

ANÁLISE COMPUTACIONAL DE PROTEÍNAS BACTERIANAS ALTAMENTE HIDROFÍLICAS DE *Ehrlichia canis* VISANDO IDENTIFICAR PADRÕES ESTRUTURAIS DE ESPIRAL ENROLADA

Gustavo J. SILVA¹; Mozart MARINS²; Ana L. FACHIN³

RESUMO

As bactérias patogênicas secretam proteínas dentro das células do hospedeiro. Essas proteínas alteram suas funções celulares podendo levá-lo à morte. O objetivo deste trabalho é investigar a presença da estrutura espiral enrolada nas dez proteínas mais hidrofílicas da bactéria *E. canis*. Para encontrar as proteínas hidrofílicas utilizou-se o método Kyte-Doolittle e para identificar as espirais utilizou-se a ferramenta de bioinformática Coiled-Coil Prediction.

INTRODUÇÃO

Uma estrutura espiral enrolada assemelha-se a uma corda. Esses motivos são encontrados em proteínas de interação, sendo responsáveis por uma variedade de funções, em sua maior parte em eucariotos. Segundo Bork (1993), as bactérias tentam imitar o comportamento do hospedeiro, dessa forma, esse domínio contribui para identificar proteínas de interação.

Para McDermott (2011), a estrutura da proteína pode apresentar relação com a sua função. Estudos recentes demonstraram a presença da estrutura espiral enrolada em proteínas de interação. Proteínas hidrofílicas também são características dessas proteínas.

Kyle e Doolittle (1982) montaram uma escala de hidropatia dos aminoácidos após observações experimentais. Nessa escala os valores são relacionados aos aminoácidos. É feito o cálculo para cada proteína, identificando as proteínas hidrofóbicas com valores positivos e proteínas hidrofílicas com valores negativos. Neste trabalho foram selecionadas as dez proteínas mais hidrofílicas. Portanto, o

¹ Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Câmpus Muzambinho. Muzambinho/MG, email: gustavo.jose@muz.ifsuldeminas.edu.br;

² Universidade de Ribeirão Preto. Ribeirão Preto/SP, email: mmarins@unaerp.br.

³ Universidade de Ribeirão Preto. Ribeirão Preto/SP, email: afachin@unaerp.br.

estudo da estrutura espiral enrolada em proteínas altamente hidrofílicas pode contribuir para identificar proteínas de interação.

MATERIAL E MÉTODOS

A ferramenta utilizada para a identificação da estrutura espiral enrolada foi a Coiled-Coil Prediction, do Pólo de Bioinformática de Lyon, na França, disponível, gratuitamente, http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.html. Essa ferramenta utiliza um banco de dados de proteínas já identificadas com domínio espiral enrolada. Mais de 200 proteínas já foram identificadas com esse domínio e são utilizadas na ferramenta para fazer comparações. A seqüência de aminoácidos foi inserida na caixa de texto específica e clicou-se no botão *SUBMIT* para enviar os dados para análise, conforme figura a seguir.

The screenshot shows the NPS@ web interface. At the top, there is a logo for 'Pôle BioInformatique Lyonnais Network Protein Sequence Analysis' with the text 'NPS@ is the IBCP contribution to PBIL in Lyon, France'. Below the logo is a navigation bar with links: [HOME], [NPS@], [SRS], [HELP], [REFERENCES], [NEWS], [MPSA], [ANTHEPROT], [Geno3D], [SuMo], [Positions], [PBIL]. A status bar below the navigation bar shows two messages: 'Monday, July 30th 2012: added support of HBVdb database (see news)' and 'Friday, January 20th 2012: upload a database: fixed error message seen with blast due to FASTA header'. The main section is titled 'COILED-COIL PREDICTION'. It contains links for '[Abstract]', '[NPS@ help]', and '[Original server]'. There is a text input field for 'Sequence name (optional)'. Below this is a section titled 'Paste a protein sequence below : help' with a text area containing a protein sequence: 'MQSIFNNKYQFKNIETKYNTLWDTTKLYKWKNSGINQFVIDTPPTISGQLHIGHVFSYC HTDFIARYQR MLGKDVFPYPIGFDDNGLPTERLVEKTKIRATDISRKEFKTICTQVSHEFRIQFKQLFQS IGISYDNDLE YHTISKDIQKISQTSFINLYNKGKLYRKLQPIFWDCIDKTAIARAEEENELSSFMNTIA FSTEAGKAIN IATTRPELMPACVAVFFNPSPDIRYQDLLGQNAIVPIFGNKVKILSDDQVKIDKGTGLVMC CTFGDEMDVY'. At the bottom of the text area are two buttons: 'SUBMIT' and 'CLEAR'. An arrow points to the 'SUBMIT' button.

