DJ* known and predicted [Fe-S] proteins into functional categories.**

Proteins are divided into groups established during the sequencing of *A. vinelandii DJ* genome. Numbers near each bar stand for functional sub-categories, as follows: amino acid biosynthesis – glutamate family (1.3), pyruvate family (1.5), serine family (1.6); Purines, pyrimidines, nucleosides and nucleotides – nucleoside and nucleotide interconversions (2.2), pyrimidine ribonucleotide biosynthesis (2.4), other (2.99); fatty acid and phospholipid metabolism – biosynthesis (3.1) ; biosynthesis of cofactors, prosthetic groups and carriers – biotin (4.1) , heme and porphyrin (4.4), molybdopterin (4.7), pyridine nucleotides (4.10), others (4.99); central intermediary metabolism – nitrogen metabolism (5.2), sulfur metabolism (5.5), nitrogen fixation (5.6), aromatic compounds (5.7), other (5.9); energy metabolism – aerobic (7.1), amino acids and amines (7.2), electron transport (7.5), Entner-Doudoroff (7.6), sugars (7.13), TCA cycle (7.14), other (7.99); transport and binding proteins – amino acids, peptides and amines (8.1), cations (8.4); DNA metabolism, replication, recombination and repair (9.2); translation – ribosomal proteins: synthesis and modification (11.5), tRNA modification (11.6); regulatory functions – Crp, Fnr (12.10) , other (12.99); cellular processes – toxin production and resistance (14.6), other (14.99); hypothetical, conserved hypothetical (16.2); undefined (18.99).

**Figure 9: Generation of an IscR-binding position-specific score matrix (PSSM).**

The matrix used here was based on the IscR-binding region upstream the *iscR* genes

of *A. vinelandii DJ* and sequenced *Pseudomonas* species. The PSSM constructed in this study (Avin_+_Pseudomonads_IscR_signatures, lower logogram) was compared to the IscR-binding matrix for *Beta* and *Gammaproteobacteria* from RegTransBase's curated database of transcription factor binding sites (IscR_Beta_Gamma, upper logogram). Three statistical methods – Pearson correlation coefficient, Euclidean distance and Sandelin-Wasserman similarity function – were then used to determine the similarity between the two matrices. Results are reported as q-values (the minimal false discovery rates at which observed similarities would be deemed significant) and converted to p-values. Overlap, query offset and orientation are position-specific parameters that help calculating p-values.

**Figure 10: Position of predicted IscR-binding motifs relative to the first codon of candidate genes.**

The distance of all elements is compared to *iscR* (*Avin40410*), a gene that has a known IscR-binding motif about 40 nucleotides upstream its start codon. All representations are drawn to the same scale.

**Figure 11: STRING analysis of *Pseudomonas aeruginosa PA14* 50S ribosomal protein L7/L12 (RplL).**

*P. aeruginosa PA14* RplL is the closest *A. vinelandii DJ* homologue for this polypeptide within STRING's database. The *rplL* gene product is involved in the complex process of ribosome assembly and translation: an extensive and highly regulated protein network. Only the 20 most direct interaction partners are shown.

**Figure 12: STRING analysis of *Pseudomonas stutzeri* NifB protein.**

*P. stutzeri* NifB is the closest *A. vinelandii DJ* homologue for this polypeptide within STRING's database. The *nifB* gene product is involved in maturation of the MoFe nitrogenase, a multi-step process dependable on many proteins. Only the 10 most direct interaction partners are shown.

**Figure 13: Conservation of the *Avin17910* to *Avin17930* operon structure of *Azotobacter vinelandii DJ* among the *Pseudomonas* genus.**

This predicted operon encodes a putative glycosyl transferase (light green) and two conserved hypothetical genes.

**Figure 1**

```
Avin06670       MSEIPYIFDVTDASFDQLVLENS-------------------------------FHKPVL 29
Avin47420       MSEL--ITNVSDASFEQDVLQA---------------------------------EGPVL 25
Avin45140       MTDALLIPCPHCHSLNRIPAARLGEAPRCGRCKSPVLPGKPFAVDAAGFAAQLKGDLPLL 60
                *::   *      *:::                                      . *:*

Avin06670       VDFWAEWCAPCKALMPLLAKIAESYQGELLLAKVNSDIEQQTVMRMGIRSLPTVVLFKDG 89
Avin47420       VDYWAEWCGPCKMIAPVLNEIAKDYEGRLKICKLNIDENQGTPAKYGVRGIPTLMLFKGG 85
Avin45140       LDVWADWCGPCKAFAPVFDQAARELAGRVRLAKLDSEAEAPLAGQLGIRSIPNLILFRAG 120
                :* **:**.*** : *:: : *..  *.: :.*:: : :     : *:*.:*.::**: *
```

**Figure 2**

```
Avin03850       MTRIPMASALQDGFGRTIDYLRLSVTDRCDFRCVYCMAEDMRFLPRRQILGLEELYRLAR 60
Avin30330       MS----DTELIDPFGRRITYLRLSVTDRCDFRCTYCMSENMVFVPREQILSLEELYDVAD 56
Avin35520       ----------------------MPIGKRCLRPSCSWPSKYLKVVN--------------- 23
                                     :.: .**   .    :: : .:

Avin03850       LFVGRGVKKIRLTGGEPLVRRDILGLCERIAALPGLRELVMTTNGSQLVRLARPLAEAGV 120
Avin30330       AFIDLGVKRIRITGGEPLVRRGLLGLLERLGARTELEDLAMTTNGSQLPFMAGDLRAAGV 116
Avin35520       -----DLRTLRITGGEPLISPLFDGFMAELG-RLGFADVSLTTNGQLLSRKLPVLLAAGI 77
                     .:: :*:******:   : *:  .:.    : :: :****. *      *  **:

Avin03850       KRLNISLDSLDPARFRAITRSGELARVLDGIEAARGAGFERIKLNTVVMQGRNADEVLDL 180
Avin30330       RRLNISLDSLRPERFAAFTRRDRLDRVLAGIEAARAAGFARIKLNSVVQRGRNDDEVLDL 176
Avin35520       ERINVSLDTLDAGALRRIARGGDLAEVLEGIDQALAAG-VRVKINMVPLRGHNREQVLPL 136
                .*:*:***:* .  :  ::* . * .** **: * .**  *:*:* *  :*:* ::** *

Avin03850       VDYAVARGLDISFIEEMPLG--EVG-RSRGEAFCSSDEVRARIASRYALLDCAEQSGGPA 237
Avin30330       VEFAVARGLDISFIEEMPLG--GISSHRRSETLCSSDEVRRRIEERFELVRSSHVSGGPS 234
Avin35520       LEYCLARGMELRFIELMRMGHLALDPVSFQRQFVGMAEILAIVSERHGFEHVPSAPDTTA 196
                :::.:***::: *** * :*   :.     . : .  *:   : .*. :   .  .. .:

Avin03850       RYVRLPAHPASRIGFISPHSHNFCASCNRLRLTVEGRLLLCLGHEQSLDLRALLRRHPTD 297
Avin30330       RYWQVVGS-RTQVGFISPHSHNFCAACNRVRVTAEGKLVLCLGHEGALDLKSLMRAHPGD 293
Avin35520       QRFAIPGR--GYFGVIANESAPFCSQCSRARLSSTGWLHGCLSSSRNHYLGDVLELPRDG 254
                :   : .     .*.*: .*  **: *.* *::  * *  **. .    *  ::.    .

Avin03850       DRPLVDAIDRALLHKPQQHDFSAGGEVQVLRFMNASGG 335
Avin30330       SARLRQALVDALRLKPERHHFEADAQVQVVRFMSMTGG 331
Avin35520       ALARIRELSARALADKQAERFSGDAMS-----MKLIGG 287
                      :      . : . *....       *.  **
```

**Figure 3**

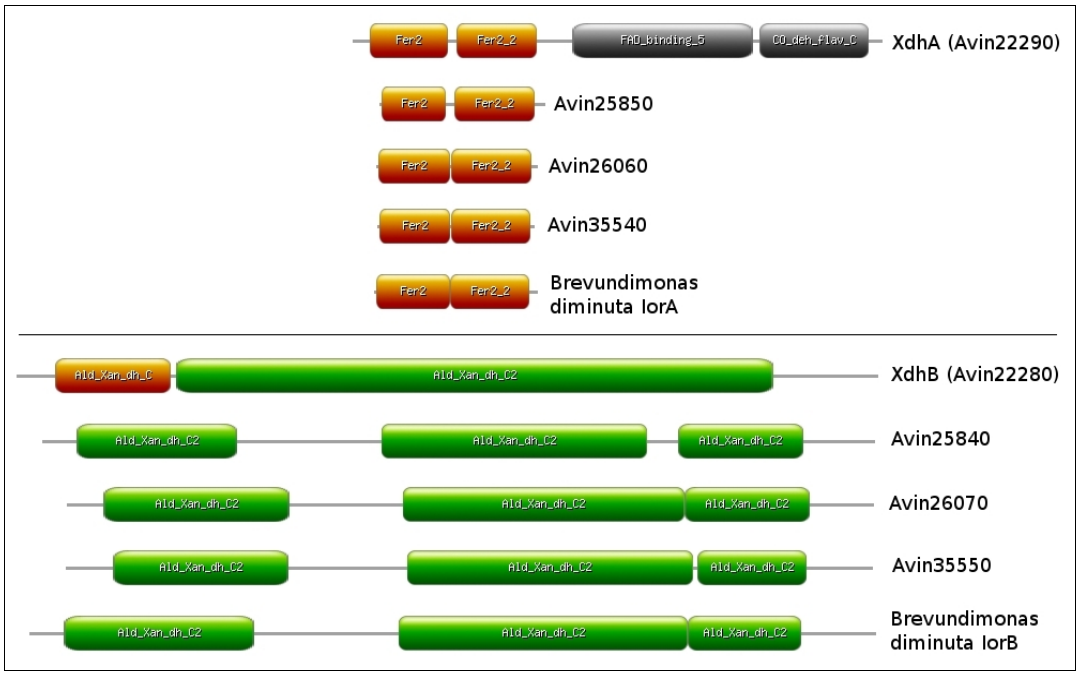
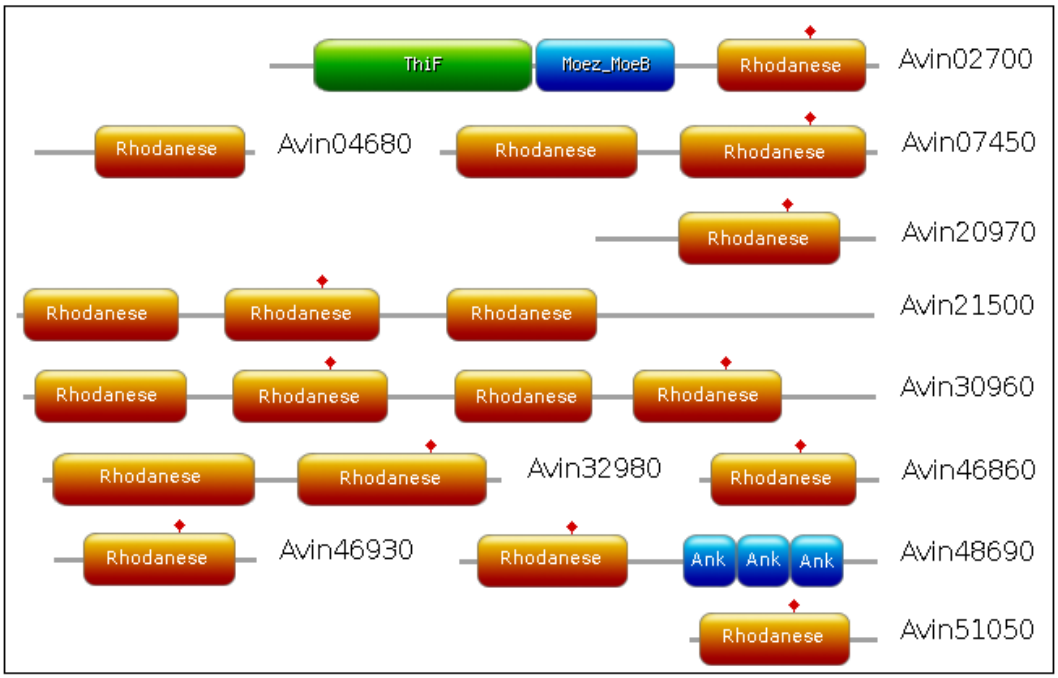
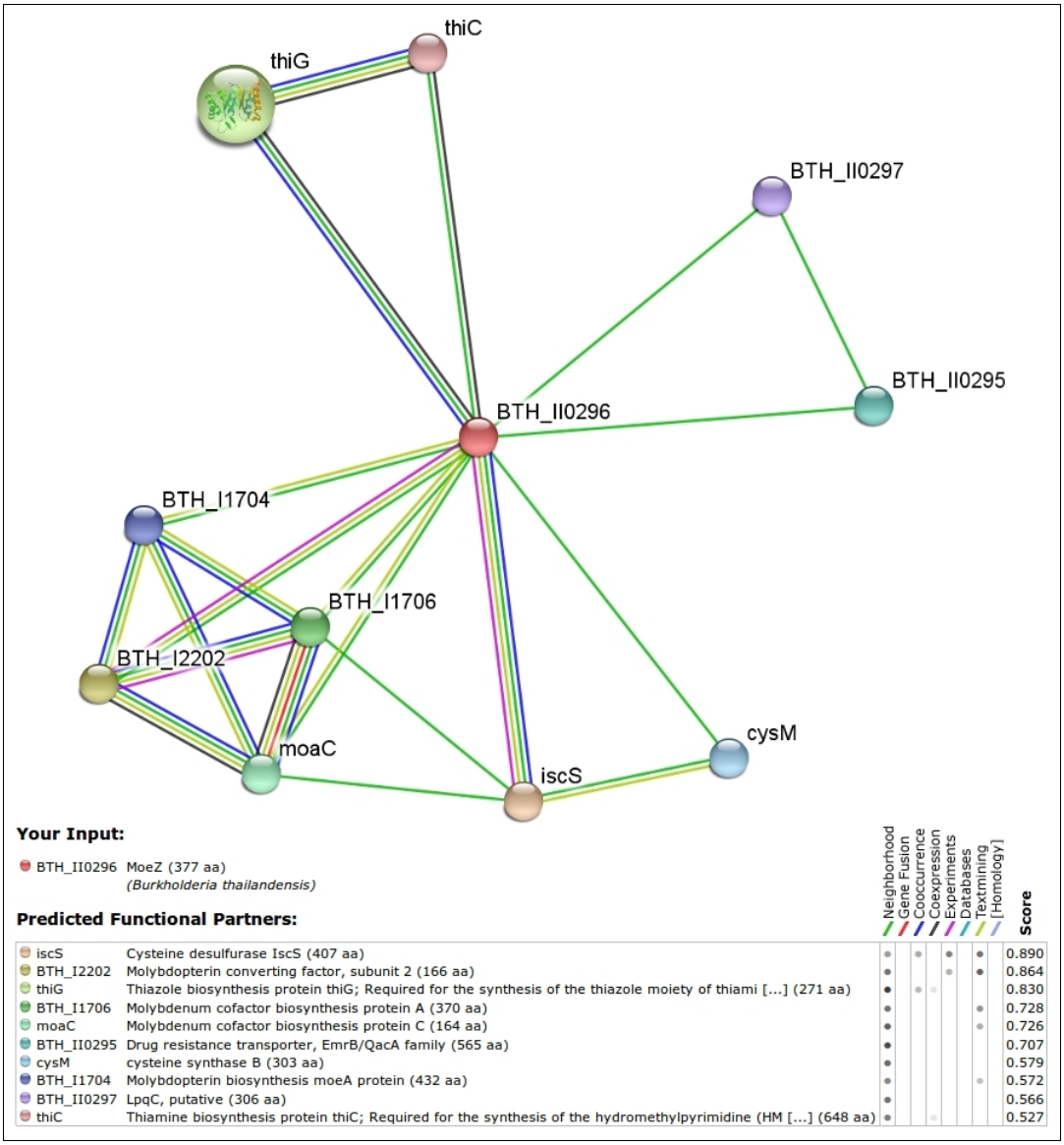



