

## Supporting Information

# **The genomic context of highly trimethoprim-resistant DfrB dihydrofolate reductases highlights an emerging threat to public health**

**Claudèle Lemay-St-Denis<sup>1,2,3</sup>, Sarah-Slim Diwan<sup>1,2,3</sup> and Joelle N. Pelletier<sup>1,2,3,4</sup>**

<sup>1</sup> Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Québec, Canada

<sup>2</sup> PROTEO, the Québec Network for Research on Protein, Function, Engineering and Applications, Qc, Canada

<sup>3</sup> CGCC, Center in Green Chemistry and Catalysis, Montréal H3A 0B8, Qc, Canada

<sup>4</sup> Chemistry Department, Université de Montréal, Montréal, Québec, Canada

**Table S1. Protein sequence identity of the ten members of the DfrB family <sup>a</sup>**

	DfrB1	DfrB2	DfrB3	DfrB4	DfrB5	DfrB6	DfrB7	DfrB9	DfrB10	DfrB11
DfrB1	100%	78%	78%	77%	88%	87%	88%	77%	82%	78%
DfrB2	17	100%	86%	74%	79%	82%	78%	84%	92%	88%
DfrB3	17	11	100%	79%	82%	83%	79%	85%	92%	90%
DfrB4	18	20	16	100%	77%	77%	76%	74%	77%	79%
DfrB5	9	16	14	18	100%	91%	94%	77%	83%	81%
DfrB6	10	14	13	18	7	100%	91%	79%	86%	83%
DfrB7	9	17	16	19	5	7	100%	76%	82%	81%
DfrB9	18	12	12	20	18	16	19	100%	88%	86%
DfrB10	14	6	6	16	13	11	14	9	100%	92%
DfrB11	17	9	8	18	15	13	15	11	6	100%

<sup>a</sup> Values in percentage correspond to protein sequence identity values; values in grey represent the number of mutations between sequences.

**Table S2. DfrB sequences**

Gene	Genbank ID	Gene sequence	Protein sequence
<i>dfrb1</i>	KAB1871659.1	ATGGAACGAAGTAGCAATGAAGTCAGTAATC CAGTTGCTGGCAATTTTGTATTCCCATCGAACG CCACGTTTGGTATGGGAGATCGCGTGCGCAAG AAATCCGGCGCCGCTGGCAAGGTCAGATTGT CGGGTGGTACTGCACAAATTTGACCCCCGAAG GCTACGCCGTCGAGTCTGAGGCTCACCCAGGC TCAGTACAGATTTATCCTGTTGCGGCGCTTGA ACGCATCAACTGA	MERSSNEVSNPV AGNFVFPNATF GMGDRVRKKSG AAWQQQIVGW YCTNLTPEGYAV ESEAHPGSVQIY PVAALERIN
<i>dfrb2</i>	FAA00064.1	ATGGGTCAAAGTAGCGATGAAGCCAACGCTC CCGTTGCAGGGCAGTTTGCCTTCCCCTGAGT GCCACCTTTGGCTTAGGGGATCGCGTACGCAA GAAATCTGGTCCGCTTGGCAGGGTCAAGTCG TCGGTTGGTATTGCACAAAACCTCACTCCTGAA GGCTATGCGGTCGAGTCCGAATCCCACCCAGG	MGQSSDEANAP VAGQFALPLSAT FGLGDRVRKKS GAAWQQQVVG WYCTKLTPGY AVESESHPGSVQ IYPVAALERVA