Figura 1 – Ferramenta para análise de domínio espiral enrolada.

O resultado é apresentado como um gráfico de probabilidades de encontrar o domínio espiral enrolada. Para identificar a espiral enrolada deve-se encontrar de duas a sete estruturas características de α -hélice (alfa hélice). A linha vermelha indica um domínio com no mínimo duas α -hélices (14 aminoácidos), a linha verde indica no mínimo três (21 aminoácidos) e a linha azul no mínimo quatro α -hélices (28 aminoácidos). Um exemplo de resultado é apresentado na figura 2.

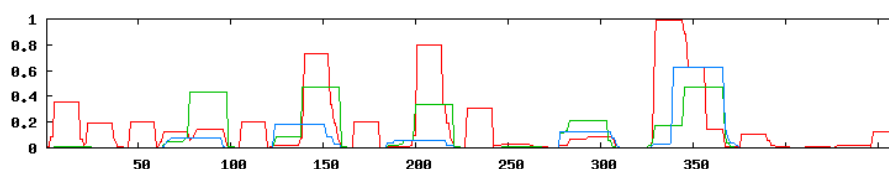


Job LUPAS (ID: 9cf2b8bb5c32) is running on NPS@ server (started on 20130808-153147).
 Results will be shown below. Please wait and don't go back.

In your publication cite :
 NPS@: Network Protein Sequence Analysis
 TIBS 2000 March Vol. 25, No 3 [291]:147-150
 Combet C., Blanchet C., Geourjon C. and Deléage G.

Coiled-Coils result for : UNK_59990

[Abstract](#) Lupas, A., Van Dyke, M., and Stock, J. Predicting Coiled Coils from Protein Sequences, Science 1991 252:1162-1164.



Under the sequence are the probabilities abbreviated to first digit for windows of 14, 21 and 28 residues and the corresponding heptads.

Figura 2 – Análise de domínio espiral enrolada.

O escopo das proteínas analisadas foram as dez proteínas mais hidrofílicas da bactéria *E. canis* encontradas pelo método Kyte-Doolittle.

RESULTADOS E DISCUSSÃO

Os cálculos da hidropatia foram baseados no método Kyle-Doolittle e foram selecionadas as dez proteínas mais hidrofílicas, conforme a Tabela 1. Estão discriminadas as informações: Nr, ID (identificação da proteína conforme descrição no proteoma), Length (número de aminoácidos) e Hydro (valor de hidropatia).

Tabela 1 – Seleção das dez proteínas mais hidrofílicas de *E. canis*

Nr	ID	Length	Hydro
1	gi73666650 ref YP_302666.1 gp140 [Ehrlichia canis str. Jake]	688	-604.20
2	gi73666693 ref YP_302709.1 hypothetical protein Ecaj_0060 [Ehrlichia canis str. Jake]	3714	-1189.70
3	gi73666759 ref YP_302775.1 hypothetical protein Ecaj_0126 [Ehrlichia canis str. Jake]	671	-580.40
4	gi73666882 ref YP_302898.1 hypothetical protein Ecaj_0253 [Ehrlichia canis str. Jake]	1206	-586.10
5	gi73667012 ref YP_303028.1 ankyrin [Ehrlichia canis str. Jake]	4245	-1862.80
6	gi73667091 ref YP_303107.1 hypothetical protein Ecaj_0472 [Ehrlichia canis str. Jake]	641	-867.90
7	gi73667147 ref YP_303163.1 putative Type IV secretory pathway VirB6 components [Ehrlichia canis str. Jake]	2030	-681.80
8	gi73667148 ref YP_303164.1 TrbL/VirB6 plasmid conjugal transfer protein [Ehrlichia canis str. Jake]	1444	-636.30
9	gi73667329 ref YP_303345.1 hypothetical protein Ecaj_0715 [Ehrlichia canis str. Jake]	1918	-634.10
10	gi73667448 ref YP_303464.1 hypothetical protein Ecaj_0835 [Ehrlichia canis str. Jake]	1267	-744.20