**Figure 4**



## Figure 5




## Figure 6

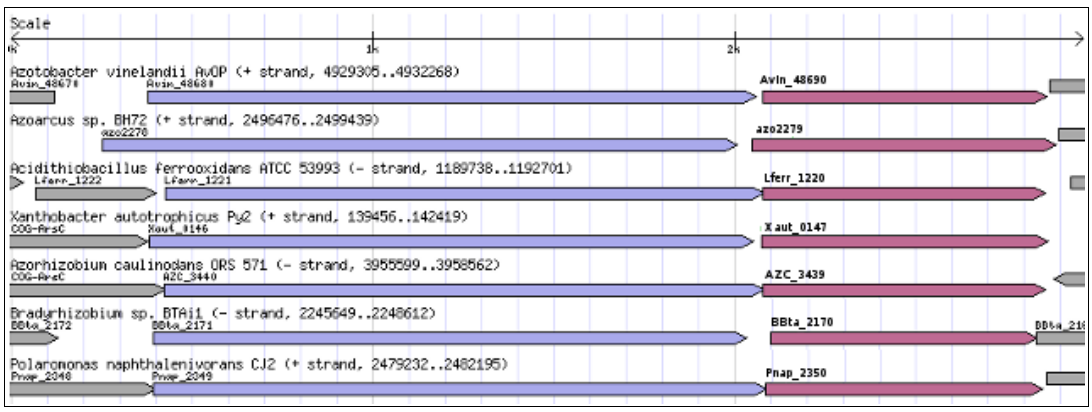


**Figure 7**

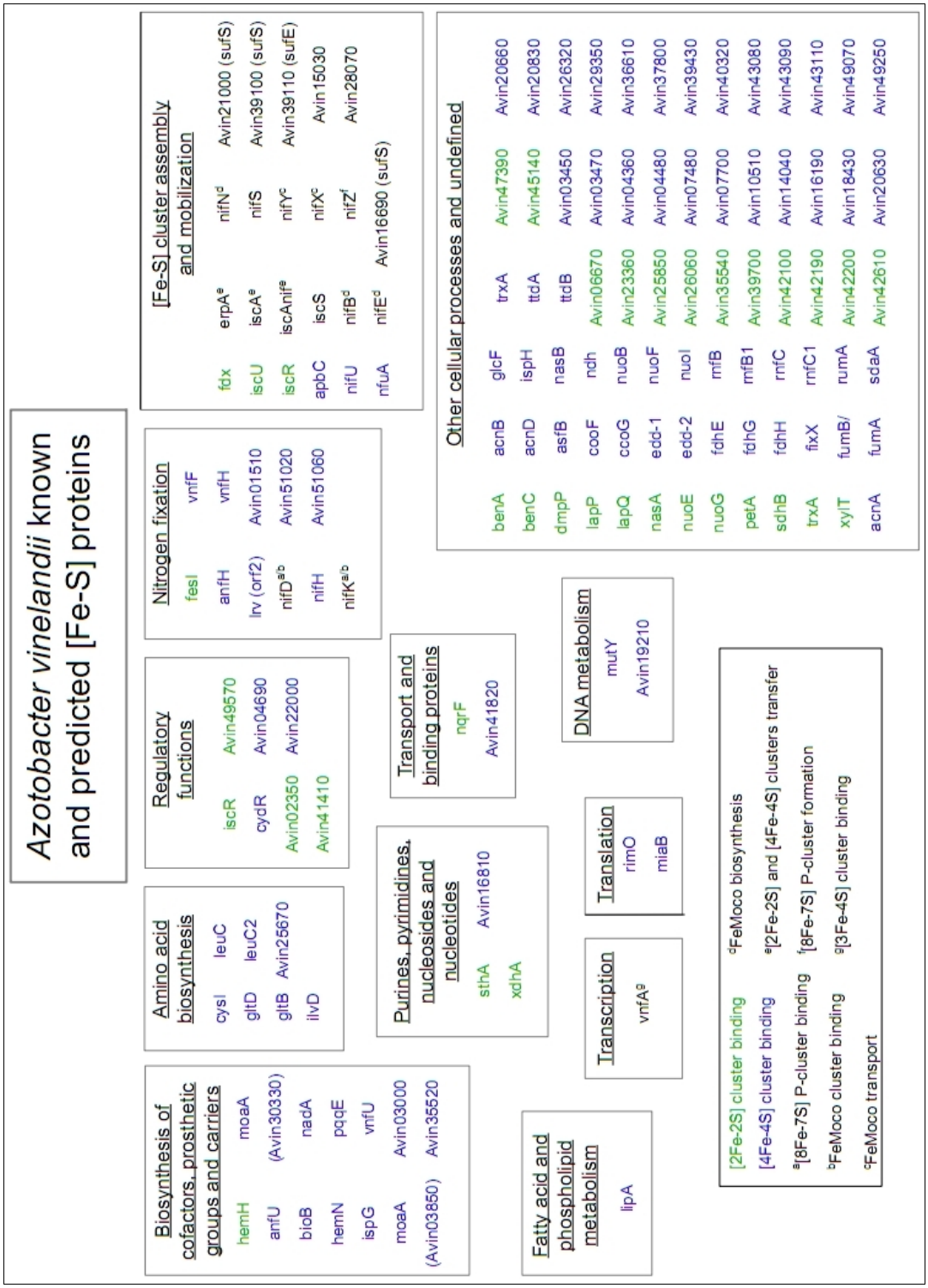

Azotobacter vinelandii known and predicted [Fe-S] proteins
[Fe-S] cluster assembly and mobilization
fdx erpA[a] nifN[d] Avin21000 (sufS)
iscU iscA[a] nifS Avin39100 (sufS)
iscR iscAnif[a] nifY[c] Avin39110 (sufE)
apbC iscS nifX[c] Avin15030
nifU nifB[d] nifZ[f] Avin28070
nfuA nifE[d] Avin16690 (sufS)
Nitrogen fixation
fesI vnfF
anfH vnfH
lrv (orf2) Avin01510
nifD[a,b] Avin51020
nifH Avin51060
nifK[a,b]
Other cellular processes and undefined
benA acnB glcF trxA Avin47390 Avin20660
benC acnD ispH ttdA Avin45140 Avin20830
dmpP asfB nasB ttdB Avin03450 Avin26320
lapP cooF ndh Avin06670 Avin03470 Avin29350
lapQ ccoG nuoB Avin23360 Avin04360 Avin36610
nasA edd-1 nuoF Avin25850 Avin04480 Avin37800
nuoE edd-2 nuoI Avin26060 Avin07480 Avin39430
nuoG fdhE mfB Avin35540 Avin07700 Avin40320
petA fdhG mfB1 Avin39700 Avin10510 Avin43080
sdhB fdhH rnfC Avin42100 Avin14040 Avin43090
trxA fixX rnfC1 Avin42190 Avin16190 Avin43110
xylT fumB/ rumA Avin42200 Avin18430 Avin49070
acnA fumA sdaA Avin42610 Avin20630 Avin49250
Regulatory functions
iscR Avin49570
cydR Avin04690
Avin02350 Avin22000
Avin41410
Amino acid biosynthesis
cysI leuC
gltD leuC2
gltB Avin25670
ilvD
Purines, pyrimidines, nucleosides and nucleotides
sthA Avin16810
xdhA
Transport and binding proteins
nqrF
Avin41820
DNA metabolism
mutY
Avin19210
Translation
rimO
miaB
Transcription
vnfA[g]
Biosynthesis of cofactors, prosthetic groups and carriers
hemH moaA
anfU (Avin30330)
bioB nadA
hemN pqqE
ispG vnfU
moaA Avin03000
(Avin03850) Avin35520
Fatty acid and phospholipid metabolism
lipA
[2Fe-2S] cluster binding
[4Fe-4S] cluster binding
[a][8Fe-7S] P-cluster binding
[b]FeMoco cluster binding
[c]FeMoco transport
[c]FeMoco biosynthesis
[e][2Fe-2S] and [4Fe-4S] clusters transfer
[f][8Fe-7S] P-cluster formation
[g][3Fe-4S] cluster binding

**Figure 8**

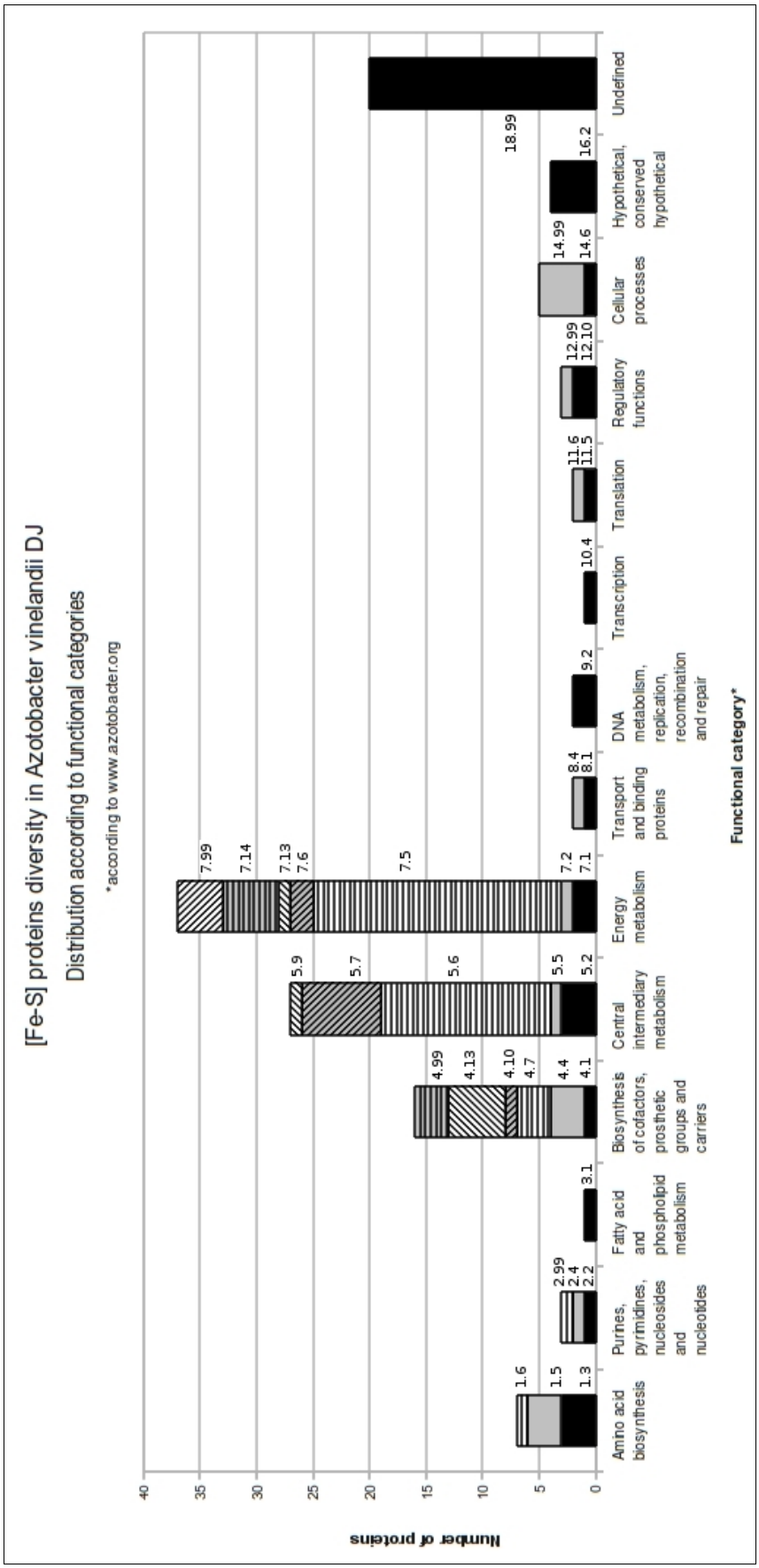

[Fe-S] proteins diversity in Azotobacter vinelandii DJ
Distribution according to functional categories
*according to www.azotobacter.org
Number of proteins
Functional category*
Amino acid biosynthesis
Purines, pyrimidines, nucleosides and nucleotides
Fatty acid and phospholipid metabolism
Biosynthesis of cofactors, prosthetic groups and carriers
Central intermediary metabolism
Energy metabolism
Transport and binding proteins
DNA metabolism, replication, recombination and repair
Transcription
Translation
Regulatory functions
Cellular processes
Hypothetical, conserved hypothetical
Undefined
1.6
1.5
1.3
2.99
2.4
2.2
3.1
4.99
4.13
4.10
4.7
4.4
4.1
5.9
5.7
5.6
5.5
5.2
7.99
7.14
7.13
7.6
7.5
7.2
7.1
8.4
8.1
9.2
10.4
11.6
11.5
12.99
12.10
14.99
14.6
18.99
16.2

**Figure 9**

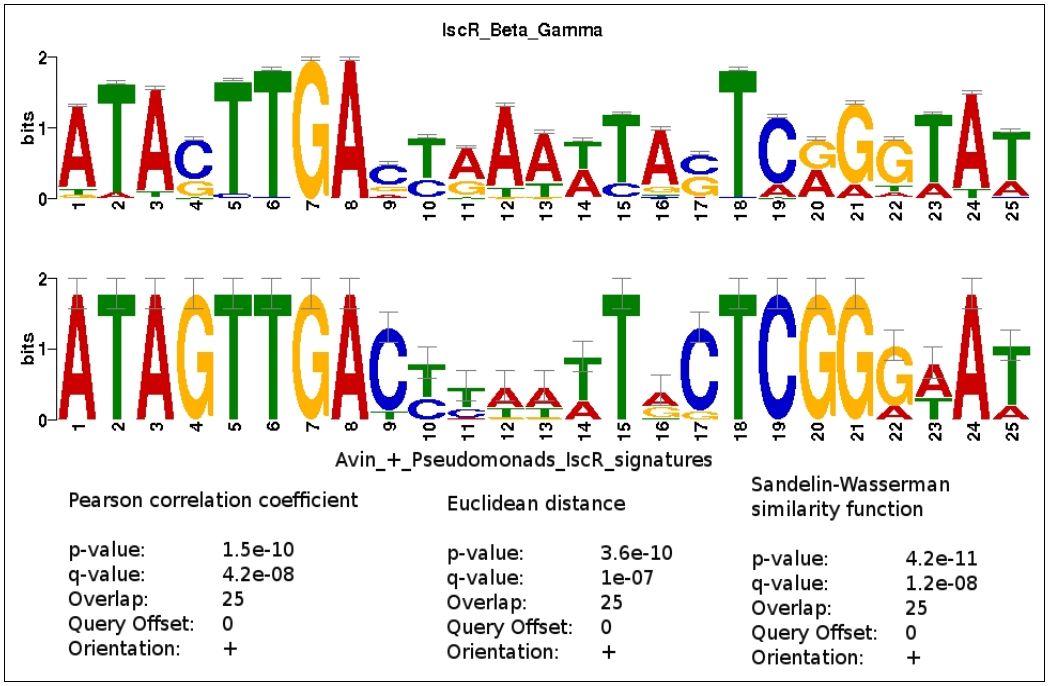

IscR_Beta_Gamma
bits
Avin_+_Pseudomonads_IscR_signatures
Pearson correlation coefficient
p-value: 1.5e-10
q-value: 4.2e-08
Overlap: 25
Query Offset: 0
Orientation: +
Euclidean distance
p-value: 3.6e-10
q-value: 1e-07
Overlap: 25
Query Offset: 0
Orientation: +
Sandelin-Wasserman similarity function
p-value: 4.2e-11
q-value: 1.2e-08
Overlap: 25
Query Offset: 0
Orientation: +

