Journal of Molecular Graphics and Modelling 71 (2017) 135–153

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Topical Perspectives

Structure-based analysis of Bacilli and plasmid dihydrofolate reductase evolution

Mona Alotaibi[a,b], Ben Delos Reyes[a], Tin Le[a], Phuong Luong[a], Faramarz Valafar[c], Robert P. Metzger[a], Gary B. Fogel[d], David Hecht[a,e,*]

[a] Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA 92182-1030, USA
[b] King Saud University, P.O. Box 245714, Riyadh 11312, Saudi Arabia
[c] Bioinformatics and Medical Informatics Research Center, San Diego State University, San Diego, CA 92182-7720, USA
[d] Natural Selection, Inc., 6480 Weathers Place, Suite 350, San Diego, CA 92121, USA
[e] Department of Chemistry, Southwestern College, 900 Otay Lakes Rd., Chula Vista, CA 91910, USA

ARTICLE INFO

Article history:
Received 25 April 2016
Received in revised form 4 October 2016
Accepted 10 October 2016
Available online 22 November 2016

Keywords:
Dihydrofolate Reductase
DHFR
Structure-based Alignment
Phylogeny
Evolution
Plasmid evolution
Horizontal Gene Transfer
Network of Coupled Protein Motion

ABSTRACT

Dihydrofolate reductase (DHFR), a key enzyme in tetrahydrofolate-mediated biosynthetic pathways, has a structural motif known to be highly conserved over a wide range of organisms. Given its critical role in purine and amino acid synthesis, DHFR is a well established therapeutic target for treating a wide range of prokaryotic and eukaryotic infections as well as certain types of cancer. Here we present a structural-based computer analysis of bacterial (Bacilli) and plasmid DHFR evolution. We generated a structure-based sequence alignment using 7 wild-type DHFR x-ray crystal structures obtained from the RCSB Protein Data Bank and 350 chromosomal and plasmid homology models we generated from sequences obtained from the NCBI Protein Database. We used these alignments to compare active site and non-active site conservation in terms of amino acid residues, secondary structure and amino acid residue class. With respect to amino acid sequences and residue classes, active-site positions in both plasmid and chromosomal DHFR are significantly more conserved than non-active site positions. Secondary structure conservation was similar for active site and non-active site positions. Plasmid-encoded DHFR proteins have greater degree of sequence and residue class conservation, particularly in sequence positions associated with a network of concerted protein motions, than chromosomal-encoded DHFR proteins. These structure-based were used to build DHFR specific phylogenetic trees from which evidence for horizontal gene transfer was identified.

1. Introduction

Both prokaryotes and eukaryotes require reduced folate cofactors for the biosynthesis of a broad range of cellular components [1]. More specifically, tetrahydrofolate cofactors are essential for the synthesis of purines, thymidine and glycine [2]. While most bacteria and plants produce folate cofactors through *de novo* biosynthesis, a variety of mammals, and some prokaryotes lack this ability and must convert dietary folates to reduced folates through the use of their enzymatic salvage pathways for survival. Most bacterial, plant, and animal species possess a pathway (Fig. 1), by which tetrahydrofolate which has been oxidized to dihydrofolate during its use in biosynthetic processes, is recycled back to tetrahydrofolate through use of NADPH and the activity of the enzyme dihydrofolate reductase (DHFR).

Folate effects the efficiency of DNA replication, repair, and methylation. Rapidly proliferating cells including leukocytes, and enterocytes require high amounts of folate. Epidemiological studies have indicated that folate deficiency is often associated with increased risk of breast cancer and birth defects [3]. Indeed, low folate homeostasis may induce hypo-methylation of DNA which may promote development of cancer [4].

There is a stark difference in tetrahydrofolate salvage pathways between bacteria and humans which makes the bacterial pathways ideal targets for antibacterial agents [5]. The development of bacterial DHFR inhibitors that will *not* affect human DHFR continues to be an area of great interest for treatment of bacterial infections [2,6,7]. Such folate analogs, (e.g., methotrexate and trimethoprim),

* Corresponding author at: Department of Chemistry, Southwestern College, 900 Otay Lakes Road, Chula Vista, CA 91910, USA.
E-mail addresses: monaotaibi@ksu.edu.sa (M. Alotaibi), faramarz@mail.sdsu.edu (F. Valafar), rmetzger@mail.sdsu.edu (R.P. Metzger), gfogel@natural-selection.com (G.B. Fogel), dhecht@swccd.edu (D. Hecht).