Após a análise probabilística na ferramenta, as proteínas foram divididas em três grupos. No Grupo 1 estão as proteínas com alta probabilidade de possuir espiral enrolada (acima de 80%), no Grupo 2 as proteínas com probabilidade média de possuir a estrutura (acima de 40% e abaixo de 80%) e no Grupo 3 as proteínas com baixa probabilidade (abaixo de 40%).

Grupo 1: proteínas identificadas com os códigos 73666693, 73667012, 73667091, 73667147, 73667148 e 73667329.

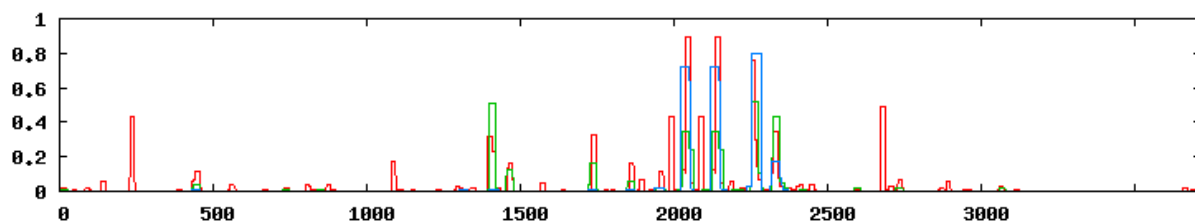


Figura 3 – Resultado de probabilidade da proteína 73666693.

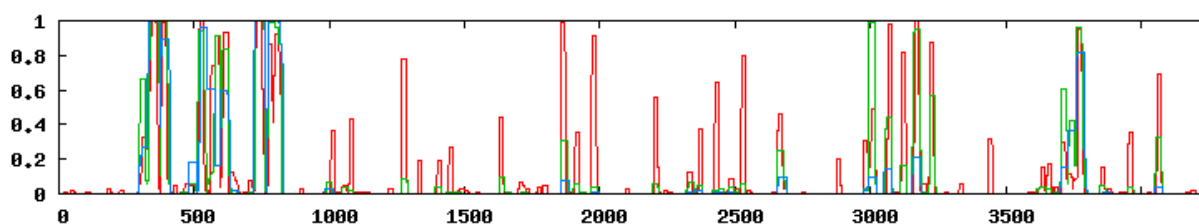


Figura 4 – Resultado de probabilidade da proteína 73667012.

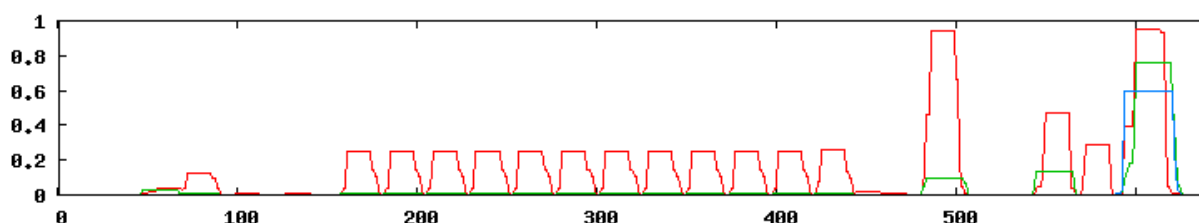


Figura 5 – Resultado de probabilidade da proteína 73667091.