**Figure 10**

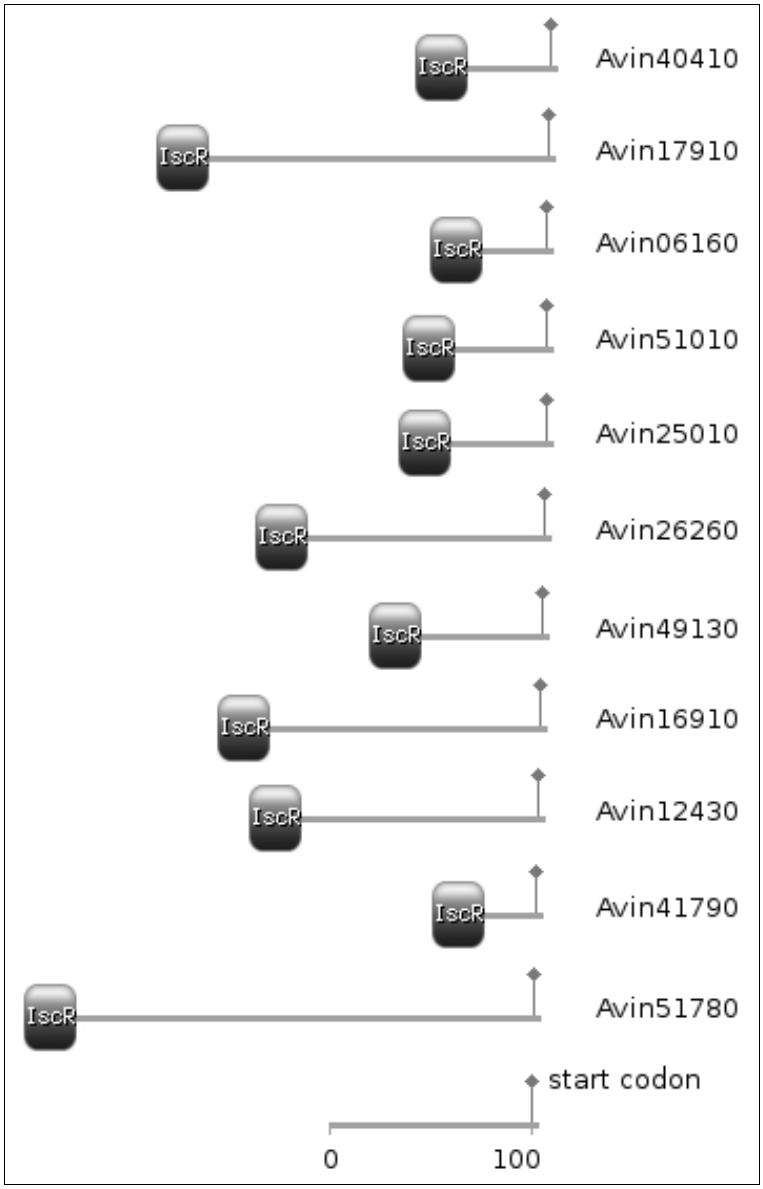
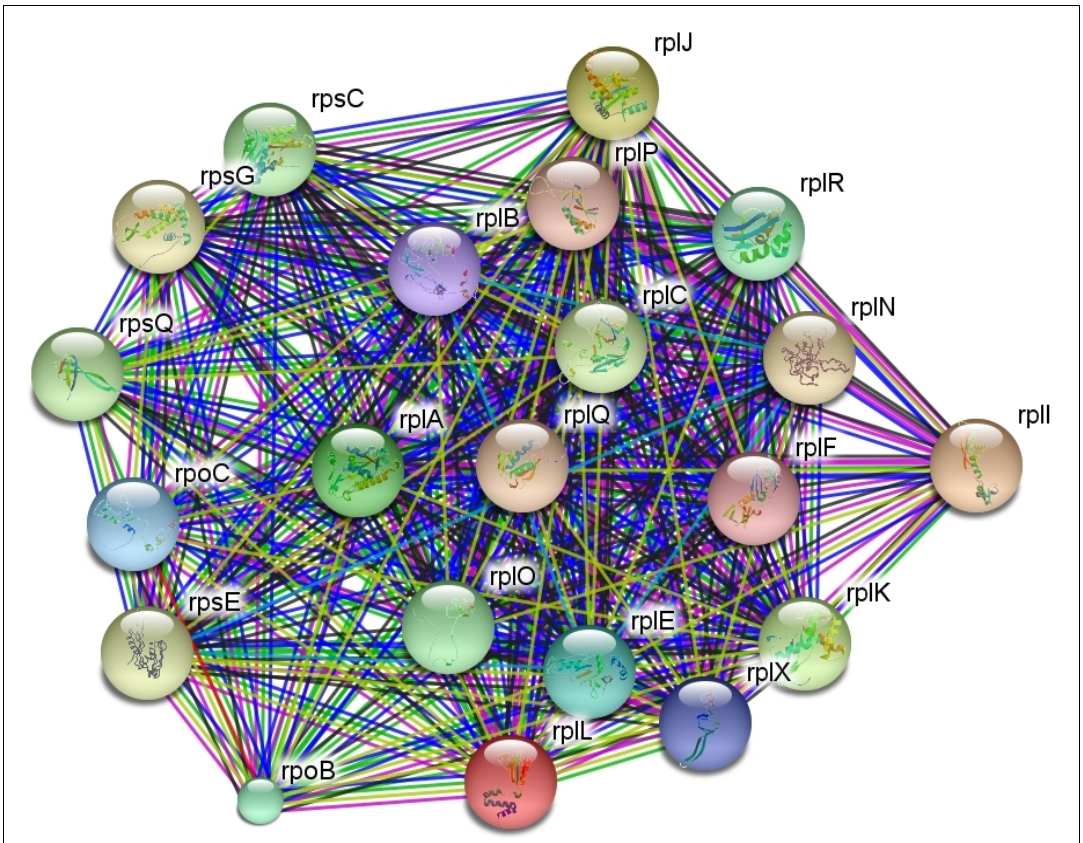

IscR
Avin40410
IscR
Avin17910
IscR
Avin06160
IscR
Avin51010
IscR
Avin25010
IscR
Avin26260
IscR
Avin49130
IscR
Avin16910
IscR
Avin12430
IscR
Avin41790
IscR
Avin51780
start codon
0
100

**Figure 11**



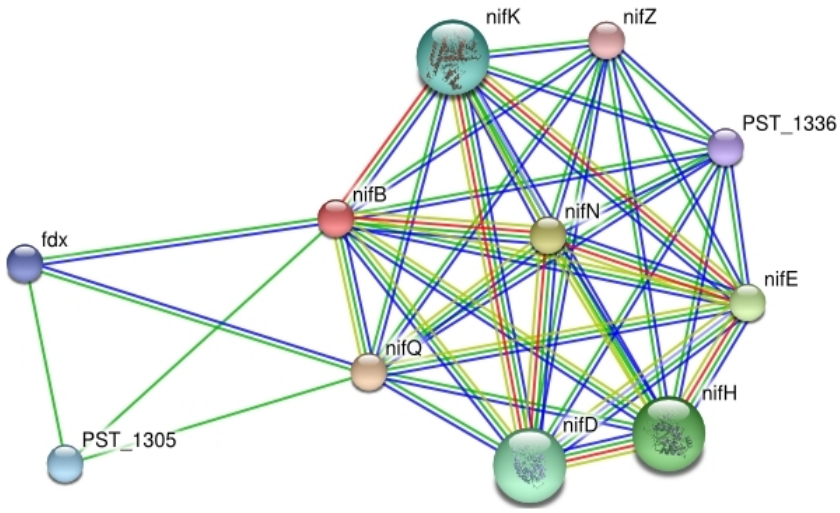

**Your Input:**

| | | Neighborhood | Gene Fusion | Cooccurrence | Coexpression | Experiments | Databases | Textmining | [Homology] | Score |
|---|---|---|---|---|---|---|---|---|---|---|
| rplL | 50S ribosomal protein L7/L12; Seems to be the binding site for several of the factors involved in protein synthesis and appears to be essential for accurate translation (By similarity) (122 aa) *(Pseudomonas aeruginosa PA14)* | | | | | | | | | |

**Predicted Functional Partners:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rplI | 50S ribosomal protein L9; Binds to the 23S rRNA (By similarity) (148 aa) | | | | | • | | | | 0.999 |
| rplJ | 50S ribosomal protein L10 (166 aa) | • | | | | • | | • | | 0.996 |
| rplK | 50S ribosomal protein L11; This protein binds directly to 23S ribosomal RNA (By similarity) (143 aa) | • | | | | • | | • | | 0.996 |
| rplA | 50S ribosomal protein L1; Binds directly to 23S rRNA. The L1 stalk is quite mobile in the ribos [...] (231 aa) | • | | | | • | | • | | 0.986 |
| rpoB | DNA-directed RNA polymerase subunit beta; DNA-dependent RNA polymerase catalyzes the transcript [...] (1357 aa) | • | | | | • | | • | | 0.985 |
| rplE | 50S ribosomal protein L5; This is 1 of the proteins that binds and probably mediates the attach [...] (179 aa) | • | | | | • | | • | | 0.985 |
| rpoC | DNA-directed RNA polymerase subunit beta'; DNA-dependent RNA polymerase catalyzes the transcrip [...] (1399 aa) | • | | | | • | | • | | 0.985 |
| rplX | 50S ribosomal protein L24; One of two assembly inititator proteins, it binds directly to the 5' [...] (104 aa) | • | | | • | • | | | | 0.982 |
| rplB | 50S ribosomal protein L2; One of the primary rRNA binding proteins. Required for association of [...] (273 aa) | • | | | | • | | | | 0.974 |
| rplF | 50S ribosomal protein L6; This protein binds to the 23S rRNA, and is important in its secondary [...] (177 aa) | • | | | | • | | | | 0.968 |
| rplP | 50S ribosomal protein L16; Binds 23S rRNA and is also seen to make contacts with the A and poss [...] (137 aa) | • | | | | • | | | | 0.967 |
| rplQ | 50S ribosomal protein L17 (129 aa) | • | | | | • | | | | 0.965 |
| rplN | 50S ribosomal protein L14; Binds to 23S rRNA. Forms part of two intersubunit bridges in the 70S [...] (122 aa) | • | | | | • | | | | 0.965 |
| rpsG | 30S ribosomal protein S7; One of the primary rRNA binding proteins, it binds directly to 16S rR [...] (156 aa) | • | | | | • | | | | 0.964 |
| rpsE | 30S ribosomal protein S5; With S4 and S12 plays an important role in translational accuracy (By [...] (166 aa) | • | | | | • | | | | 0.959 |
| rplC | 50S ribosomal protein L3; One of the primary rRNA binding proteins, it binds directly near the [...] (211 aa) | • | | | | • | | | | 0.959 |
| rpsQ | 30S ribosomal protein S17; One of the primary rRNA binding proteins, it binds specifically to t [...] (88 aa) | • | | | | | | • | | 0.958 |
| rpsC | 30S ribosomal protein S3; Binds the lower part of the 30S subunit head. Binds mRNA in the 70S r [...] (228 aa) | • | | | | • | | | | 0.956 |
| rplO | 50S ribosomal protein L15; Binds to the 23S rRNA (By similarity) (144 aa) | • | | | | • | | | | 0.946 |
| rplR | 50S ribosomal protein L18; This is one of the proteins that binds and probably mediates the att [...] (116 aa) | • | | | | • | | | | 0.939 |

**Figure 12**



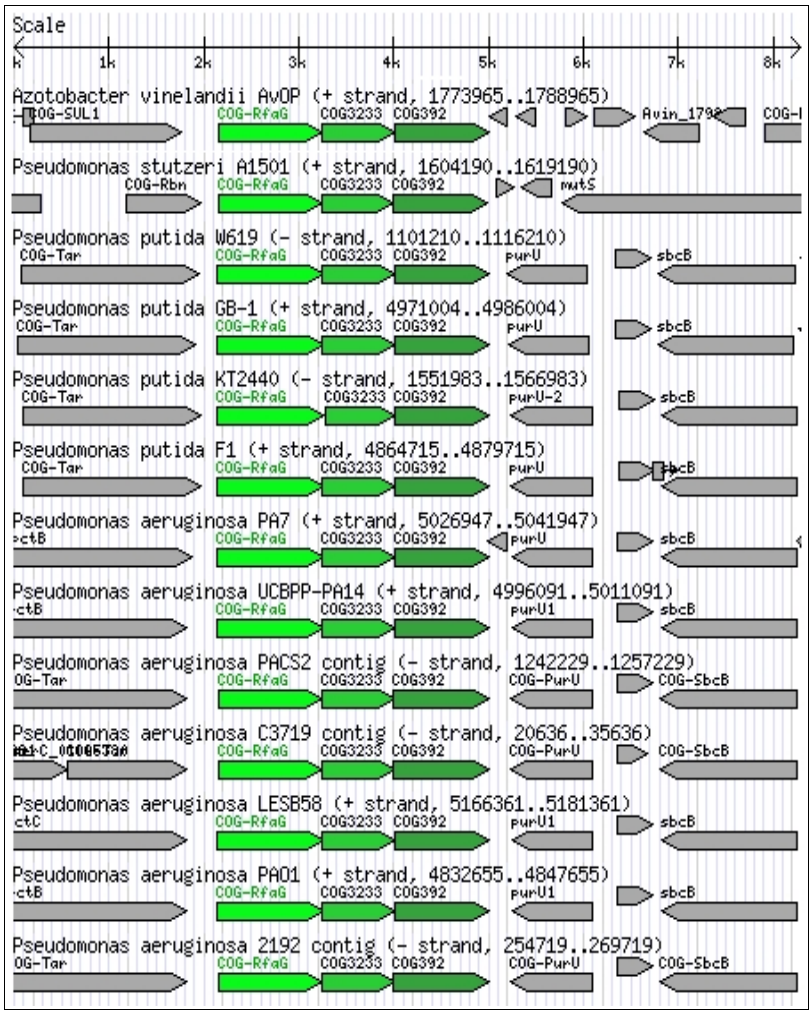

**Your Input:**

| | | Neighborhood | Gene Fusion | Cooccurrence | Coexpression | Experiments | Databases | Textmining | [Homology] | Score |
|---|---|---|---|---|---|---|---|---|---|---|
| nifB | FeMo cofactor biosynthesis protein NifB (447 aa) *(Pseudomonas stutzeri)* | | | | | | | | | |

**Predicted Functional Partners:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| nifQ | Nitrogen fixation protein NifQ (185 aa) | • | | • | | | | • | | 0.989 |
| nifN | Nitrogenase iron-molybdenum cofactor biosynthesis protein (458 aa) | • | • | • | | | | • | | 0.965 |
| nifE | Nitrogenase iron-molybdenum cofactor biosynthesis protein (472 aa) | • | | • | | | | • | | 0.955 |
| nifH | Nitrogenase iron protein; The key enzymatic reactions in nitrogen fixation are catalyzed by the [...] (293 aa) | • | | • | | | | • | | 0.954 |
| nifD | Nitrogenase protein alpha chain (493 aa) | • | | • | | | | • | | 0.950 |
| nifK | MoFe protein, beta subunit (523 aa) | • | • | • | | | | | | 0.948 |
| PST_1305 | arsenate reductase related protein (148 aa) | • | | | | | | | | 0.944 |
| fdx | Ferredoxin-like protein (92 aa) | • | | • | | | | | | 0.938 |
| PST_1336 | protein of unknown function DUF269 (159 aa) | • | | • | | | | | | 0.919 |
| nifZ | Fe-S cofactor synthesis protein (156 aa) | • | | • | | | | | | 0.909 |

**Figure 13**


Scale
k
1k
2k
3k
4k
5k
6k
7k
8k
Azotobacter vinelandii AvOP (+ strand, 1773965..1788965)
COG-SUL1
COG-RfaG
COG3233 COG392
Avin_1798
Pseudomonas stutzeri A1501 (+ strand, 1604190..1619190)
COG-Rbn
COG-RfaG
COG3233 COG392
mutS
Pseudomonas putida W619 (- strand, 1101210..1116210)
COG-Tar
COG-RfaG
COG3233 COG392
purU
sbcB
Pseudomonas putida GB-1 (+ strand, 4971004..4986004)
COG-Tar
COG-RfaG
COG3233 COG392
purU
sbcB
Pseudomonas putida KT2440 (- strand, 1551983..1566983)
COG-Tar
COG-RfaG
COG3233 COG392
purU-2
sbcB
Pseudomonas putida F1 (+ strand, 4864715..4879715)
COG-Tar
COG-RfaG
COG3233 COG392
purU
sbcB
Pseudomonas aeruginosa PA7 (+ strand, 5026947..5041947)
ctB
COG-RfaG
COG3233 COG392
purU
sbcB
Pseudomonas aeruginosa UCBPP-PA14 (+ strand, 4996091..5011091)
ctB
COG-RfaG
COG3233 COG392
purU1
sbcB
Pseudomonas aeruginosa PACS2 contig (- strand, 1242229..1257229)
OG-Tar
COG-RfaG
COG3233 COG392
COG-PurU
COG-SbcB
Pseudomonas aeruginosa C3719 contig (- strand, 20636..35636)
COG-RfaG
COG3233 COG392
COG-PurU
COG-SbcB
Pseudomonas aeruginosa LESB58 (+ strand, 5166361..5181361)
ctC
COG-RfaG
COG3233 COG392
purU1
sbcB
Pseudomonas aeruginosa PAO1 (+ strand, 4832655..4847655)
ctB
COG-RfaG
COG3233 COG392
purU1
sbcB
Pseudomonas aeruginosa 2192 contig (- strand, 254719..269719)
OG-Tar
COG-RfaG
COG3233 COG392
COG-PurU
COG-SbcB

# Tables

**Table 1: *Azotobacter vinelandii DJ* NADH-uibiquinone oxidoreductases encoding [Fe-S] subunits.**

Listed are the the ATP-uncoupled Nqr (membrane-associated) and Ndh systems, as well as the ATP-coupled Nuo-type oxidoreductase. Products labeled in gray denote predicted [2Fe-2S] proteins, while those in black stand for predicted [4Fe-4S] proteins. Some terms are abbreviated. Pfam: protein motifs description.

**Table 2: Rnf electron complex systems from *Azotobacter vinelandii DJ*.**

In both operons listed, subunits B and C, labeled in black, are characterized as [4Fe-4S] proteins. Some terms are abbreviated. Pfam: protein motifs description.

**Table 3: Cytochrome c reductase and cytochrome c oxidase cbb3 systems from *Azotobacter vinelandii DJ*.**

The Rieske subunit of complex III, PetA (Avin13060) is a [2Fe-2S] protein – labeled in gray. The accessory protein CcoG (Avin19970) to the terminal oxidase cbb3 is a predicted [4Fe-4S] protein, as depicted in black. Some terms are abbreviated. Pfam: protein motifs description.

**Table 4: Membrane bound Ni-dependent carbon monoxide dehydrogenase and accessory proteins from *Azotobacter vinelandii DJ*.**

Also shown is the transcriptional regulator *cooA* (Avin47010), distantly located from the *coo* operon. CooF (Avin04500) is a known [4Fe-4S] protein, while Avin04480 is

predicted to bind a cluster of the same nature; both proteins are highlighted in black. Some terms are abbreviated. Pfam: protein motifs description.

**Table 5: Formate dehydrogenase (FDH) and accessory proteins from *Azotobacter vinelandii DJ*.**

FdhH (Avin03820) is a [4Fe-4S] protein and the β subunit of FDH. FdhE (Avin03840) is also a [4Fe-4S] protein accessory to the process of FDH assembly. Some terms are abbreviated. Pfam: protein motifs description.

**Table 6: Thioredoxin-like genes encoded in Azotobacter vinelandii DJ genome.**

Products labeled in gray denote predicted [2Fe-2S] proteins. Some terms are abbreviated. Pfam: protein motifs description.

**Table 7: Glutaredoxin (Grx) homologs in *Azotobacter vinelandii DJ*.**

Avin04690 and Avin14040 are predicted to harbor [2Fe-2S] cofactors and are shaded in gray. Avin51060, the *nif*-associated Grx, is likely to ligate a [4Fe-4S] cluster (black). Some terms are abbreviated. Pfam: protein motifs description.

