

## Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes

TIM SODERBERG

Division of Science and Mathematics, University of Minnesota, Morris, 600 E. 4th Street, Morris, MN 56267, USA (soderbt@mrs.umn.edu)

Received June 7, 2004; accepted September 15, 2004; published online October 12, 2004

**Summary** A phylogenetic analysis of the genes encoding enzymes in the pentose phosphate pathway (PPP), the ribulose monophosphate (RuMP) pathway, and the chorismate pathway of aromatic amino acid biosynthesis, employing data from 13 complete archaeal genomes, provides a potential explanation for the enigmatic phylogenetic patterns of the PPP genes in archaea. Genomic and biochemical evidence suggests that three archaeal species (*Methanocaldococcus jannaschii*, *Thermoplasma acidophilum* and *Thermoplasma volcanium*) produce ribose-5-phosphate via the nonoxidative PPP (NOPPP), whereas nine species apparently lack an NOPPP but may employ a reverse RuMP pathway for pentose synthesis. One species (*Halobacterium* sp. NRC-1) lacks both the NOPPP and the RuMP pathway but may possess a modified oxidative PPP (OPPP), the details of which are not yet known. The presence of transketolase in several archaeal species that are missing the other two NOPPP genes can be explained by the existence of differing requirements for erythrose-4-phosphate (E4P) among archaea: six species use transketolase to make E4P as a precursor to aromatic amino acids, six species apparently have an alternate biosynthetic pathway and may not require the ability to make E4P, and one species (*Pyrococcus horikoshii*) probably does not synthesize aromatic amino acids at all.

**Keywords:** aromatic amino acid biosynthesis, chorismate, genomic analysis, nucleotide biosynthesis, pentose phosphate pathway, ribulose-5-phosphate, transketolase.

### Introduction

The pentose phosphate pathway (PPP) is a ubiquitous metabolic pathway in bacteria and eukaryotes. The function of the PPP is threefold: to generate NADPH for reducing power in reductive biosynthesis, to convert three- and six-carbon sugar phosphates from the glycolytic pathway into ribose-5-phosphate (R5P) for nucleotide biosynthesis, and to provide erythrose-4-phosphate (E4P), a precursor for aromatic amino acid biosynthesis. The classical PPP is divided into two phases (Figure 1). In the oxidative PPP (OPPP), glucose-6-phosphate (G6P) undergoes oxidative decarboxylation to form ribulose-5-phosphate (Ru5P) with the concomitant reduction of two

molecules of NADP<sup>+</sup> to NADPH. In the nonoxidative phase (NOPPP), R5P isomerase converts Ru5P to R5P, while the enzymes Ru5P 3-epimerase, transketolase and transaldolase act in concert to shuttle excess Ru5P back into three- and six-carbon glycolytic intermediates, with E4P formed in the process.

The role of the PPP in archaea has been an open question for more than a decade. Before the advent of whole genome sequencing projects, it was reported that activities of NOPPP enzymes could be detected in the methanogenic archaeon *Methanococcus maripaludis*, but OPPP enzyme activities were not observed (Yu et al. 1994). The authors proposed that R5P synthesis in methanococci occurs solely via the NOPPP enzymes acting in the 'reverse' direction (F6P and GAP to Ru5P), and not by oxidative decarboxylation of G6P. This hypothesis was further supported by studies with *Methanococcus voltae* (Choquet et al. 1994) and *M. maripaludis* (Tumbula et al. 1997), in which <sup>13</sup>C labeling patterns were consistent with R5P and E4P synthesis via the NOPPP. Further genomic analysis indicated that *Methanocaldococcus jannaschii* has a complete NOPPP but is missing the OPPP entirely (Selkov et al. 1997). Conversely, <sup>13</sup>C labeling studies with other methanogenic archaea including *Methanobacterium thermoautotrophicum* (Ekiel et al. 1983, Eisenreich et al. 1991, Choquet et al. 1994) showed ribose labeling patterns consistent with the synthesis of R5P by the OPPP. The same studies also revealed labeling patterns in phenylalanine and tyrosine that were inconsistent with the biosynthesis of these aromatic amino acids starting from E4P, calling into question the role of E4P and the NOPPP in these species.

Recent progress in the sequencing of archaeal genomes has revealed a confusing picture of the PPP in the archaeal domain (Table 1). Among the thirteen completed archaeal genomes included in the Clusters of Orthologous Groups (COG) database (Tatusov et al. 1997, 2001; <http://www.ncbi.nlm.nih.gov/COG/>), orthologs of genes encoding enzymes of the OPPP are virtually nonexistent: only one species, *Halobacterium* sp. NRC-1, has a recognizable ortholog of 6-phosphogluconate dehydrogenase (6PGDH), whereas glucose-6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate lactonase (6PGL) orthologs are missing in all 13 genomes. Nonoxidative PPP orthologs are present to a varying extent in archaea: all 13 genomes contain R5P isomerase orthologs, but only three species (*M. jannaschii*, *Thermoplasma volcanium* and



Table 1. Phylogenetic patterns of archaeal gene orthologs from the pentose phosphate pathway (PPP) and the ribulose monophosphate (RuMP) pathway. GenBank accession numbers are listed for each ortholog.