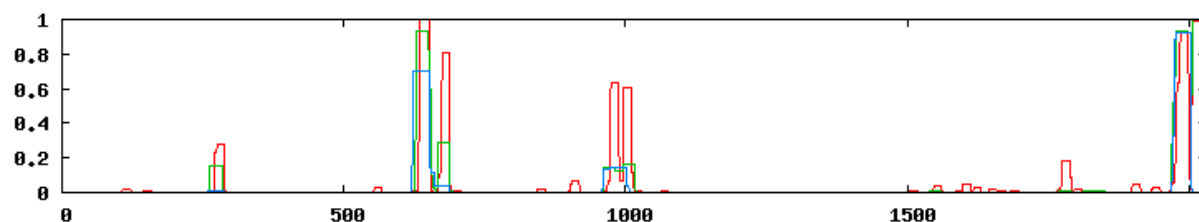


Figura 6 – Resultado de probabilidade da proteína 73667147.

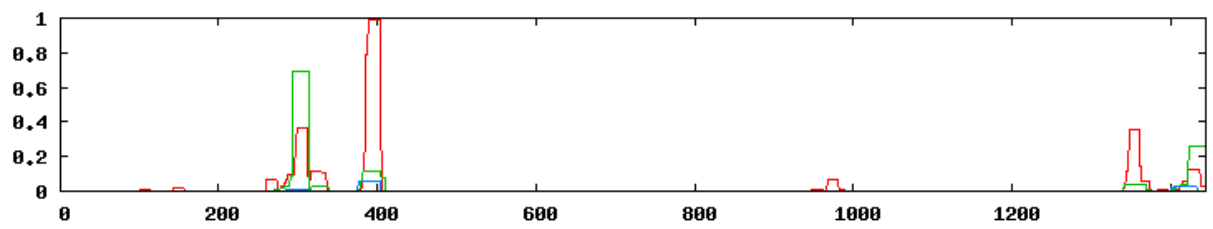


Figura 7 – Resultado de probabilidade da proteína 73667148.

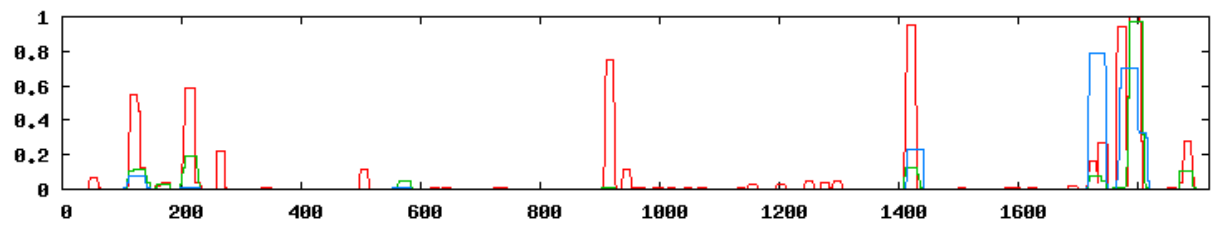


Figura 8 – Resultado de probabilidade da proteína 73667329.

Grupo 2: proteínas identificadas com os códigos 73667448 e 73666759.

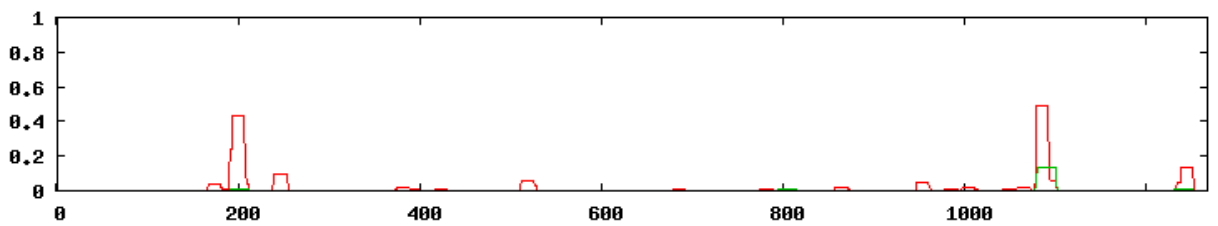


Figura 9 – Resultado de probabilidade da proteína 73667448.