**Table 8: Molybdenum cofactor biosynthesis protein A (MoaA) homologs in *Azotobacter vinelandii DJ*.**

Avin03850 and Avin30330 are two known [4Fe-4S]-containing MoaA homologs, while Avin35520 is a similar protein with the same domains as the MoaA-like protein. Two other molecules share a more distant similarity with MoaA (Avin20630

and Avin20660) and together with Avin35520 are predicted to bind a [4Fe-4S] cluster. Pfam: protein motifs description.

**Table 9: *Azotobacter vinelandii DJ xdh* operon and three predicted xanthine dehydrogenase-like encoding operons.**

The three putative transcription units ( *Avin25820* to *Avin25850*, *Avin26060* and *Avin26070*, and *Avin35540* to *Avin35570*) bear consistent homology to *Brevundimonas diminuta* isoquinoline 1-oxidoreductase genes.

**Table 10: Genomic neighborhood of rhodanese-domain proteins in *Azotobacter vinelandii DJ*.**

The predicted sulfurtransferase elements whitin each operon are highlighted in gray. See text for details. Some terms are abbreviated. Pfam: protein motifs description.

**Table 11: *Azotobacter vinelandii DJ*'s genome scanning for IscR-regulated genes.**

(A): *A. vinelandii DJ* candidates whose genes are predicted to encode IscR-binding motifs, as computed by the MEME Suite. From 36 putative sites identified using the PSSM generated, those having p-values below a $1x10^{-5}$ threshold are listed here. (B): The procedure described in (A) was also done using the MREC tool. Candidates having similar IscR-binding motifs were grouped by MREC on a high significance dataset with a single p-value. Members predicted by both MEME Suite and MREC – depicted in gray in (A) and (B) – were designated as 'best hits' and were chosen for further computational investigations.

**Table 12: Genomic neighborhood analysis of the 10 genes having the best scored IscR-binding motifs in *Azotobacter vinelandii DJ*.**

The computations were carried out by both MEME Suite and MREC tool. Candidates are depicted in gray, while putative transcriptional partners are in white.

**Table 1**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| | | Membrane associated | |
| Avin14590 | nqrA | NQRA | NADH:ubq oxidored. Na(+) sub. A |
| Avin14600 | nqrB | NQR2_RnfD_RnfE | NADH:ubq oxidored. Na(+) sub. B |
| Avin14610 | nqrC | FMN_bind | NADH:ubq oxidored. Na(+) sub. C |
| Avin14620 | nqrD | Rnf-Nqr | NADH:ubq oxidored. Na(+) sub. D |
| Avin14630 | nqrE | Rnf-Nqr | NADH:ubq oxidored. Na(+) sub. E |
| Avin14640 | nqrF | Fer2 + FAD_6 + NAD_1 | NADH:ubq oxidored. Na(+) sub. F |
| | | Cytosolic | |
| Avin12000 | ndh | Pyr_redox_2 | Uncoupled NADH:quinone oxidored. |
| | | ATP-coupled | |
| Avin28440 | nuoA | Oxidored_q4 | NADH:ubq oxidored, sub. A |
| Avin28450 | nuoB | Oxidored_q6 | NADH:ubq oxidored, sub. B |
| Avin28460 | nuocd | Complex1_30kDa +_49kDa | NADH:ubq oxidored, sub. CD |
| Avin28470 | nuoE | Complex1_24kDa | NADH:ubq oxidored, sub. E |
| Avin28480 | nuoF | Complex1_51K | NADH:ubq oxidored, sub. F |
| Avin28490 | nuoG | Fer2 + Molybdop_Fe4S4 | NADH:ubq oxidored, sub. G |
| Avin28500 | nuoH | NADHdh | NADH:ubq oxidored, sub. H |
| Avin28510 | nuoI | Fer4 + Fer4 | NADH:ubq oxidored, sub. I |
| Avin28520 | nuoJ | Oxidored_q3 | NADH:ubq oxidored, sub. J |
| Avin28530 | nuoK | Oxidored_q2 | NADH:ubq oxidored, sub. K |
| Avin28540 | nuoL | Oxidored_q1_N + Oxidored_q1 | NADH:ubq oxidored, sub. L |
| Avin28550 | nuoM | Oxidored_q1 | NADH:ubq oxidored, sub. M |
| Avin28560 | nuoN | Oxidored_q1 | NADH:ubq oxidored, sub. N |

**Table 2**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| | | Rnf2 | |
| Avin19220 | rnfE | Rnf-Nqr | Electron transp. complex, sub. E |
| Avin19230 | rnfG | FMN_bind | Electron transp. complex, sub. G |
| Avin19240 | rnfD | NQR2_RnfD_RnfE | Electron transp. complex, sub. D |
| Avin19250 | rnfC | Complex1_51K + Fer4 + Fer4 | Electron transp. complex, sub. C |
| Avin19260 | rnfB | FeS + Fer4 + Fer4 | Electron transp. complex, sub. B |
| Avin19270 | rnfA | Rnf-Nqr | Electron transp. complex, sub. A |
| | | Rnf1 | |
| Avin50900 | | Nitro_FeMo-Co | $N_2$ fixation-related protein |
| Avin50910 | nafY | Nitro_FeMo-Co | $N_2$ fixation-related protein, γ sub. |
| Avin50920 | rnfH | UPF0125 | Electron transp. complex, sub. H |
| Avin50930 | rnfE1 | Rnf-Nqr | Electron transp. complex, sub. E |
| Avin50940 | rnfG1 | FMN_bind | Electron transp. complex, sub. G |
| Avin50950 | rnfD1 | NQR2_RnfD_RnfE | Electron transp. complex, sub. D |
| Avin50960 | rnfC1 | Complex1_51K + Fer4 + Fer4 | Electron transp. complex, sub. C |
| Avin50970 | rfnB1 | FeS + Fer4 + Fer4 | Electron transp. complex, sub. B |
| Avin50980 | rnfA1 | Rnf-Nqr | Electron transp. complex, sub. A |

**Table 3**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| | | Cytochrome c reductase | |
| Avin13060 | petA | Rieske | Ubq.-cyt. c reductase, [Fe-S] sub. |
| Avin13070 | petB | Cyt._B_N + Cyt._B_C | Cytochrome b/b6 |
| Avin13080 | petC | Cytochrom_C1 | Ubq.-cyt. c reductase, cytochrome c1 |
| | | Cytochrome terminal oxidase | |
| Avin19940 | ccoS | CcoS | Cytochrome c oxidase maturation ptn. |
| Avin19950 | ccoI | HMA + E1-E2_ATPase + Hydr. | Copper-translocating P-type ATPase |
| Avin19960 | ccoH | No | Cytochrome c oxidase accessory ptn. |
| Avin19970 | ccoG | Fer4 | Cytochrome c oxidase accessory ptn. |
| Avin19980 | ccoP | Cytochrom_C (2x) | Cytochrome c oxidase cbb3, sub. III |
| Avin19990 | ccoQ | No | Cytochrome c oxidase cbb3, sub. IV |
| Avin20000 | ccoO | FixO | Cytochrome c oxidase cbb3, sub. II |
| Avin20010 | ccoN | COX1 | Cytochrome c oxidase cbb3, sub. I |

**Table 4**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin04450 | | HupE_UreJ | Nickel transp. for CO dehydrogenase |
| Avin04460 | | No | Hypothetical protein |
| Avin04470 | cooC | CbiA | CO dehydrogenase assesory protein |
| Avin04480 | | Pyr_redox_2 | FAD/pir. nucleotide-disulph. Oxidored. |
| Avin04490 | cooS | Prismane | CO dehydrogenase, catalytic subunit |
| Avin04500 | cooF | Fer4 | Fe-S binding protein |
| Avin47010 | cooA | Crp | CO oxidation regulator, Crp/Fnr |

**Table 5**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin03800 | fdhD | FdhD-NarQ | Formate dehydr. accessory ptn. |
| Avin03810 | fdhG | Mo_Fe4S4 + Mo + Mo_binding | Formate dehydrogenase, α subunit |
| Avin03820 | fdhH | Fer4 | Formate dehydrogenase, β subunit |
| Avin03830 | fdhI | Cytochrom_B_N | Formate dehydrogenase, γ subunit |
| Avin03840 | fdhE | FdhE | Formate dehydr. accessory ptn. |

**Table 6**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin06670 | | Thioredoxin | Thioredoxin protein |
| Avin45140 | | Thioredoxin | Thioredoxin 2, TrxC |
| Avin47420 | trxA | Thioredoxin | Thioredoxin 1, Trx1 |

**Table 7**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin04680 | | Rhodanese | Rhodanese-like protein |
| Avin04690 | | Glutaredoxin | Dithiol-glutaredoxin protein |
| Avin04700 | secB | SecB | Protein-export chaperone |
| Avin14040 | | Glutaredoxin | Glutaredoxin 4 protein |
| Avin50990 | nifL | PAS + PAS + HATPase_c | Regulatory protein |
| Avin51000 | nifA | GAF + Sigma54_activat + HTH_8 | Sigma54 transc. Activator |
| Avin51010 | nifB | Radical_SAM + Nitro_FeMo-Co | Nitrogenase cofactor |
| Avin51020 | | Fer4 | Ferredoxin |
| Avin51030 | nifO | ArsC | Nitrogenase-associated |
| Avin51050 | rhdN | Rhodanese | Rhodanese/ sulfurtransferase |
| Avin51060 | | Glutaredoxin | Glutaredoxin-related protein |

**Table 8**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin03850 | moaA | Radical_SAM + Mob_synth_C | Mo cofactor biosynthesis protein A |
| Avin20620 | | Glycos_transf_2 | Glycosyl transferase, family 2 |
| Avin20630 | | Radical_SAM | MoaA, NifB, PqqE family protein |
| Avin20650 | | Glycos_transf_2 | Glycosyl transferase, family 2 |
| Avin20660 | | Radical_SAM | MoaA, NifB, PqqE family protein |
| Avin20670 | | YdjC | Conserved hypothetical protein |
| Avin20680 | | No | Conserved hypothetical protein |
| Avin30330 | moaA | Radical_SAM + Mob_synth_C | Mo cofactor biosynthesis protein A |
| Avin35520 | | Radical_SAM + Mob_synth_C | Mo cofactor biosynthesis protein |

**Table 9**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin22270 | xdhC | XdhC_CoxI | Xanthine dehydr. accessory ptn. |
| Avin22280 | xdhB | Ald_Xan_dh_C + _C2 | Xanthine dehydr. Mo-pterin bind. Sub. |
| Avin22290 | xdhA | Fer2 + _2 + CO_deh_flav_C | Xanthine dehydr. small sub. |
| Avin25820 | | No | Conserved hypothetical protein |
| Avin25830 | | XdhC_CoxI | Conserved hypothetical protein |
| Avin25840 | | Ald_Xan_dh_C2 (3x) | Oxidored. Mo-pterin-binding subunit |
| Avin25850 | | Fer2 + Fer2_2 | Ferredoxin |
| Avin26060 | | Fer2 + Fer2_2 | Oxidoreductase Fe-S binding subunit |
| Avin26070 | | Ald_Xan_dh_C2 (3x) | Oxidoreductase |
| Avin35540 | | Fer2 + Fer2_2 | Fe-S binding oxidoreductase |
| Avin35550 | | Ald_Xan_dh_C2 (3x) | Oxidored., xanthine dehydr.-like |
| Avin35560 | | No | Hypothetical protein |
| Avin35570 | | Cupin_2 | Conserved hypothetical protein |

**Table 10**

| Locus Tag | Gene | Pfam | Product |
|---|---|---|---|
| Avin02700 | moeB2 | ThiF + MoeZ_MoeB + Rhodan. | Molybdopterin biosyn. protein |
| Avin02710 | | ThiF + MoeZ_MoeB | Molybdopterin biosyn. protein |
| Avin02720 | | ThiS | Molybdopterin biosyn. protein |
| Avin02730 | | No | Conserved hypothetical protein |
| Avin04680 | | Rhodanese | Rhodanese-like protein |
| Avin04690 | | Glutaredoxin | Dithiol-glutaredoxin protein |
| Avin04700 | secB | SecB | Protein-export chaperone |
| Avin07440 | psd | PS_Decarbxylase | Phosphatidylserine decarboxylase |
| Avin07450 | rhdA | Rhodanese + Rhodanese | Rhodanese |
| Avin20950 | | PP2C | Ser/Thr phosphatase, 2C-like |
| Avin20960 | | No | Hypothetical protein |
| Avin20970 | rhdE | Rhodanese | Rhodanese-like |
| Avin20980 | cysE3 | Hexapep (4x) | Serine O-acetyltransferase |
| Avin21500 | | Rhodanese (3x) | Rhodanese-like protein |
| Avin21510 | metC | Cys_Met_Meta_PP | Cystathionine beta-lyase |
| Avin30940 | | No | Hypothetical protein |
| Avin30950 | | No | Hypothetical protein |
| Avin30960 | | Rhodanese (4x) | Rhodanese-like protein |
| Avin30970 | | CDO_I | Cysteine dioxygenase type I |
| Avin32980 | | Rhodanese + Rhodanese | 3-mercaptopyruvate sulfurtransferase |
| Avin46800 | lptD | OstA_C | Organic solvent tolerance |
| Avin46810 | | Rotamase + Rotamase | Pep-prolyl cis-trans isomerase |
| Avin46820 | pdxA | PdxA | PLP biosynthetic protein |
| Avin46830 | ksgA | RrnaAD | dimethyladenosine transf. |
| Avin46840 | apaG | DUF525 | ApaG-protein |
| Avin46850 | apaH | Metallophos | bis(5-nucleosyl)-tetraphos. |
| Avin46860 | glpE | Rhodanese | Thiosulfate sulfurtransferase |
| Avin46910 | | Lactamase_B | Beta-lactamase-like protein |
| Avin46920 | | STAS | Sulphate transporter |
| Avin46930 | | Rhodanese | Rhodanese-domain protein |
| Avin48680 | | Nitroreductase (2x) | McbC-like oxidoreductase |
| Avin48690 | rhdK | Rhodanese + Ank (3x) | Rhodanese with ankyrin repeat |
| Avin48700 | | Nitro_FeMo-Co | Metallocluster binding |
| Avin48710 | | No | Conserved hyp. protein |
| Avin48720 | | GST_N + GST_C | Glutathione S-transferase |
| Avin50990 | nifL | PAS + PAS + HATPase_c | Nitrogen fixing regulatory ptn. |
| Avin51000 | nifA | GAF + Sigma54_activat + HTH_8 | Nif-specif. Sigma54 act. |
| Avin51010 | nifE | Radical_SAM + Nitro_FeMo-Co | Nitrogenase cofac. Biosyn. |
| Avin51020 | | Fer4 | Ferredoxin |
| Avin51030 | nifO | ArsC | Nitrogenase-associated |
| Avin51050 | rhdN | Rhodanese | Rhodanese/sulfurtransferase |
| Avin51060 | | Glutaredoxin | Glutaredoxin-related |

## Table 11

**A) MEME order of significance ($10^{-5}$ treshold)**

| Product | Pfam | Locus tag | p-value |
|---|---|---|---|
| IscR | Rrf2 | Avin40410 | 3,47E-011 |
| Glycosyl transferase, group 1 | Glycos_transf_1 | Avin17910 | 1,09E-009 |
| 50S ribosomal protein L7/L12 | Ribosomal_L12 | Avin06160 | 7,97E-008 |
| Nitrogenase cofactor NifB | Radical_SAM + Nitro_FeMo-Co | Avin51010 | 4,71E-007 |
| Glutathione-disulfide reductase | Pyr_redox_2 + Pyr_redox_dim | Avin25010 | 8,97E-007 |
| Hypothetical protein | None | Avin26260 | 1,22E-006 |
| Hypothetical protein | None | Avin49130 | 1,91E-006 |
| Hypothetical protein | None | Avin16910 | 2,94E-006 |
| TonB-dependent siderophore receptor | Plug + TonB_dep_Rec | Avin12430 | 3,88E-006 |
| Cell division protein FtsK | FtsK_SpoIIIE | Avin28220 | 6,62E-006 |
| Carbohydrate transport membrane protein | MSF_1 | Avin41790 | 1,10E-005 |
| UDP-glucose 4-epimerase | Epimerase | Avin44740 | 1,24E-005 |
| Transcriptional regulator, LacI family | LacI + Peripla_BP_1 | Avin27300 | 1,78E-005 |
| Hypothetical protein | None | Avin40100 | 1,78E-005 |
| Alpha-glucosidase | Glyco_hydro_97 | Avin51780 | 2,00E-005 |
| Membrane protein, Auxin Efflux Carrier | Auxin_eff | Avin12580 | 2,25E-005 |
| General glycosylation protein | None | Avin33460 | 3,15E-005 |
| AA ABC transporter periplasmic | None | Avin16838 | 4,83E-005 |
| Transposase | None | Avin31320 | 5,95E-005 |
| ABC transp. substrate binding, family 3 | SBP_bac_3 | Avin21410 | 9,75E-005 |