Genome	Oxidative PPP		Nonoxidative PPP				RuMP pathway	
	Glucose-6-phosphate dehydrogenase COG 0364	6-Phospho-gluconate dehydrogenase COG 1023	Ribose-5-phosphate isomerase COG 0120	Transketolase (N- and C-terminal subunits) COG 3959/3958	Transaldolase COG 0176	Ribulose-5-phosphate-3-epimerase COG 0036	Hexulose-6-phosphate synthase COG 0269	Hexulose-6-phosphate isomerase COG 0794
<i>M. jannaschii</i>			NP_248616	NP_247665 NP_247663	NP_247955	NP_247664	NP_248451	NP_248242
<i>T. acidophilum</i>			NP_394337	NP_394090 NP_394089	NP_394088	NP_394770		
<i>T. volcanium</i>			NP_111236	NP_111187 NP_111186	NP_111185	NP_110783		
<i>P. abyssi</i>			NP_126456	NP_126136 NP_126137			NP_127388	NP_127388
<i>A. pernix</i>			NP_147400	NP_147340 NP_147339			NP_147623	NP_148269
<i>P. aerophilum</i>			NP_559026	NP_559642 NP_559641			NP_559470	NP_559333
<i>S. solfataricus</i>			NP_342463	NP_341848 NP_341849			NP_341758	NP_341708
<i>P. horikoshii</i>			NP_143254				NP_143767	NP_143767
<i>M. kandleri</i>			NP_614948				NP_613440	NP_614946
<i>M. thermo-autotrophicum</i>			NP_275751				NP_276588	NP_276659
<i>A. fulgidus</i>			NP_069776				NP_070134	NP_070624
<i>M. acetivorans</i>			NP_616610				NP_619467	NP_616320
<i>Halobacterium</i>		NP_281126	NP_280908					

2000, von Mering et al. 2003; <http://string.embl.de/>) was used to compare phylogenetic profiles and gene position. Homology searches were performed with the BLASTP program on the NCBI website.

## Results and discussion

### *Biosynthesis of ribose-5-phosphate*

The genomic evidence strongly suggests that in archaea, as in bacteria and eukaryotes, Ru5P is the original source of the pentose moiety in nucleotides. Orthologs of R5P isomerase (COG 0120), which catalyzes the reversible interconversion of R5P and Ru5P, are present in all 13 archaeal genomes compared in this study (Table 1). Also ubiquitous in archaeal genomes are most of the genes for nucleotide biosynthesis, including putative genes for phosphoribosylpyrophosphate (PRPP) synthase (COG 0462), orotate phosphoribosyltransferase (COG 0461) and glutamine phosphoribosyl amidotransferase (COG 0034), the key enzymes responsible for in-

corporating R5P into nucleotides. The question thus becomes how Ru5P is synthesized by those archaea that lack a complete set of NOPPP orthologs.

The RuMP pathway, operating in the 'reverse' direction, is a possible alternate source of Ru5P. The two reactions of the RuMP pathway—an aldol condensation and a 3-ketose to 2-ketose isomerization—are thermodynamically reversible. Furthermore, it appears that in archaeal genomes, the phylogenetic profiles of the RuMP pathway and the NOPPP are for the most part complementary. Orthologs for *hps* and *phi* (COG 0269 and COG 0794, respectively), are present in 10 of the 13 genomes (Table 1), and all but one of these (*M. jannaschii*) lack a complete NOPPP, which is the source of Ru5P in the 'forward' RuMP pathway of methylotrophic bacteria. This indicates that the role of the RuMP pathway in these archaeal species may be reversed: that is, its function may be to produce Ru5P from F6P.

The operation of a reverse RuMP pathway for the production of Ru5P implies the concurrent production of formaldehyde, which is toxic to cells. Some mechanism must therefore

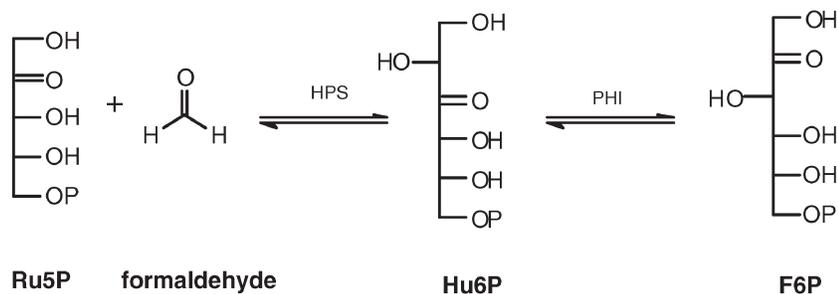


Figure 2. The reactions of the ribulose monophosphate (RuMP) pathway. Abbreviations: Ru5P = ribulose-5-phosphate; HPS = 3-hexulose-6-phosphate synthase; Hu6P = 3-hexulose-6-phosphate; PHI = 3-hexulose-6-phosphate isomerase; and F6P = fructose-6-phosphate.