		CTCAGTGCAAATTTATCCTGTGGCTGCACTTGA ACGTGTGGCCTAA	
<i>dfrb3</i>	ACR57831.1	ATGGACCAACACAACAATGGAGTCAGTACTC TAGTTGCTGGCCAGTTTGCCTCCCATCGCAC GCCACGTTTGGCCTGGGAGATCGCGTGCGCAA GAAATCTGGCGCCGCTTGGCAGGGTCAAGTTG TCGGGTGGTACTGCACAAAAGTACCCCTGAA GGCTATGCCGTCGAGTCCGAGTCTACCCCGG TTCAGTACAGATTTATCCTGTGGCTGCGCTTGA ACGCGTGGCCTGA	MDQHNNGVSTL VAGQFALPSHA TFGLGDRVRKKS GAAWQGVVVG WYCTKLTPEGY AVESESHPGSVQ IYPVAALERVA
<i>dfrb4</i>	ALF62656.1 <sup>a</sup>	ATGAATGAAGGAAAAAATGAGGTCAGTACTT CAGCTGCTGGCCGGTTTCGATTCCCATCAAAC GCCACGTTTGCCTTGGGGGATCGCGTACGCAA GAAGTCTGGCGCTGCTTGGCAGGGGCGCATTG TCGGGTGGTACTGCACAACACTTACCCCTGAA GGCTACGCCGTCGAGTCCGAATCTACCCAGG CTCAGTCCAGATTTATCCCATGACTGCGCTTG AACGGGTGGCCTGA	MNEGKNEVSTS AAGRFAFSPNAT FALGDRVRKKS GAAWQGRIVG WYCTTLTPEGYA VESESHPGSVQIY PMTALERVA
<i>dfrb5</i>	AAX46054.1	ATGGACCAAGGCAGAAGTGAAGTCAGTAATC CAGTTGCTGGCCAGTTTGCCTTCCCTTCAAAC GCCGCGTTCGGAATGGGAGATCGCGTGCGCA AGAAATCTGGCGCCGCTTGGCAAGGCCAGATT GTCGGGTGGTACTGCACAAAATTGACCCCTGA AGGGTACGCTGTGAGTCTGAGGCTCACCCCTG GCTCGGTACAGATTTATCCTGTTGCGGCACTG GAACGCATCAACTGA	MDQGRSEVSNP VAGQFAPSNA AFMGDRVRKKS SGAAWQGQIVG WYCTKLTPEGY AVESEAHPGSV QIYPVAALERIN
<i>dfrb6</i>	ADO00942.1	ATGGACCAAGGTAGCAATGAAGTCATTAATC CAGTCGCTGGCCAGTTTGCCTCCCATCGAAC GCCACGTTTGGTATGGGAGATCGCGTGCGCAA GAAATCTGGCGCCGCTTGGCAAGGTCAGATTG TCGGGTGGTACAGCACAAGTTGACCCCTGA AGGCTACGCTGTGAGTCTGAGGCTCACCCCTG GCTCGGTACAGATTTATCCTGTTGCCGCGCTTG AACCGTCAACTGA	MDQGSNEVINP VAGQFASPSNA TFMGDRVRKKS SGAAWQGQIVG WYSTKLTPEGYA VESEAHPGSVQI YPVAALERN
<i>dfrb7</i>	ADB54781.1	ATGGACCAAGGTAGCAATGAAGTCGGTAATC CAGTTGCGGGCCAGTTTTCGTTCCCATCGAAC GCCGCGTTTAGTATGGGAGATCGCGTGCGCAA GAAATCGGGCGCCGCTTGGCAAGGTCAGATT GTCGGGTGGTACTGCACAAAAGTTGACCCCTGA AGGCTACGCTGTGAGTCTGAGGCTCACCCCTG GCTCGGTACAGATTTATCCTGTTGCCGCGCTTG AACCGTCAACTGA	MDQGSNEVGNP VAGQFSFSPNAA FSMGDRVRKKS GAAWQGVVVG WYCTKLTPEGY AVESEAHPGSV QIYPVAALERIN GVQG
<i>dfrb9</i>	AGM20434.1	ATGAATCAAAGTAGCAATTGCATCAGCACTCC AGTTGTTGGACAGTTTGCCTGCCATTTCAACC CACGTTCCGCTGGGAGATCGCGTACGCAAG AAGTCTGGCGCCGCTTGGCAAGGTAAGTTGT CGGCTGGTACTGCACAAAATTAACCCCTGAAG GCTACGCGTTCGAGTCCGAAGCTCATCCAGGC TCAGTGCAGATTTATCCTGTGGCTGCGCTTGA ACGCGTGGCCTAA	MNQSSNCISTPV VGQFALPFQPTF GLGDRVRKKS AAWQGVVVG YCTKLTPEGYAV ESEAHPGSVQIY PVAALERVA
<i>dfrb10</i>	ALZ46148.1	ATGGATCAAAGTAGCAATGAAGTCAGCACTC CAGTTGCTGGCCAGTTTGCCTCCCATTGCGC	MDQSSNEVSTPV AGQFALPLRATF

		GCCACGTTTGGCCTGGGAGATCGCGTACGCAA GAAATCTGGCGCCGCTTGGCAAGGTCAAGTTG TCGGCTGGTACTGCACAAAAGTACCCCTGAA GGCTATGCAGTCGAGTCCGAGTCTACCCAGG CTCAGTACAGATTTATCCTGTGGCTGCGCTTGA ACGCGTGGCCTAA	GLGDRVRRKKS AAWQGVVGV YCTKLTPEGYAV ESESHPGSVQIYP VAALERVA
<i>dfrb11</i>	PKO69073.1	ATGGATCAAAGTAGTAAAGAGGTTGGCACTC CCGTTGTTGGCCAGTTTGCCTCCCGTTCGCAC GCCACGTTTGGCCTTGGAGACCGCGTTCGCAA GAAATCGGGCGCCGCTTGGCAGGGTCAAGTT GTGGGCTGGTATTGCACAAAGCTGACCCCTGA AGGCTATGCCGTCGAGTCCGAGTCTACCCAG GCTCGGTACAAATTTATCCAGTGAATGCGCTT GAACCGTGGCCTGA	MDQSSKEVGT VVGQFALPSHA TFGLGDRVRRKKS GAAWQGVVGV WYCTKLTPEGY AVESESHPGSVQ IYPVNALERVA

<sup>a</sup> This sequence differs from CARD's reference sequence ABY55281.1 at the highlighted region (here Gly instead of Asp). Since we previously characterized DfrB4 using ALF62656.1 sequence [1], we used it in this study.

1. Toulouse, J.L.; Shi, G.; Lemay-St-Denis, C.; Ebert, M.C.C.J.C.; Deon, D.; Gagnon, M.; Ruediger, E.; Saint-Jacques, K.; Forge, D.; Vanden Eynde, J.J.; et al. Dual-Target Inhibitors of the Folate Pathway Inhibit Intrinsically Trimethoprim-Resistant DfrB Dihydrofolate Reductases. *ACS Med. Chem. Lett.* **2020**, acsmedchemlett.0c00393, doi:10.1021/acsmchemlett.0c00393.