**B) MREC order of significance**

p-value = 1,75E-015

| Product | Pfam | Locus tag |
|---|---|---|
| IscR | Rrf2 | Avin40410 |
| TonB-dependent siderophore receptor | Plug + TonB_dep_Rec | Avin12430 |
| Hypothetical protein | None | Avin26260 |
| 50S ribosomal protein L7/L12 | Ribosomal_L12 | Avin06160 |
| Carbohydrate transport membrane protein | MSF_1 | Avin41790 |
| Hypothetical protein | None | Avin16910 |
| Glycosyl transferase, group 1 | Glycos_transf_1 | Avin17910 |
| Nitrogenase cofactor NifB | Radical_SAM + Nitro_FeMo-Co | Avin51010 |
| Glutathione-disulfide reductase | Pyr_redox_2 + Pyr_redox_dim | Avin25010 |
| Molybdate ABC transporter | None | Avin50730 |
| Sensory histidine protein kinase | PAS (2x) + HisKA + HATPase_c | Avin25270 |
| Transcriptional regulator protein DgoR | IclR | Avin51350 |
| Sugar isomerase (SIS) | HTH_6 + SIS | Avin50020 |
| Lysine 2,3-aminomutase | None | Avin41820 |
| ABC transp. aliph. sulfonate subs.-binding | None | Avin31680 |
| Hypothetical protein | None | Avin49130 |
| Bacterial regulatory protein, GntR family | GntR + FCD | Avin26860 |
| Alpha-glucosidase | Glyco_hydro_97 | Avin51780 |
| Sigma54-dep. Activ. XylR/DmpR family | XylR_N + V4R + Sigma54_activat | Avin08880 |

**Table 12**

| Product | Pfam | Locus tag |
|---|---|---|
| Glycosyl transferase, group 1 | Glycos_transf_1 | Avin17910 |
| Conserved hypothetical protein | DUF2334 | Avin17920 |
| Conserved hypothetical protein | UPF0104 | Avin17930 |
| Preprotein translocase, SecE subunit | SecE | Avin06100 |
| Transcription term./antiterm. factor NusG | NusG + KOW | Avin06110 |
| 50S ribosomal protein L11 | Ribosomal_L11_N + Ribosomal_L11 | Avin06120 |
| ribosomal protein L1 | Ribosomal_L1 | Avin06130 |
| 50S ribosomal protein L10 | Ribosomal_L10 | Avin06150 |
| 50S ribosomal protein L7/L12 | Ribosomal_L12 | Avin06160 |
| P26 sRNA | | Avin65060 |
| Regulatory protein | PAS + PAS + HATPase_c | Avin50990 |
| Sigma54 transc. Activator | GAF + Sigma54_activat + HTH_8 | Avin51000 |
| Nitrogenase cofactor NifB | Radical_SAM + Nitro_FeMo-Co | Avin51010 |
| Ferredoxin | Fer4 | Avin51020 |
| Nitrogenase-associated | ArsC | Avin51030 |
| Rhodanese/sulfurtransferase | Rhodanese | Avin51050 |
| Glutaredoxin-related | Glutaredoxin | Avin51060 |
| Glutathione-disulfide reductase | Pyr_redox_2 + Pyr_redox_dim | Avin25010 |
| Hypothetical protein | None | Avin26260 |
| Hypothetical protein | None | Avin49130 |
| Hypothetical protein | None | Avin16910 |
| TonB-dependent siderophore receptor | Plug + TonB_dep_Rec | Avin12430 |
| Carbohydrate transport membrane protein | MFS_1 | Avin41790 |
| Alpha-glucosidase | Glyco_hydro_97 | Avin51780 |

# Conclusão e perspectivas

Utilizando uma ampla abordagem computacional, o presente trabalho dá um passo à frente na caracterização funcional do genoma da bactéria fixadora de nitrogênio *A. vinelandii*. Do ponto de vista do metabolismo de cofatores [Fe-S], novos elementos putativos foram identificados e catalogados. Além disso, buscou-se compreender melhor a rede regulatória da proteína IscR, cujo mecanismo de ação envolve um centro [2Fe-2S]. A partir deste estudo, será possível investigar a atividade predita de genes candidatos, assim como avaliar novos processos metabólicos dependentes de grupamentos [Fe-S].

Como perspectiva de validação dos resultados aqui apresentados, entre outros objetivos, nosso grupo de pesquisa desenvolve no momento um grande projeto de transcriptômica de *A. vinelandii*. Com o advento de novas tecnologias de sequenciamento de ácidos nucleicos capazes de gerar milhões de leituras em poucas horas, estudos práticos de genomas inteiros são agora possíveis. Além disso, considerável avanço no campo de enriquecimento de RNAs mensageiros procarióticos a partir de extratos de RNA total tornou a transcriptômica bacteriana em larga escala uma realidade (85). Ainda que tal estudo contribua de forma significativa para o melhor entendimento do metabolismo de cofatores [Fe-S], um experimento dessa natureza certamente irá agregar substancial conhecimento à biologia geral de *A. vinelandii*.

Referências bibliográficas adicionais

1. Johnson, D.C., Dean, D.R., Smith, A.D. & Johnson, M.K. Structure, function, and formation of biological iron-sulfur clusters. Annu. Rev. Biochem. 74, 247-81(2005).

2. Beinert, H., Holm, R.H. & Münck, E. Iron-sulfur clusters: nature's modular, multipurpose structures. Science 277, 653-9(1997).

3. Beinert, H. A tribute to sulfur. Eur. J. Biochem. 267, 5657-64(2000).

4. Glaser, T., Hedman, B., Hodgson, K.O. & Solomon, E.I. Ligand K-edge X-ray absorption spectroscopy: a direct probe of ligand-metal covalency. Acc. Chem. Res. 33, 859-68(2000).

5. Peters, J.W., Stowell, M.H., Soltis, S.M., Finnegan, M.G., Johnson, M.K. & Rees, D.C. Redox-dependent structural changes in the nitrogenase P-cluster. Biochemistry 36, 1181-7(1997).

6. Beinert, H., Kennedy, M.C. & Stout, C.D. Aconitase as Ironminus signSulfur Protein, Enzyme, and Iron-Regulatory Protein. Chem. Rev. 96, 2335-2374(1996).

7. Jarrett, J.T. The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. Curr Opin Chem Biol 7, 174-82(2003).

8. Cheek, J. & Broderick, J.B. Adenosylmethionine-dependent iron-sulfur enzymes: versatile clusters in a radical new role. J. Biol. Inorg. Chem. 6, 209-26(2001).

9. Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R. & Meyer, O. Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. Science 293, 1281-5(2001).

10. Crane, B.R., Siegel, L.M. & Getzoff, E.D. Sulfite reductase structure at 1.6 A: evolution and catalysis for reduction of inorganic anions. Science 270,

59-67(1995).

11. Darnault, C., Volbeda, A., Kim, E.J., Legrand, P., Vernède, X., Lindahl, P.A. et al. Ni-Zn-[Fe4-S4] and Ni-Ni-[Fe4-S4] clusters in closed and open subunits of acetyl-CoA synthase/carbon monoxide dehydrogenase. Nat. Struct. Biol. 10, 271-9(2003).

12. Svetlitchnyi, V., Dobbek, H., Meyer-Klaucke, W., Meins, T., Thiele, B., Römer, P. et al. A functional Ni-Ni-[4Fe-4S] cluster in the monomeric acetyl-CoA synthase from Carboxydothermus hydrogenoformans. Proc. Natl. Acad. Sci. U.S.A. 101, 446-51(2004).

13. Doukov, T.I., Iverson, T.M., Seravalli, J., Ragsdale, S.W. & Drennan, C.L. A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. Science 298, 567-72(2002).

14. Nicolet, Y., Cavazza, C. & Fontecilla-Camps, J.C. Fe-only hydrogenases: structure, function and evolution. J. Inorg. Biochem. 91, 1-8(2002).

15. Peters, J.W., Lanzilotta, W.N., Lemon, B.J. & Seefeldt, L.C. X-ray crystal structure of the Fe-only hydrogenase (CpI) from Clostridium pasteurianum to 1.8 angstrom resolution. Science 282, 1853-8(1998).

16. Plank, D.W., Kennedy, M.C., Beinert, H. & Howard, J.B. Cysteine labeling studies of beef heart aconitase containing a 4Fe, a cubane 3Fe, or a linear 3Fe cluster. J. Biol. Chem. 264, 20385-93(1989).

17. Golinelli, M.P., Chatelet, C., Duin, E.C., Johnson, M.K. & Meyer, J. Extensive ligand rearrangements around the [2Fe-2S] cluster of Clostridium pasteurianum ferredoxin. Biochemistry 37, 10429-37(1998).

18. Cunningham, R.P., Asahara, H., Bank, J.F., Scholes, C.P., Salerno, J.C., Surerus, K. et al. Endonuclease III is an iron-sulfur protein. Biochemistry 28, 4450-5(1989).

19. Kuo, C.F., McRee, D.E., Fisher, C.L., O'Handley, S.F., Cunningham,

R.P. & Tainer, J.A. Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. Science 258, 434-40(1992).

20. Demple, B., Ding, H. & Jorgensen, M. Escherichia coli SoxR protein: sensor/transducer of oxidative stress and nitric oxide. Meth. Enzymol. 348, 355-64(2002).

21. Kiley, P.J. & Beinert, H. The role of Fe-S proteins in sensing and regulation in bacteria. Curr. Opin. Microbiol. 6, 181-5(2003).

22. Malkin, R. & Rabinowitz, J.C. The reconstitution of clostridial ferredoxin. Biochem. Biophys. Res. Commun. 23, 822-7(1966).

23. Rees, D.C. & Howard, J.B. Nitrogenase: standing at the crossroads. Curr Opin Chem Biol 4, 559-66(2000).

24. Arnold, W., Rump, A., Klipp, W., Priefer, U.B. & Pühler, A. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae. J. Mol. Biol. 203, 715-38(1988).

25. Jacobson, M.R., Cash, V.L., Weiss, M.C., Laird, N.F., Newton, W.E. & Dean, D.R. Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii. Mol. Gen. Genet. 219, 49-57(1989).

26. Jacobson, M.R., Brigle, K.E., Bennett, L.T., Setterquist, R.A., Wilson, M.S., Cash, V.L. et al. Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii. J. Bacteriol. 171, 1017-27(1989).

27. Zheng, L., White, R.H., Cash, V.L., Jack, R.F. & Dean, D.R. Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 90, 2754-8(1993).

28. Evans, D.J., Jones, R., Woodley, P.R., Wilborn, J.R. & Robson, R.L. Nucleotide sequence and genetic analysis of the Azotobacter chroococcum nifUSVWZM gene cluster, including a new gene (nifP) which encodes a serine acetyltransferase. J. Bacteriol. 173, 5457-69(1991).

29. Fu, W., Jack, R.F., Morgan, T.V., Dean, D.R. & Johnson, M.K. nifU gene product from Azotobacter vinelandii is a homodimer that contains two identical [2Fe-2S] clusters. Biochemistry 33, 13455-63(1994).

30. Beynon, J., Ally, A., Cannon, M., Cannon, F., Jacobson, M., Cash, V. et al. Comparative organization of nitrogen fixation-specific genes from Azotobacter vinelandii and Klebsiella pneumoniae: DNA sequence of the nifUSV genes. J. Bacteriol. 169, 4024-9(1987).

31. Hwang, D.M., Dempsey, A., Tan, K.T. & Liew, C.C. A modular domain of NifU, a nitrogen fixation cluster protein, is highly conserved in evolution. J. Mol. Evol. 43, 536-40(1996).

32. Agar, J.N., Yuvaniyama, P., Jack, R.F., Cash, V.L., Smith, A.D., Dean, D.R. et al. Modular organization and identification of a mononuclear iron-binding site within the NifU protein. J. Biol. Inorg. Chem. 5, 167-77(2000).

33. Dos Santos, P.C., Smith, A.D., Frazzon, J., Cash, V.L., Johnson, M.K. & Dean, D.R. Iron-sulfur cluster assembly: NifU-directed activation of the nitrogenase Fe protein. J. Biol. Chem. 279, 19705-11(2004).

34. Nishio, K. & Nakai, M. Transfer of iron-sulfur cluster from NifU to apoferredoxin. J. Biol. Chem. 275, 22615-8(2000).

35. Léon, S., Touraine, B., Ribot, C., Briat, J. & Lobréaux, S. Iron-sulphur cluster assembly in plants: distinct NFU proteins in mitochondria and plastids from Arabidopsis thaliana. Biochem. J. 371, 823-30(2003).

36. Tong, W., Jameson, G.N.L., Huynh, B.H. & Rouault, T.A. Subcellular compartmentalization of human Nfu, an iron-sulfur cluster scaffold protein, and its ability to assemble a [4Fe-4S] cluster. Proc. Natl. Acad. Sci. U.S.A. 100, 9762-7(2003).

37. Olson, J.W., Agar, J.N., Johnson, M.K. & Maier, R.J. Characterization of the NifU and NifS Fe-S cluster formation proteins essential for viability in Helicobacter pylori. Biochemistry 39, 16213-9(2000).

38. Ali, V., Shigeta, Y., Tokumoto, U., Takahashi, Y. & Nozaki, T. An intestinal parasitic protist, Entamoeba histolytica, possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions. J. Biol. Chem. 279, 16863-74(2004).

39. Zheng, L., Cash, V.L., Flint, D.H. & Dean, D.R. Assembly of iron-sulfur clusters. Identification of an iscSUA-hscBA-fdx gene cluster from Azotobacter vinelandii. J. Biol. Chem. 273, 13264-72(1998).

40. Schwartz, C.J., Giel, J.L., Patschkowski, T., Luther, C., Ruzicka, F.J., Beinert, H. et al. IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins. Proc. Natl. Acad. Sci. U.S.A. 98, 14895-900(2001).

41. Frazzon, J. & Dean, D.R. Feedback regulation of iron-sulfur cluster biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 98, 14751-3(2001).

42. Krebs, C., Agar, J.N., Smith, A.D., Frazzon, J., Dean, D.R., Huynh, B.H. et al. IscA, an alternate scaffold for Fe-S cluster biosynthesis. Biochemistry 40, 14069-80(2001).

43. Vinella, D., Brochier-Armanet, C., Loiseau, L., Talla, E. & Barras, F. Iron-sulfur (Fe/S) protein biogenesis: phylogenomic and genetic studies of A-type carriers. PLoS Genet. 5, e1000497(2009).

44. Jung, Y.S., Gao-Sheridan, H.S., Christiansen, J., Dean, D.R. & Burgess, B.K. Purification and biophysical characterization of a new [2Fe-2S] ferredoxin from Azotobacter vinelandii, a putative [Fe-S] cluster assembly/repair protein. J. Biol. Chem. 274, 32402-10(1999).

45. Tokumoto, U. & Takahashi, Y. Genetic analysis of the isc operon in Escherichia coli involved in the biogenesis of cellular iron-sulfur proteins. J Biochem 130, 63-71(2001).

46. Tokumoto, U., Nomura, S., Minami, Y., Mihara, H., Kato, S., Kurihara, T. et al. Network of protein-protein interactions among iron-sulfur cluster assembly proteins in Escherichia coli. J Biochem 131, 713-9(2002).

47. Shimomura, Y., Takahashi, Y., Kakuta, Y. & Fukuyama, K. Crystal structure of Escherichia coli YfhJ protein, a member of the ISC machinery involved in assembly of iron-sulfur clusters. Proteins 60, 566-9(2005).

48. Pastore, C., Adinolfi, S., Huynen, M.A., Rybin, V., Martin, S., Mayer, M. et al. YfhJ, a molecular adaptor in iron-sulfur cluster formation or a frataxin-like protein?. Structure 14, 857-67(2006).

49. Kawula, T.H. & Lelivelt, M.J. Mutations in a gene encoding a new Hsp70 suppress rapid DNA inversion and bgl activation, but not proU derepression, in hns-1 mutant Escherichia coli. J. Bacteriol. 176, 610-9(1994).

50. Seaton, B.L. & Vickery, L.E. A gene encoding a DnaK/hsp70 homolog in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 91, 2066-70(1994).

51. Lelivelt, M.J. & Kawula, T.H. Hsc66, an Hsp70 homolog in Escherichia coli, is induced by cold shock but not by heat shock. J. Bacteriol. 177, 4900-7(1995).

52. Vickery, L.E., Silberg, J.J. & Ta, D.T. Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system from Escherichia coli. Protein Sci. 6, 1047-56(1997).

53. Silberg, J.J., Hoff, K.G. & Vickery, L.E. The Hsc66-Hsc20 chaperone system in Escherichia coli: chaperone activity and interactions with the DnaK-DnaJ-grpE system. J. Bacteriol. 180, 6617-24(1998).

54. Nakamura, M., Saeki, K. & Takahashi, Y. Hyperproduction of recombinant ferredoxins in escherichia coli by coexpression of the ORF1-ORF2-iscS-iscU-iscA-hscB-hs cA-fdx-ORF3 gene cluster. J Biochem 126, 10-8(1999).

55. Takahashi, Y. & Nakamura, M. Functional assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster involved in the assembly of Fe-S clusters in Escherichia coli. J Biochem 126, 917-26(1999).

56. Hoff, K.G., Ta, D.T., Tapley, T.L., Silberg, J.J. & Vickery, L.E. Hsc66

substrate specificity is directed toward a discrete region of the iron-sulfur cluster template protein IscU. J. Biol. Chem. 277, 27353-9(2002).