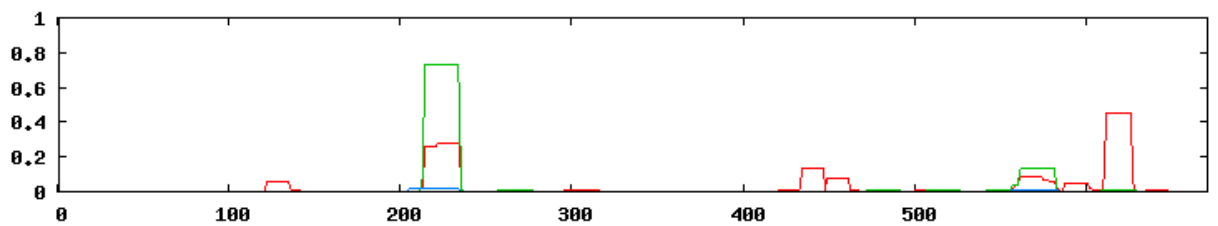


Figura 10 – Resultado de probabilidade da proteína 73666759.

Grupo 3: proteínas identificadas com os códigos 73666650 e 73666882

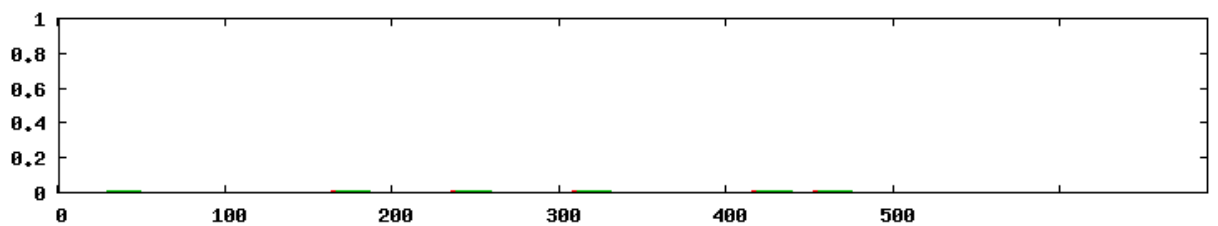


Figura 11 – Resultado de probabilidade da proteína 73666650.

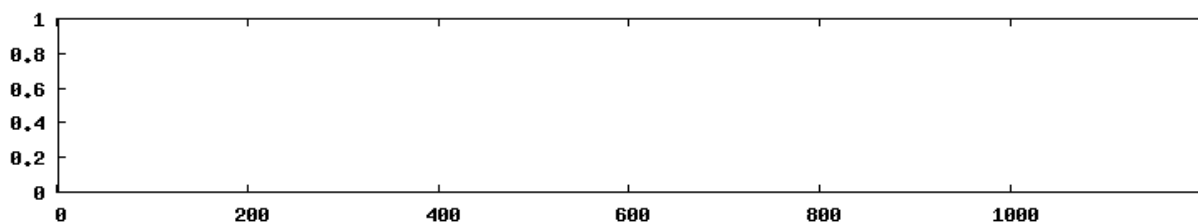


Figura 12 – Resultado de probabilidade da proteína 73666882.

CONCLUSÕES

As características quanto à hidropatia e presença da estrutura espiral enrolada são importantes para identificar proteínas de interação. Os resultados apresentaram grande probabilidade de possuir a estrutura espiral enrolada em seis das dez proteínas analisadas. Portanto, essa característica demonstrou-se presente em proteínas hidrofílicas, assim, deve ser investigada em mais proteínas hidrofílicas visando selecionar proteínas para testes laboratoriais e desenvolvimento de fármacos.

REFERÊNCIAS BIBLIOGRÁFICAS

LUPAS, A., VAN DYKE, M., STOCK, J. Predicting Coiled Coils from Protein Sequences. **Science** v 252, p 1162-1164, 1991.

KYTE, J., DOOLITTLE, R. A simple method for displaying the hydropathic character of a protein. **Journal of Molecular Biology**, v 157, p 105-132, 1982.

MCDERMOTT, J.E., et al., Computational prediction of type III and IV secreted effectors in gram-negative bacteria. **Infect Immun** v 79, p 23-32, 2011.

BORK, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally. **Proteins** v 17, p 363–364, 1993.