http://dx.doi.org/10.1016/j.jmgm.2016.10.011
1093-3263/

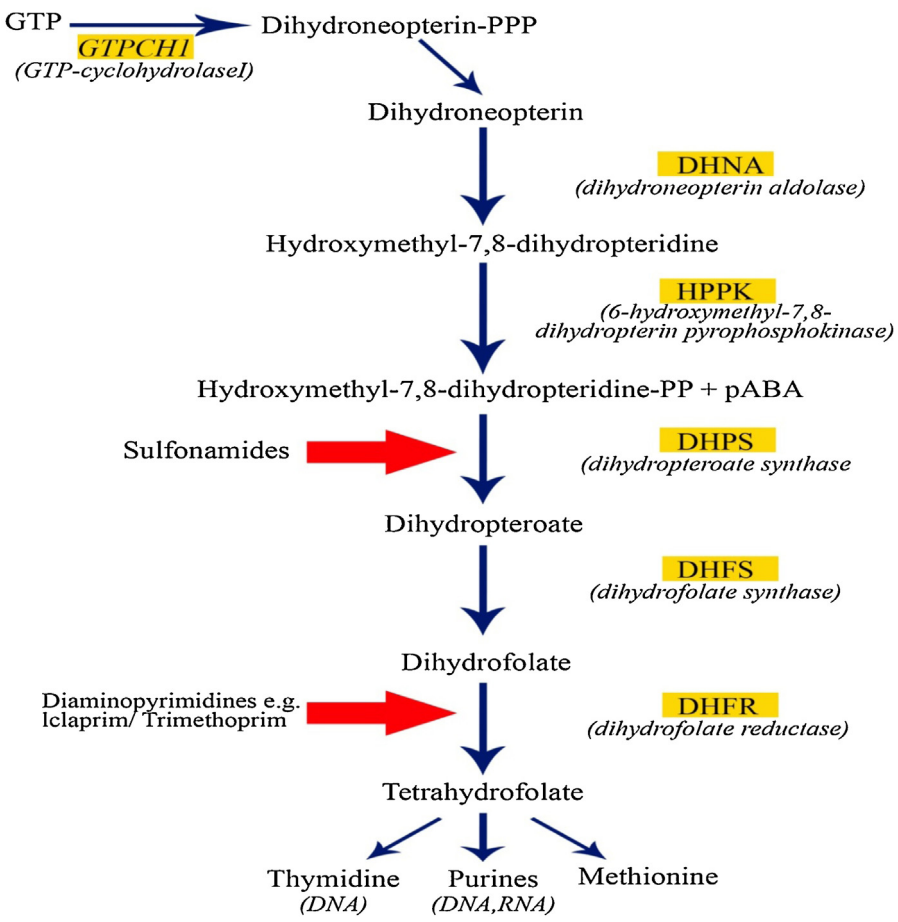

Fig. 1. Reactions in the folate pathway [1,2].

are known to inhibit DHFR, resulting in the death of cancerous human cells and bacterial cells, respectively. These agents bind tightly to the catalytic site of the enzyme in a region known to form a hydrophobic dihydrofolate-binding site [8].

However, the ability of bacteria to adapt rapidly to new conditions has led to the development of resistance to antifolate agents. The most common resistance results from mutations leading to the synthesis of an altered form of DHFR with reduced affinity for an inhibitor or inhibitors [9]. Acquired resistance is also associated with mobile elements temporarily located on the chromosome by transposon movement, referred to as plasmid-encoded DHFR [10–14].

Plasmids containing genes encoding drug resistant DHFR variants were first discovered in *Escherichia coli* and *Klebsiella aerogenes* by Fleming [15]. To date, more than 30 different plasmid-encoded DHFR genes have been identified. Plasmid DHFR genes encode two distinct DHFR types, I and II, and have therefore been divided into two families, *dfrA* and *dfrB* [16]. The majority of the *dfrA* family is carried as a gene cassette encoding DHFR variants ranging between 157 and 187 amino acid residues. There is a high degree of conservation within the N-terminus, with 12 residues strictly conserved across all enzymes [17]. The second family (*dfrB1*, *dfrB2*, *dfrB3*, *dfrB4*, *dfrB5* and *dfrB6*) also forms part of a gene cassette, and encodes the enzymes DHFR IIa, IIb and IIc, respectively. These enzymes have significantly shorter amino acid sequences of 78 residues which is roughly 42–50% of the amino acid residues of the enzymes encoded by the *dfrA* family. Members of the *dfrB* family are completely non-responsive to trimethoprim, methotrexate, and other folate analogs and appear to have a distinctive evolutionary origin [18,19].

Phylogenetic analyses of DHFR [20–23] as well as studies and simulations of evolutionary processes [14,24–30] have been performed to study the rapid development of resistance to antifolate agents in response to the excessive use of antibiotics [31]. Our efforts focus on predictive models of the evolution of drug resistance in bacterial pathogens having the goal of developing novel therapeutics targeting drug resistant strains *in anticipation* of their development in the wild [26,27,32]. This proactive approach to drug discovery requires a thorough understanding of the evolutionary processes specific to the gene of interest. To this end we have performed a structural based analysis of sequence and structure variation in *Bacilli* DHFR. *Bacilli* are a prokaryotic class belonging to the taxonomic phylum *Firmicutes* [33–35]. Members of this class are generally Gram-positive bacteria. The *Bacilli* class consists of two orders *Bacillales* and *Lactobacillales* (Fig. 2).

Here we present an analysis of DHFR evolution based on a structure-based alignment of 273 *Bacilli* species and 77 *dfrA* family DHFR protein sequences of which five come from *Bacilli* species (*Planococcus citreus, Listeria monocytogenes, Staphylococcus aureus (2), Enterococcus faecium*). The remainder of the plasmid sequences include non-*Bacilli* sequences of which *E. coli.* (30) and *Klebsiella pneumoniae* (8) were the most common. Using this structure-based alignment we evaluated the degree of site-specific variability/conservation of amino acid sequence, secondary structure, and residue class. Using the structure-based alignment, we generated phylogenetic trees and explored the evolution of *Bacilli* DHFR in terms of sequence and structure conservation between chromosomal and plasmid DHFR, *Bacillales* and *Lactobacillales*, in order to attempt to correlate the observed differences to their environment.