57. Hoff, K.G., Cupp-Vickery, J.R. & Vickery, L.E. Contributions of the LPPVK motif of the iron-sulfur template protein IscU to interactions with the Hsc66-Hsc20 chaperone system. J. Biol. Chem. 278, 37582-9(2003).

58. Hoff, K.G., Silberg, J.J. & Vickery, L.E. Interaction of the iron-sulfur cluster assembly protein IscU with the Hsc66/Hsc20 molecular chaperone system of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 97, 7790-5(2000).

59. Mühlenhoff, U., Gerber, J., Richhardt, N. & Lill, R. Components involved in assembly and dislocation of iron-sulfur clusters on the scaffold protein Isu1p. EMBO J. 22, 4815-25(2003).

60. Schwartz, C.J., Djaman, O., Imlay, J.A. & Kiley, P.J. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 97, 9009-14(2000).

61. Patzer, S.I. & Hantke, K. SufS is a NifS-like protein, and SufD is necessary for stability of the [2Fe-2S] FhuF protein in Escherichia coli. J. Bacteriol. 181, 3307-9(1999).

62. Müller, K., Matzanke, B.F., Schünemann, V., Trautwein, A.X. & Hantke, K. FhuF, an iron-regulated protein of Escherichia coli with a new type of [2Fe-2S] center. Eur. J. Biochem. 258, 1001-8(1998).

63. Zheng, M., Wang, X., Templeton, L.J., Smulski, D.R., LaRossa, R.A. & Storz, G. DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. J. Bacteriol. 183, 4562-70(2001).

64. Yeo, W., Lee, J., Lee, K. & Roe, J. IscR acts as an activator in response to oxidative stress for the suf operon encoding Fe-S assembly proteins. Mol. Microbiol. 61, 206-18(2006).

65. Rincon-Enriquez, G., Crété, P., Barras, F. & Py, B. Biogenesis of Fe/S proteins and pathogenicity: IscR plays a key role in allowing Erwinia

chrysanthemi to adapt to hostile conditions. Mol. Microbiol. 67, 1257-73(2008).

66. Wang, T., Shen, G., Balasubramanian, R., McIntosh, L., Bryant, D.A. & Golbeck, J.H. The sufR gene (sll0088 in Synechocystis sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria. J. Bacteriol. 186, 956-67(2004).

67. Nachin, L., El Hassouni, M., Loiseau, L., Expert, D. & Barras, F. SoxR-dependent response to oxidative stress and virulence of Erwinia chrysanthemi: the key role of SufC, an orphan ABC ATPase. Mol. Microbiol. 39, 960-72(2001).

68. Nachin, L., Loiseau, L., Expert, D. & Barras, F. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. EMBO J. 22, 427-37(2003).

69. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. & Barras, F. Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. J. Biol. Chem. 278, 38352-9(2003).

70. Outten, F.W., Wood, M.J., Munoz, F.M. & Storz, G. The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in Escherichia coli. J. Biol. Chem. 278, 45713-9(2003).

71. Rangachari, K., Davis, C.T., Eccleston, J.F., Hirst, E.M.A., Saldanha, J.W., Strath, M. et al. SufC hydrolyzes ATP and interacts with SufB from Thermotoga maritima. FEBS Lett. 514, 225-8(2002).

72. Layer, G., Gaddam, S.A., Ayala-Castro, C.N., Ollagnier-de Choudens, S., Lascoux, D., Fontecave, M. et al. SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly. J. Biol. Chem. 282, 13342-50(2007).

73. Takahashi, Y. & Tokumoto, U. A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids. J. Biol. Chem.

277, 28380-3(2002).

74. Albrecht, A.G., Netz, D.J.A., Miethke, M., Pierik, A.J., Burghaus, O., Peuckert, F. et al. SufU is an essential iron-sulfur cluster scaffold protein in Bacillus subtilis. J. Bacteriol. 192, 1643-51(2010).

75. Riboldi, G.P., Verli, H. & Frazzon, J. Structural studies of the Enterococcus faecalis SufU [Fe-S] cluster protein. BMC Biochem. 10, 3(2009).

76. Bandyopadhyay, S., Naik, S.G., O'Carroll, I.P., Huynh, B., Dean, D.R., Johnson, M.K. et al. A proposed role for the Azotobacter vinelandii NfuA protein as an intermediate iron-sulfur cluster carrier. J. Biol. Chem. 283, 14092-9(2008).

77. Angelini, S., Gerez, C., Ollagnier-de Choudens, S., Sanakis, Y., Fontecave, M., Barras, F. et al. NfuA, a new factor required for maturing Fe/S proteins in Escherichia coli under oxidative stress and iron starvation conditions. J. Biol. Chem. 283, 14084-91(2008).

78. Skovran, E. & Downs, D.M. Lack of the ApbC or ApbE protein results in a defect in Fe-S cluster metabolism in Salmonella enterica serovar Typhimurium. J. Bacteriol. 185, 98-106(2003).

79. Boyd, J.M., Lewis, J.A., Escalante-Semerena, J.C. & Downs, D.M. Salmonella enterica requires ApbC function for growth on tricarballylate: evidence of functional redundancy between ApbC and IscU. J. Bacteriol. 190, 4596-602(2008).

80. Boyd, J.M., Pierik, A.J., Netz, D.J.A., Lill, R. & Downs, D.M. Bacterial ApbC can bind and effectively transfer iron-sulfur clusters. Biochemistry 47, 8195-202(2008).

81. Lineweaver, H. & Burk, D. The determination of enzyme dissociation constants. J. Am. Chem. Soc. 56, 658-66(1934).

82. Ortiz, P.J. & Ochoa, S. Studies on polynucleotides synthesized by

polynucleotide phosphorylase. IV. P32-labeled ribonucleic acid. J. Biol. Chem. 234, 1208-12(1959).

83. Frazzon, J. & Dean, D.R. Biosynthesis of the nitrogenase iron-molybdenum-cofactor from Azotobacter vinelandii. Met Ions Biol Syst 39, 163-86(2002).

84. Setubal, J.C., dos Santos, P., Goldman, B.S., Ertesvåg, H., Espin, G., Rubio, L.M. et al. Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. J. Bacteriol. 191, 4534-45(2009).

85. Sorek, R., Cossart, P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet. 11, 9-16(2010).

# Anexos

Complete list of *Azotobacter vinelandii* known and predicted [Fe-S] proteins.

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin01510 | | 156 | 51 | Fer4 | + | [4Fe-4S] Ferredoxin-like |
| Avin03000 | | 1122 | 373 | Radical_SAM + HemN_C | - | O2-indep. coproporphyrinogen III oxidase |
| Avin03010 | rdgB | 597 | 198 | Ham1p_like | - | Non-canonical purine NTP pyrophosphatase |
| Avin03020 | | 420 | 139 | No | - | Conserved hypothetical protein |
| Avin03030 | metW | 600 | 199 | MetW | - | Methionine biosynthesis protein |
| Avin03040 | metX | 1140 | 379 | Abhydrolase_1 | - | Homoserine O-acetyltransferase |
| Avin03310 | ilvD | 1839 | 612 | ILVD_EDD | - | Dihydroxy-acid dehydratase |
| Avin03450 | | 1356 | 451 | CCG | - | Conserved hypothetical protein |
| Avin03460 | | 1704 | 567 | G_glu_transpept | - | Gamma-glutamyltranspeptidase precursor |
| Avin03470 | | 252 | 83 | Fer4 | - | 4Fe-4S Ferredoxin |
| Avin03800 | fdhD | 828 | 275 | FdhD-NarQ | + | Formate dehydrogenase family accessory protein |
| Avin03810 | fdhG | 3057 | 1018 | Molybdop_Fe4S4 + Molyb. + Molydop_binding | + | Formate dehydrogenase, α subunit |
| Avin03820 | fdhH | 894 | 297 | Fer4 | + | Formate dehydrogenase, β subunit |
| Avin03830 | fdhI | 648 | 215 | Cytochrom_B_N | + | Formate dehydrogenase, γ subunit |
| Avin03840 | fdhE | 915 | 304 | FdhE | + | Formate dehydrogenase accessory protein |
| Avin03850 | moaA | 1008 | 335 | Radical_SAM + Mob_synth_C | + | Molybdenum cofactor biosynthesis protein A |
| Avin04360 | | 1149 | 382 | Fer4 | + | Soluble hydrogenase, α or β chain |
| Avin04370 | hoxI | 459 | 152 | cNMP_binding | + | Cyclic nucleotide-binding, hydrogenase accessory |
| Avin04380 | hoxG | 831 | 276 | FAD_binding_6 + NAD_binding_1 | + | Soluble hydrogenase γ subunit |
| Avin04390 | hoxY | 771 | 256 | Oxidored_q6 | + | Soluble hydrogenase δ subunit |
| Avin04400 | hoxH | 1260 | 419 | NiFeSe_Hases | + | Soluble nickel-dependent hydrog., large subunit |
| Avin04410 | hoxW | 483 | 160 | HycI | + | Soluble hydrogenase processing peptidase |
| Avin04450 | | 564 | 187 | HupE_UreJ | - | Nickel transp. for carbon monoxide dehydrogenase |
| Avin04460 | | 222 | 73 | No | - | Hypothetical protein |
| Avin04470 | cooC | 801 | 266 | CbiA | - | Carbon monoxide dehydrogenae assesory protein |
| Avin04480 | | 1233 | 410 | Pyr_redox_2 | - | FAD-dep. pridine nucleotide-disulph. Oxidoreductase |
| Avin04490 | cooS | 1899 | 632 | Prismane | - | Carbon-monoxide dehydrogenase, catalytic subunit |
| Avin04500 | cooF | 558 | 185 | Fer4 | - | Fe-S binding protein |
| Avin04520 | yggX | 273 | 90 | DUF495 | - | Fe(II) trafficking YggX |
| Avin04530 | mutY | 1089 | 362 | HhH-GPD + HHH | - | A/G-specific adenine glycosylase |
| Avin04540 | | 2247 | 748 | AsmA | - | Conserved hypothetical protein |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin04680 | | 414 | 137 | Rhodanese | + | Rhodanese-like protein |
| Avin04690 | | 255 | 84 | Glutaredoxin | + | Dithiol-glutaredoxin protein |
| Avin04700 | secB | 492 | 163 | SecB | + | Protein-export chaperone |
| Avin05980 | bioB | 1056 | 351 | Radical_SAM + BATS | + | **Biotin synthase** |
| Avin05990 | bioF | 1176 | 392 | Aminotran_1_2 | + | 8-amino-7-oxononanoate synthase |
| Avin06000 | bioH | 723 | 240 | Abhydrolase_1 | + | Biotin biosynthesis protein |
| Avin06010 | bioC | 804 | 267 | Methyltransf_11 | + | Biotin synthesis protein |
| Avin06020 | | 693 | 230 | CDP-OH_P_transf | + | CDP-alcohol phosphatidyltransferase |
| Avin06030 | | 693 | 230 | Acyltransferase | + | Phospholipid/glycerol acyltransferase |
| Avin06040 | | 933 | 310 | CTP_transf_1 | + | Phosphatidate cytidylyltransferase |
| Avin06670 | | 870 | 289 | Thioredoxin | - | Thioredoxin protein |
| Avin07480 | | 1065 | 354 | Fer4 | - | **Fe-S binding protein** |
| Avin07700 | | 2274 | 757 | Radical_SAM | - | **Elongator protein 3, MiaB/NifB** |
| Avin08400 | rodA | 1149 | 382 | FTSW_RODA_SPOVE | + | Rod shape-determining protein |
| Avin08410 | mltB | 1029 | 342 | No | + | Lytic murein transglycosylase B |
| Avin08420 | | 978 | 325 | DPBB_1 + SPOR (2x) | + | Rare lipoprotein, RlpA family |
| Avin08430 | dacA | 1215 | 404 | Peptidase_S11 + PBP5_C | + | D-Alanyl-D-Alanine carboxypeptidase |
| Avin08440 | | 279 | 92 | DUF493 | + | Conserved hypothetical protein |
| Avin08450 | lipB | 657 | 218 | BPL_LipA_LipB | + | Lipoate-protein ligase B |
| Avin08460 | lipA | 978 | 325 | Radical_SAM | + | **Lipoic acid synthetase** |
| Avin08570 | benR | 960 | 319 | AraC_binding + HTH_AraC (2x) | + | Transcriptional activator, BenR (AraC family) |
| Avin08590 | benA | 1362 | 453 | Rieske + Ring_hydroxyl_A | + | Benzoate-1,2-dioxygenase oxygenase, α subunit |
| Avin08600 | benB | 489 | 162 | Ring_hydroxyl_B | + | Benzoate-1,2-dioxygenase oxygenase, β subunit |
| Avin08610 | benC | 1011 | 336 | Fer2 + FAD_binding_6 + NAD_binding_1 | + | Benzoate 1,2-dioxygenase reductase |
| Avin08620 | benD | 774 | 257 | adh_short | + | cis-1,2-dihydroxycyclohexa-3,4-diene carboxylate dehydrog. |
| Avin08630 | | 1332 | 443 | MFS_1 | + | Benzoate transporter |
| Avin08660 | | 1272 | 423 | OprD | + | Benzoate-specific outer membrane porin |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin08670 | xylJ | 786 | 261 | Hyratase_decarb | + | 4-oxo-4-pentenoate hydratase |
| Avin08680 | xylI | 789 | 262 | Hyratase_decarb | + | 4-oxalocrotonate decarboxylase |
| Avin08690 | xylH | 183 | 60 | Tautomerase | + | 4-oxalocrotonate tautomerase |
| Avin08700 | xylQ | 957 | 318 | Semialdhyde_dh | + | Acetaldehyde dehydrogenase |
| Avin08710 | xylT | 315 | 104 | Fer2 | + | Ferredoxin |
| Avin08720 | xylE | 924 | 307 | Glyoxalase + Glyoxalase | + | Catechol-2,3-dioxygenase |
| Avin08730 | xylG | 1461 | 486 | Aldedh | + | 2-hydroxymuconic semi-aldehyde dehydrogenase |
| Avin08740 | xylK | 1050 | 349 | HMGL-like + DmpG_comm | + | 4-hydroxy-2-oxovalerate aldolase |
| Avin08750 | | 450 | 149 | DUF336 | + | Conserved hypothetical protein |
| Avin08760 | | 921 | 306 | HTH_1 + LysR_substrate | + | Transcriptional regulator, LysR family |
| Avin08810 | dmpP | 1062 | 353 | Fer2 + FAD_binding_6 + NAD_binding_1 | - | Phenol hydroxylase, ferredoxin subunit |
| Avin08820 | dmpO | 360 | 119 | Phenol_monoox | - | Phenol hydroxylase subunit P4 |
| Avin08830 | dmpN | 1536 | 511 | Phenol_Hydrox + YHS | - | Phenol hydroxylase subunit P3 |
| Avin08840 | dmpM | 270 | 89 | MmoB_DmpM | - | Phenol hydroxylase subunit P2 |
| Avin08850 | dmpL | 999 | 332 | Phenol_Hydrox | - | Phenol hydroxylase subunit P1 |
| Avin08860 | dmpK | 258 | 85 | Phenol_hyd_sub | - | Phenol hydroxylase subunit |
| Avin09030 | lnt | 1551 | 516 | CN_hydrolase | - | Apolipoprotein N-acyltransferase |
| Avin09040 | | 843 | 280 | CBS + CorC_HlyC | - | CBS domain-containing transporter |
| Avin09050 | | 489 | 162 | UPF0054 | - | Conserved hypothetical protein |
| Avin09060 | | 1011 | 336 | PhoH | - | PhoH-like protein |
| Avin09070 | | 129 | 42 | No | - | Hypothetical protein |
| Avin09080 | miaB | 1329 | 442 | UPF0004 + Radical_SAM + TRAM | - | **tRNA-i(6)A37 thiotransferase enzyme** |
| Avin09800 | leuD2 | 609 | 202 | Aconitase_C | - | 3-isopropylmalate dehydratase small subunit |
| Avin09810 | leuC2 | 1422 | 473 | Aconitase | - | **3-isopropylmalate dehydratase large subunit** |
| Avin10510 | | 219 | 72 | Fer4 | + | **[4Fe-4S] Ferredoxin-like** |
| Avin10520 | fixA | 846 | 281 | ETF | + | Electron transfer flavoprotein β subunit |
| Avin10530 | fixB | 1083 | 360 | ETF + ETF_alpha | + | Electron transfer flavoprotein, α subunit |
| Avin10540 | fixC | 1299 | 432 | DAO | + | Electron-transferring flavoprotein dehydrogenase |
| Avin10550 | fixX | 285 | 94 | No (others: [4Fe-4S ferredoxin]) | + | **FixX Ferredoxin-like protein** |
| Avin10560 | | 459 | 152 | Rubrerythrin | + | Ferritin-like protein |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin11760 | mviN | 1545 | 514 | MVIN | + | Integral membrane protein |
| Avin11770 | ribF | 825 | 274 | FAD_syn + Flavokinase | + | Riboflavin biosynthesis protein |
| Avin11780 | ileS | 2832 | 943 | TRNA-synt_1 + Anticodon_1 + zf-FPG_IleRS | + | Isoleucyl-tRNA synthetase |
| Avin11790 | lspA | 504 | 167 | Peptidase_A8 | + | Lipoprotein signal peptidase |
| Avin11800 | | 438 | 145 | FKBP_C | + | Peptidylprolyl isomerase, FKBP-type |
| Avin11810 | ispH | 948 | 315 | LYTB | + | 4-hydroxy-3-methylbut-2-enyl diphosphate red. |
| Avin12000 | ndh | 1299 | 432 | Pyr_redox_2 | - | Uncoupled NADH:quinone oxidoreductase |
| Avin13060 | petA | 405 | 134 | Rieske | + | Ubiquinol-cytochrome c reductase, Fe-S subunit |
| Avin13070 | petB | 1212 | 403 | Cytochrom_B_N + Cytochrom_B_C | + | Cytochrome b/b6 |
| Avin13080 | petC | 780 | 259 | Cytochrom_C1 | + | Ubiquinol-cytochrome c reductase, cytochrome c1 |
| Avin13090 | sspA | 618 | 205 | GST_N + GST_C | + | Stringent starvation protein A |
| Avin13100 | sspB | 408 | 135 | SspB | + | Stringent starvation protein B |
| Avin14040 | | 327 | 108 | Glutaredoxin | + | Glutaredoxin4 protein |
| Avin14590 | nqrA | 1338 | 445 | NQRA | + | NADH:ubiquinone oxidoreductase, Na(+)-translocating, subunit A |
| Avin14600 | nqrB | 1212 | 403 | NQR2_RnfD_RnfE | + | NADH:ubiquinone oxidoreductase, Na(+)-translocating, subunit B |
| Avin14610 | nqrC | 789 | 262 | FMN_bind | + | NADH:ubiquinone oxidoreductase, Na(+)-translocating, subunit C |
| Avin14620 | nqrD | 669 | 222 | Rnf-Nqr | + | NADH:ubiquinone oxidoreductase, Na(+)-translocating, subunit D |
| Avin14630 | nqrE | 609 | 202 | Rnf-Nqr | + | NADH:ubiquinone oxidoreductase, Na(+)-translocating, subunit E |
| Avin14640 | nqrF | 1224 | 407 | Fer2 + FAD_binding_6 + NAD_binding_1 | + | NADH:ubiquinone oxidoreductase, Na(+)-translocating, subunitF |
| Avin14650 | apbE | 1014 | 337 | ApbE | + | Thiamine biosynthesis lipoprotein |
| Avin14660 | | 234 | 77 | DUF539 | + | Conserved hypothetical protein |
| Avin14670 | sthA | 1395 | 464 | Pyr_redox_2 + Pyr_redox_dim | + | Pyridine nucleotide transhydrogenase |
| Avin15680 | edd-1 | 1827 | 608 | ILVD_EDD | + | 6-phosphogluconate dehydratase |
| Avin15690 | glk-1 | 969 | 322 | Gloucokinase | + | Glucokinase |
| Avin16190 | | 1413 | 470 | Pyr_redox | + | Conserved hypothetical protein |
| Avin16810 | | 999 | 332 | DHO_dh | - | Dihydroorotate dehydrogenase |
| Avin18430 | | 1140 | 379 | Oxidored_FMN | - | NADH:flavin oxidoreductase/NADH oxidase family |
| Avin18440 | | 1500 | 499 | Aldedh | - | Aldehyde dehydrogenase family protein |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin19200 | | 180 | 59 | No | - | Conserved hypothetical protein |
| Avin19210 | | 639 | 212 | HhH-GPD + HHH | - | Endonuclease III/Nth |
| Avin19220 | rnfE | 714 | 237 | Rnf-Nqr | - | Electron transport complex, subunit E |
| Avin19230 | rnfG | 627 | 208 | FMN_bind | - | Electron transport complex, subunit G |
| Avin19240 | rnfD | 1032 | 343 | NQR2_RnfD_RnfE | - | Electron transport complex, subunit D |
| Avin19250 | rnfC | 2067 | 688 | Complex1_51K + Fer4 + Fer4 | - | Electron transport complex, subunit C |
| Avin19260 | rnfB | 564 | 187 | FeS + Fer4 + Fer4 | - | Electron transport complex, subunit B |
| Avin19270 | rnfA | 585 | 194 | Rnf-Nqr | - | Electron transport complex, subunit A |
| Avin19910 | cydR | 735 | 244 | cNMP_binding + Crp | - | Fnr-like negative transc. regulator of CydAB |
| Avin19920 | hemN | 1377 | 458 | Radical_SAM + HemN_C | - | O2-indep. coproporphyrinogen III oxidase |
| Avin19930 | | 870 | 289 | No | - | Conserved hypothetical protein |
| Avin19940 | ccoS | 222 | 73 | CcoS | - | cbb3-type cytochrome oxidase maturation protein |
| Avin19950 | ccoI | 2403 | 800 | HMA + E1-E2_ATPase + Hydrolase | - | Copper-translocating P-type ATPase |
| Avin19960 | ccoH | 504 | 167 | No | - | Cytochrome c oxidase accessory protein |
| Avin19970 | ccoG | 1413 | 470 | Fer4 | - | Cytochrome c oxidase accessory protein |
| Avin19980 | ccoP | 921 | 306 | Cytochrom_C (2x) | - | Cytochrome c oxidase, cbb3-type, subunit III |
| Avin19990 | ccoQ | 186 | 61 | No | - | Cytochrome c oxidase, cbb3-type, subunit IV |
| Avin20000 | ccoO | 609 | 202 | FixO | - | Cytochrome c oxidase, cbb3-type, subunit II |
| Avin20010 | ccoN | 1443 | 480 | COX1 | - | Cytochrome c oxidase, cbb3-type, subunit I |
| Avin20040 | acnA | 2688 | 895 | Aconitase + Aconitase_C | - | Aconitate hydratase 1 |
| Avin20620 | | 1107 | 368 | Glycos_transf_2 | - | Glycosyl transferase, family 2 |
| Avin20630 | | 1134 | 377 | Radical_SAM | - | MoaA, NifB, PqqE family protein |
| Avin20650 | | 1161 | 386 | Glycos_transf_2 | + | Glycosyl transferase, family 2 |
| Avin20660 | | 1413 | 470 | Radical_SAM | + | MoaA, NifB, PqqE family protein |
| Avin20670 | | 708 | 235 | YdjC | + | Conserved hypothetical protein |
| Avin20680 | | 1002 | 333 | No | + | Conserved hypothetical protein |
| Avin20820 | | 1356 | 451 | Lipase_GDSL | + | Lipolytic enzyme, G-D-S-L domain protein |
| Avin20830 | | 1140 | 379 | Oxidored_FMN | + | NADH:flavin oxidoreductase/NADH oxidase family |
| Avin20840 | | 1212 | 403 | MFS_1 | + | Multidrug/chloramphenicol efflux transporter MFS1 |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin21880 | | 687 | 228 | AhpC-TSA | - | Peroxiredoxin, thioredoxin-like fold protein |
| Avin21890 | | 1332 | 443 | Bac_luciferase | - | Monooxygenase, NtaA/SnaA/SoxA family |
| Avin21900 | nemA | 1071 | 356 | Oxidored_FMN | - | **NADH:flavin oxidoreductase NADH oxidase family** |
| Avin21910 | | 990 | 329 | NMT1 | - | ABC transp., subs.-bind., aliphatic sulphonates |
| Avin21920 | asfB | 393 | 130 | Fer4 | - | **[4Fe-4S] Ferredoxin-like** |
| Avin21930 | asfA | 1638 | 545 | GIDA | - | Oxidoreductase flavoprotein |
| Avin22270 | xdhC | 843 | 280 | XdhC_CoxI | - | Xanthine dehydrogenase accessory protein |
| Avin22280 | xdhB | 2394 | 797 | Ald_Xan_dh_C + Ald_Xan_dh_C2 | - | Xanthine dehydrogenase molybdopterin bind. subunit |
| Avin22290 | xdhA | 1491 | 496 | Fer2 + Fer2_2 + FAD_binding_5 + CO_deh_flav_C | - | Xanthine dehydrogenase small subunit |
| Avin23200 | gntR | 723 | 240 | GntR + FCD | + | Transcriptional regulator |
| Avin23210 | prpB | 888 | 295 | ICL | + | Methylisocitrate lyase |
| Avin23220 | prpC | 1164 | 387 | Citrate_synt | + | 2-methylcitrate synthase |
| Avin23230 | acnD | 2604 | 867 | Aconitase + Aconitase_C | + | **2-methylisocitrate dehydratase** |
| Avin23250 | prpF | 1194 | 397 | DUF453 | + | acnD accessory protein |
| Avin23340 | nasH | 738 | 245 | TP_methylase | - | Uroporphyrin-III C-methyltransferase, NasH |
| Avin23350 | nasB | 2706 | 901 | Molybdop_Fe4S4 + Molybdopterin + Molydop_binding + Fer2_BFD | - | **Assimilatory nitrate reductase** |
| Avin23360 | | 327 | 108 | Rieske | - | Assimilatory nitrite reductase, small subunit |
| Avin23370 | nasA | 2451 | 816 | Pyr_redox_2 + Fer2_BFD (2) + NIR_SIR_ferr + NIR_SIR | - | Assimilatory nitrite reductase, large subunit |
| Avin23470 | acnB | 2610 | 869 | Aconitase_2_N + Aconitase | - | **Aconitate hydratase 2** |
| Avin25670 | | 1614 | 537 | Glu_synthase | + | **Glutamate synthase** |
| Avin25820 | | 603 | 200 | No | - | Conserved hypothetical protein |
| Avin25830 | | 990 | 329 | XdhC_CoxI | - | Conserved hypothetical protein |
| Avin25840 | | 2313 | 770 | Ald_Xan_dh_C + Ald_Xan_dh_C2 | - | Oxidoreductase molybdopterin-binding subunit |
| Avin25850 | | 471 | 156 | Fer2 + Fer2_2 | - | Ferredoxin |
| Avin26040 | sdaA | 1377 | 458 | SDH_beta + SDH_alpha | + | **L-serine dehydratase** |
| Avin26050 | gcvT3 | 1125 | 374 | GCV_T | + | Glycine cleavage system T protein |
| Avin26060 | | 459 | 152 | Fer2 + Fer2_2 | + | Oxidoreductase Fe-S binding subunit |
| Avin26070 | | 2241 | 746 | Ald_Xan_dh_C2 | + | Oxidoreductase |
| Avin26320 | | 1566 | 521 | Pyr_redox_2 | + | **FAD-dep. Pyridine nucleotide-disulph. Oxidored.** |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin27280 | edd-2 | 1827 | 608 | ILVD_EDD | - | 6-phosphogluconate dehydratase |
| Avin27520 | cysI | 1659 | 552 | NIR_SIR_ferr (2x) + NIR_SIR (2x) | + | Sulfite reductase, beta subunit (hemoprotein) |
| Avin27530 | | 498 | 165 | DUF934 | + | Conserved hypothetical protein |
| Avin28440 | nuoA | 414 | 137 | Oxidored_q4 | + | NADH-ubiquinone oxidoreductase, chain A |
| Avin28450 | nuoB | 675 | 224 | Oxidored_q6 | + | NADH-ubiquinone oxidoreductase, chain B |
| Avin28460 | nuocd | 1782 | 593 | Complex1_30kDa + Complex1_49kDa | + | NADH-ubiquinone oxidoreductase, chain CD |
| Avin28470 | nuoE | 495 | 164 | Complex1_24kDa | + | NADH-ubiquinone oxidoreductase, chain E |
| Avin28480 | nuoF | 1350 | 449 | Complex1_51K | + | NADH-ubiquinone oxidoreductase, chain F |
| Avin28490 | nuoG | 2712 | 903 | Fer2 + Molybdop_Fe4S4 + Molybdopterin | + | NADH-ubiquinone oxidoreductase, chain G |
| Avin28500 | nuoH | 990 | 329 | NADHdh | + | NADH-ubiquinone oxidoreductase, chain H |
| Avin28510 | nuoI | 549 | 182 | Fer4 + Fer4 | + | NADH-quinone oxidoreductase, chain I |
| Avin28520 | nuoJ | 501 | 166 | Oxidored_q3 | + | NADH-ubiquinone/plastoquinone oxidoreductase, chain J |
| Avin28530 | nuoK | 309 | 102 | Oxidored_q2 | + | NADH-ubiquinone oxidoreductase, chain K |
| Avin28540 | nuoL | 1848 | 615 | Oxidored_q1_N + Oxidored_q1 | + | NADH-plastoquinone oxidoreductase, chain L |
| Avin28550 | nuoM | 1530 | 509 | Oxidored_q1 | + | Proton-translocating NADH-quinone oxidoreductase, chain M |
| Avin28560 | nuoN | 1473 | 490 | Oxidored_q1 | + | Proton-translocating NADH-quinone oxidoreductase, chain N |
| Avin29350 | | 1035 | 344 | Radical_SAM | + | Fe-S cluster redox enzyme |
| Avin29780 | sdhB | 702 | 233 | Fer2 | - | Succinate dehydrogenase Fe-S subunit |
| Avin29790 | sdhA | 1320 | 439 | FAD_binding_2 + Succ_DH_flav_C | - | Succinate dehydrogenase flavoprotein subunit |
| Avin29800 | sdhD | 369 | 122 | No | - | Succinate dehydrogenase cytochrome b556 small subunit |
| Avin29810 | sdhC | 234 | 77 | Sdh_cyt | - | Succinate dehydrogenase cytochrome b556 large subunit |
| Avin30330 | moaA | 996 | 331 | Radical_SAM + Mob_synth_C | - | Molybdenum cofactor biosynthesis protein A |
| Avin30560 | lapQ | 366 | 121 | Fer2 | - | Ferredoxin |
| Avin30570 | lapG | 1053 | 350 | HMGL-like + DmpG_comm | - | 4-hydroxy-2-oxovalerate aldolase |
| Avin30580 | lapF | 915 | 304 | Semialdhyde_dh | - | Acetaldehyde dehydrogenase (acetylating) |
| Avin30590 | lapI | 210 | 69 | Tautomerase | - | 4-oxalocrotonate isomerase |
| Avin30600 | lapH | 804 | 267 | Hyratase_decarb | - | 4-oxalocrotonate decarboxylase |
| Avin30610 | lapE | 798 | 265 | Hyratase_decarb | - | 2-Hydroxypent-2,4-dienoate hydratase |
| Avin30620 | lapC | 1461 | 486 | Aldedh | - | 2-hydroxymuconic semialdehyde dehydrogenase |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin30710 | lapP | 1062 | 353 | Fer2 + FAD_binding_6 + NAD_binding_1 | - | **Multi-component phenol hydoxylase, reductase subunit** |
| Avin30720 | lapO | 357 | 118 | Phenol_monoox | - | Multi-component phenol hydoxylase, γ subunit |
| Avin30730 | lapN | 1518 | 505 | Phenol_Hydrox + YHS | - | Multi-component phenol hydoxylase, α subunit |
| Avin30740 | lapM | 270 | 89 | MmoB_DmpM | - | Multi-component phenol hydoxylase, activator subunit |
| Avin30750 | lapL | 993 | 330 | Phenol_Hydrox | - | Multi-component phenol hydoxylase, β subunit |
| Avin30760 | lapK | 267 | 88 | Phenol_hyd_sub | - | Multi-component phenol hydoxylase, assembly subunit |
| Avin30780 | lapB | 930 | 309 | Glyoxalase + Glyoxalase | - | Catechol 2,3-dioxygenase |
| Avin31190 | ttdB | 609 | 202 | Fumerase_C | - | **L(+)-tartrate dehydratase beta subunit protein** |
| Avin31200 | ttdA | 861 | 286 | Fumerase | - | **L-tartrate dehydratase alpha subunit** |
| Avin31210 | dctA-3 | 1326 | 441 | SDF | - | C4-dicarboxylate transport protein |
| Avin34240 | | 1011 | 336 | Semialdhyde_dh + Semialdhyde_dhC | - | Aspartate-semialdehyde dehydrogenase |
| Avin34250 | asd | 1053 | 350 | Semialdhyde_dh + Semialdhyde_dhC | - | Aspartate-semialdehyde dehydrogenase |
| Avin34260 | leuB | 1083 | 360 | Iso_dh | - | 3-isopropylmalate dehydrogenase |
| Avin34270 | leuD | 648 | 215 | Aconitase_C | - | |
| Avin34280 | leuC | 1425 | 474 | Aconitase | - | **3-isopropylmalate dehydratase large subunit** |
| Avin35390 | nadA | 1059 | 352 | NadA | - | **Quinolinate synthetase complex, A subunit** |
| Avin35520 | | 864 | 287 | Radical_SAM + Mob_synth_C | - | **Molybdenum cofactor biosynthesis protein** |
| Avin35540 | | 447 | 148 | Fer2 + Fer2_2 | + | Fe-S binding oxidoreductase |
| Avin35550 | | 2241 | 746 | Ald_Xan_dh_C2 + Ald_Xan_dh_C2 | + | Oxidoreductase, xanthine dehydrogenase-like |
| Avin35560 | | 108 | 35 | No | + | Hypothetical protein |
| Avin35570 | | 318 | 105 | Cupin_2 | + | Conserved hypothetical protein |
| Avin36600 | | 675 | 224 | ExsB | - | ExsB protein |
| Avin36610 | | 582 | 193 | Radical_SAM | - | **Radical SAM protein** |
| Avin37060 | ivlM | 2250 | 749 | RelA_SpoT + TGS + ACT | - | RelA/SpoT protein |
| Avin37070 | rumA | 1365 | 454 | TRAM + Met_10 | - | **23S rRNA (uracil-5-)-methyltransferase RumA** |
| Avin37080 | cysM | 900 | 299 | PALP | - | Cysteine synthase K/M:Cysteine synthase M |
| Avin37650 | rimO | 1323 | 440 | UPF0004 + Radical_SAM + TRAM | + | **Ribosomal protein S12 methylthiotransferase** |
| Avin37790 | | 303 | 100 | HTH_5 | + | Bacterial regulatory protein, ArsR family |
| Avin37800 | | 1050 | 349 | Oxidored_FMN | + | **NADH:flavin xenobiotic reductase** |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin38440 | fumB/fumA | 1512 | 503 | Fumerase + Fumerase_C | + | **Fumarate hydratase, class I** |
| Avin39420 | | 726 | 241 | DUF533 | - | Conserved hypothetical protein |
| Avin39430 | | 1107 | 368 | Oxidored_FMN | - | **NADH:flavin oxidoreductase/NADH oxidase family** |
| Avin39700 | | 369 | 122 | Fer2 | - | Ferredoxin |
| Avin40260 | | 1155 | 384 | PQQ | - | YfgL-like phosphoquinolipoprotein kinase |
| Avin40270 | | 645 | 214 | TPR_2 | - | YfgM-like protein |
| Avin40280 | hisS | 1287 | 428 | tRNA-synt_2b + HGTP_anticodon | - | Histidyl-tRNA synthetase |
| Avin40290 | ispG | 1113 | 370 | GcpE | - | **4-hydroxy-3-methylbut-2-en-1-yl diphosphate synt.** |
| Avin40300 | | 1029 | 342 | No | - | Conserved hypothetical protein |
| Avin40310 | pilF | 759 | 252 | TPR_1 (3x) | - | Type IV pilus biogenesis protein |
| Avin40320 | | 1146 | 381 | Radical_SAM | - | **Conserved hypothetical protein** |
| Avin40330 | ndk | 432 | 143 | NDK | - | Nucleoside-diphosphate kinase |
| Avin41370 | hemH | 1017 | 338 | Ferrochelatase | - | Ferrochelatase |
| Avin41380 | | 954 | 317 | Epimerase | - | Sugar nucleotide epimerase |
| Avin41610 | pqqF | 2532 | 843 | Peptidase_M16 + Peptidase_M16_C | - | Coenzyme PQQ biosynthesis protein F |
| Avin41620 | | 1956 | 651 | Peptidase_S9 | - | Peptidase |
| Avin41640 | pqqE | 1152 | 383 | Radical_SAM | - | **Coenzyme PQQ biosynthesis protein E** |
| Avin41650 | pqqD | 279 | 92 | PqqD | - | Coenzyme PQQ biosynthesis protein D |
| Avin41660 | pqqC | 753 | 250 | TENA_THI-4 | - | Coenzyme PQQ biosynthesis protein C |
| Avin41670 | pqqB | 909 | 302 | No | - | Coenzyme PQQ biosynthesis protein B |
| Avin41820 | | 1302 | 433 | No | - | **Lysine 2,3-aminomutase** |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin42100 | | 1002 | 333 | Fer2 + FAD_binding_6 + NAD_binding_1 | - | Ferredoxin:oxidoreductase FAD/NAD(P)-binding |
| Avin42110 | | 1224 | 407 | MFS_1 | - | Major facilitator superfamily protein |
| Avin42120 | | 1023 | 340 | HMGL-like + DmpG_comm | - | 4-hydroxy-2-oxovalerate aldolase |
| Avin42130 | | 939 | 312 | Semialdhyde_dh | - | Acetaldehyde dehydrogenase |
| Avin42140 | | 1860 | 619 | AMP_binding | - | AMP-dependent synthetase and ligase |
| Avin42150 | | 786 | 261 | Hyratase_decarb | - | 2-hydroxypent-2,4-dienoate hydratase |
| Avin42160 | | 978 | 325 | NPD | - | Dioxygenase related to 2-nitropropane dioxygenase |
| Avin42170 | | 813 | 270 | adh_short | - | Dihydrodiol dehydrogenase |
| Avin42180 | | 522 | 173 | Ring_hydroxyl_B | - | Aromatic-Ring-hydroxylating dioxygenase, β subunit |
| Avin42190 | | 1347 | 448 | Rieske + Ring_hydroxyl_A | - | Aromatic ring hydroxylating dioxygenase, α subunit |
| Avin42200 | | 369 | 122 | Rieske | - | Rieske type ferredoxin |
| Avin42210 | | 855 | 284 | LigB | - | Extradiol ring-cleavage dioxygenase, class III, subunit B |
| Avin42220 | | 297 | 98 | No | - | Extradiol ring-cleavage dioxygenase, class III, subunit A |
| Avin42230 | | 816 | 271 | Abhydrolase_1 | - | 2-hydroxymuconic semialdehyde hydrolase |
| Avin42610 | | 318 | 105 | Rieske | + | Conserved hypothetical protein |
| Avin43070 | lctP | 1689 | 562 | Lactate_perm | + | L-lactate permease |
| Avin43080 | | 825 | 274 | CCG (2x) | + | **Conserved hypothetical protein** |
| Avin43090 | | 1449 | 482 | Fer4 | + | **Fe-S cluster binding protein-like** |
| Avin43100 | | 672 | 223 | DUF162 | + | Conserved hypothetical protein |
| Avin43110 | | 2817 | 938 | FAD_binding_4 + FAD-oxidase_C + Fer4 + CCG | + | **Fe-S/FAD domain protein** |
| Avin43290 | glcB | 2178 | 725 | Malate_synthase | - | Malate synthase G |
| Avin43300 | glcG | 366 | 121 | DUF336 | - | Glyoxylate glcG protein |
| Avin43310 | glcF | 1218 | 405 | Fer4 (2x) + CCG (2x) | - | **Glycolate oxidase Fe-S subunit, glcF** |
| Avin43320 | glcE | 1056 | 351 | FAD_binding_4 + FAD-oxidase_C | - | Glycolate oxidase FAD binding subunit |
| Avin43330 | glcD | 1500 | 499 | FAD_binding_4 + FAD-oxidase_C | - | Glycolate oxidase, subunit GlcD |
| Avin45190 | hemE | 1068 | 355 | URO-D | - | Uroporphyrinogen decarboxylase |
| Avin45200 | gltD | 1422 | 473 | Pyr_redox_2 | - | **Glutamate synthase small subunit 2** |
| Avin45210 | gltB | 4443 | 1480 | Glu_synthase | - | **Glutamate synthase large subunit protein** |
| Avin47390 | | 969 | 322 | Fer2 + FAD_binding_6 + NAD_binding_1 | - | Oxidoreductase protein |
| Avin47400 | ubiD | 1467 | 488 | UbiD | - | 3-octaprenyl-4-hydroxybenzoate carboxy-lyase |
| Avin47420 | trxA | 327 | 108 | Thioredoxin | - | Thioredoxin 1, trx1 |

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin49070 | | 1095 | 364 | Pyr_redox_2 | - | **Pyridine nucleotide-disulphide oxidoreductase** |
| Avin49080 | | 1452 | 483 | Aldedh | - | Aldehyde dehydrogenase |
| Avin49250 | | 1737 | 578 | ILVD_EDD | - | **L-arabonate dehydratase** |
| Avin50900 | | 396 | 131 | Nitro_FeMo-Co | - | Nitrogen fixation-related protein |
| Avin50910 | nafY | 732 | 243 | Nitro_FeMo-Co | - | Nitrogen fixation-related protein, γ subunit |
| Avin50920 | rnfH | 261 | 86 | UPF0125 | - | Electron transp. Comp. RnfABCDGE type, H subunit |
| Avin50930 | rnfE1 | 717 | 238 | Rnf-Nqr | - | Electron transp. Comp. RnfABCDGE type, E subunit |
| Avin50940 | rnfG1 | 690 | 229 | FMN_bind | - | Electron transp. Comp. RnfABCDGE type, G subunit |
| Avin50950 | rnfD1 | 1101 | 366 | NQR2_RnfD_RnfE | - | Electron transp. Comp. RnfABCDGE type, D subunit |
| Avin50960 | rnfC1 | 1491 | 496 | Complex1_51K + Fer4 + Fer4 | - | **Electron transp. Comp. RnfABCDGE type, C subunit** |
| Avin50970 | rfnB1 | 525 | 174 | FeS + Fer4 + Fer4 | - | **Electron transp. Comp. RnfABCDGE type, B subuni** |
| Avin50980 | rnfA1 | 573 | 190 | Rnf-Nqr | - | Electron transp. Comp. RnfABCDGE type, A subunit |

**Nif system**

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin01360 | hyp (orf13) | 564 | 167 | No | - | Conserved hypothetical protein |
| Avin01370 | hyp (orf12) | 864 | 287 | No | - | Conserved hypothetical protein |
| Avin01380 | nifH | 873 | 290 | Fer4_NifH | + | **Nitrogenase iron protein** |
| Avin01390 | nifD | 1479 | 492 | Oxidored_nitro | + | Nitrogenase I, α-chain[a,b] |
| Avin01400 | nifK | 1572 | 523 | Oxidored_nitro | + | Nitrogenase I, β-chain[a,b] |
| Avin01410 | nifT | 219 | 72 | NifT | + | Nitrogen fixation protein |
| Avin01420 | nifY | 729 | 242 | Nitro_FeMo-Co | + | Nitrogenase FeMo biosynthesis[c] |
| Avin01430 | hyp (orf1) | 270 | 89 | No | + | Conserved hypothetical |
| Avin01440 | lrv (orf2) | 735 | 244 | LRV_FeS + LRV (4) | + | **Leu rich variant repeat [4Fe-4S]** |
| Avin01450 | nifE | 1428 | 475 | Oxidored_nitro | + | Nitrogenase FeMo biosynthesis[d] |
| Avin01470 | nifN | 1377 | 458 | Oxidored_nitro | + | Nitrogenase FeMo biosynthesis[d] |
| Avin01480 | nifX | 477 | 158 | Nitro_FeMo-Co | + | Nitrogenase FeMo biosynthesis[c] |
| Avin01490 | hyp (orf3) | 483 | 160 | DUF269 | + | Conserved hypothetical protein |
| Avin01500 | hyp (orf4) | 210 | 69 | DUF683 | + | Conserved hypothetical protein |
| Avin01510 | hyp | 156 | 51 | Fer4 | + | **[4Fe-4S] ferredoxin-like** |
| Avin01520 | fesI | 324 | 107 | No | + | [2Fe-2S] ferredoxin |
| Avin01530 | hyp | 501 | 166 | No | + | Conserved hypothetical protein |
| Avin01540 | hyp | 378 | 125 | No | + | Conserved hypothetical protein |
| Avin01550 | hyp | 750 | 249 | DUF364 | + | Conserved hypothetical protein |
| Avin01560 | hyp | 408 | 135 | No | + | Conserved hypothetical protein |
| Avin01570 | hyp (orf11) | 219 | 72 | Nif11 | - | Conserved hypothetical protein |
| Avin01580 | hyp (orf10) | 1077 | 358 | ABC_tran + TOBE | - | ABC transporter, ATP-binding |
| Avin01600 | hyp (orf5) | 762 | 253 | No | + | Conserved hypothetical protein |
| Avin01610 | iscAnif | 324 | 107 | Fe-S_biosyn | + | Fe-S cluster assembly[e] |
| Avin01620 | nifU | 939 | 312 | NifU_N + Fer2_BFD + NifU | + | **Fe-S cluster scaffold** |
| Avin01630 | nifS | 1209 | 402 | Aminotran_5 | + | Cysteine desulfurase |
| Avin01640 | nifV | 1155 | 384 | HMGL-like | + | Homocitrate synthase |
| Avin01650 | cysE1nif | 798 | 265 | Hexapep | + | Serine O-acetyltransferase |
| Avin01660 | hyp (orf8nif) | 540 | 179 | No | + | Conserved hypothetical protein |
| Avin01670 | nifW | 348 | 115 | NifW | + | Cis-trans peptidyl prolyl isom. |
| Avin01680 | nifZ | 480 | 159 | NifZ | + | Nitrogen fixation protein NifZ[f] |
| Avin01690 | nifM | 882 | 293 | Rotamase | + | Cis-trans peptidyl prolyl isom. |
| Avin01700 | clpXnif | 1323 | 440 | zf-C4_ClpX + AAA_2 | + | Nitrogen fixation protein orf9, ClpX |
| Avin01710 | nifF | 543 | 180 | Flavodoxin_1 | + | Flavodoxin |

[a][8Fe-7S] P-cluster binding [b]FeMoco cluster binding [c]FeMoco transport [d]FeMoco biosynthesis [e][2Fe-2S] and [4Fe-4S] clusters transfer [f][8Fe-7S] P-cluster formation

**Nif system**

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin50990 | nifL | 1560 | 519 | PAS + PAS + HATPase_c | + | Regulatory protein |
| Avin51000 | nifA | 1569 | 522 | GAF + Sigma54_activat + HTH_8 | + | Sigma54 transc. Activator |
| Avin51010 | nifB | 1512 | 503 | Radical_SAM + Nitro_FeMo-Co | + | Nitrogenase FeMoco biosynthesis[d] |
| Avin51020 | Avin51020 | 279 | 92 | Fer4 | + | Ferredoxin |
| Avin51030 | nifO | 444 | 147 | ArsC | + | Nitrogenase-associated |
| Avin51050 | rhdN | 363 | 120 | Rhodanese | + | Rhodanese/sulfurtransferase |
| Avin51060 | | 330 | 109 | Glutaredoxin | + | Glutaredoxin-related |

**Vnf system**

| | | | | | | |
|---|---|---|---|---|---|---|
| Avin02590 | vnfK | 1428 | 475 | Oxidored_nitro | - | V nitrogenase, β subunit |
| Avin02600 | vnfG | 342 | 113 | AnfG_VnfG | - | V nitrogenase, δ subunit |
| Avin02610 | vnfD | 1425 | 474 | Oxidored_nitro | - | V nitrogenase, α subunit |
| Avin02620 | | 177 | 58 | No | + | Hypothetical protein |
| Avin02630 | | 84 | 27 | No | - | Hypothetical protein |
| Avin02640 | | 219 | 72 | No | + | Hypothetical protein |
| Avin02650 | vnfF | 198 | 65 | Fer4 | - | V nitrogenase ferredoxin |
| Avin02660 | vnfH | 873 | 290 | Fer4_NifH | - | V nitrogenase iron protein |
| Avin02740 | vnfX | 549 | 182 | Nitro_FeMo-Co | - | V nitrogenase biosynthesis protein |
| Avin02750 | vnfN | 1383 | 460 | Oxidored_nitro | - | V nitrogenase biosynthesis protein |
| Avin02770 | vnfE | 1410 | 469 | Oxidored_nitro | - | V nitrogenase biosynthesis protein |
| Avin02780 | vnfA | 1569 | 522 | GAF + Sigma54_activat + HTH_8 | - | Sigma54 transc. Activator[g] |
| Avin02790 | vnfU | 291 | 96 | NifU | - | NifU C-terminal |

**Anf system**

| | | | | | | |
|---|---|---|---|---|---|---|
| Avin48950 | anfR | 552 | 183 | No | - | Fe-only nitrogenase, accessory |
| Avin48960 | anfO | 738 | 245 | No | - | Fe-only nitrogenase, accessory |
| Avin48970 | anfK | 1389 | 462 | Oxidored_nitro | - | Fe-only nitrogenase, β-subunit |
| Avin48980 | anfG | 399 | 132 | AnfG_VnfG | - | Fe-only nitrogenase, δ-subunit |
| Avin48990 | anfD | 1557 | 518 | Oxidored_nitro | - | Fe-only nitrogenase, α-subunit |
| Avin49000 | anfH | 828 | 275 | Fer4_NifH | - | Fe-only nitrogenase protein |
| Avin49020 | anfA | 1614 | 537 | GAF + Sigma54_activat + HTH_8 | - | Sigma54 transc. Activator |
| Avin49030 | anfU | 291 | 96 | NifU | + | NifU C-terminal |

[d]FeMoco biosynthesis [g][3Fe-4S] cluster binding

**Isc system**

| Locus Tag | Gene | bp | aa | Pfam | Strand | Product |
|---|---|---|---|---|---|---|
| Avin40340 | hyp (orf3) | 207 | 68 | DUF528 | - | IscX |
| Avin40350 | fdx | 342 | 113 | Fer2 | - | Isc ferredoxin |
| Avin40360 | hscA | 1866 | 621 | HSP70 | - | Fe-S chaperone |
| Avin40370 | hscB | 522 | 173 | DnaJ + HSCB_C | - | Co-chaperone Hsc20 |
| Avin40380 | iscA | 324 | 107 | Fe-S_biosyn | - | Iron-sulfur biogenesis* |
| Avin40390 | iscU | 387 | 128 | NifU_N | - | Fe-S assembly scaffold |
| Avin40400 | iscS | 1215 | 404 | Aminotran_5 | - | Cysteine desulfurase |
| Avin40410 | iscR | 495 | 164 | Rrf2 | - | Transcription factor |
| Avin40420 | cysE2 | 780 | 259 | Hexapep | - | Serine O-acetyltransferase |

*[2Fe-2S] and [4Fe-4S] clusters transfer