exist to remove formaldehyde in all species that synthesize RuMP in this manner. Five of the 13 genomes analyzed in this study (the four methanogens plus the sulfate reducer *Archaeoglobus fulgidus*) contain members of COG 1795, a formaldehyde-activating enzyme first characterized in the methylotroph *Methylobacterium extorquens* (Vorholt et al. 2000). It is significant that in four of these five species this gene is fused to *hps*, possibly encoding a bifunctional enzyme that activates formaldehyde for oxidation as soon as it is produced as a by-product of Ru5P synthesis. Also present in nine of the 13 genomes are members of COG 2414, a tungsten-dependent formaldehyde ferredoxin oxidoreductase that has been characterized in *Pyrococcus furiosus* (Roy et al. 1999). Among the nine archaeal genomes that contain the RuMP pathway genes, only *Sulfolobus solfataricus* is missing both COG 1795 and COG 2414. This genome does, however, contain an open reading frame (GenBank accession no. AAK40795) that appears to be an ortholog of a glutathione-independent formaldehyde dehydrogenase characterized from *Pseudomonas putida* (Ogushi et al. 1984, Tanaka et al. 2002).

Experimental confirmation of the biochemical function of the protein encoded by *hps* in archaea will be critical to confirming or denying the role of the RuMP pathway in ribose synthesis. The homology between some archaeal *hps* orthologs and the biochemically characterized *hps* gene in *B. subtilis* is relatively low, with expected values ranging from  $10^{-43}$  for the *Pyrococcus abyssi* gene to  $10^{-14}$  for the *M. jannaschii* gene. In addition, *hps* orthologs are fused in several archaeal genomes to members of COG 0684, which encodes demethylmenaquinone methyltransferase. This fusion event further clouds predictions of the function of *hps* in archaea.

An alternate interpretation of results from previous isotopic labeling studies can be provided in light of the genomic data now available. Previous reports (Eisenreich et al. 1991, Choquet et al. 1994) stated that  $^{13}\text{C}$  labeling patterns in several methanogenic archaea (but not in methanococci) pointed to the action of OPPP enzymes in the conversion of F6P to Ru5P. However, the same labeling pattern observed in these studies would also be expected if the F6P to Ru5P conversion was carried out instead by a reverse RuMP pathway, a more likely scenario given the lack of OPPP orthologs in methanogenic genomes.

Although both *T. volcanium* and *T. acidophilum* are missing the RuMP pathway genes, these species do have a complete NOPPP that could provide a source of Ru5P. *Methanocaldococcus jannaschii* has both a complete NOPPP and an RuMP

pathway, therefore Ru5P could derive from either or both of these pathways. However, labeling studies (Choquet et al. 1994, Tumbula et al. 1997) suggest that the NOPPP alone is responsible for ribose biosynthesis in methanococci.

Only *Halobacterium* lacks both an NOPPP and an RuMP pathway. However, *Halobacterium* is also the only archaeal genome that contains a recognizable OPPP gene (the ortholog of genes encoding 6PGDH (COG 1023)), which suggests that pentose synthesis in *Halobacterium* may take place via a modified OPPP. Although it is not yet clear how this could occur, there are several reasonable possibilities. *Halobacterium* may have a novel G6PDH that is not recognizable by sequence homology to known enzymes. 6-Phosphogluconate could be produced by oxidation of glucose to gluconate by glucose dehydrogenase (COG 1063, GenBank accession no. AAG18991), followed by phosphorylation of gluconate by an as yet uncharacterized putative sugar kinase (e.g., COG 0524, GenBank accession no. AAG20057). Alternatively, the oxidative carboxylation of gluconate to ribulose could be followed by phosphorylation to Ru5P by an as yet unidentified ribulokinase, if the predicted 6PGDH enzyme in *Halobacterium* is capable of oxidizing a nonphosphorylated substrate. Biochemical characterization of the predicted 6PGDH enzyme in *Halobacterium*, as well as experimental confirmation of the presence or absence of G6PDH, gluconate-6-kinase and ribulokinase activities in cell extracts, will be necessary to address the questions surrounding the origin of Ru5P in *Halobacterium*.

#### Biosynthesis of erythrose-4-phosphate and aromatic amino acids

Another critical function of the NOPPP in bacteria and eukaryotes is the generation of E4P as a precursor for aromatic amino acid biosynthesis. The curious NOPPP phylogeny in archaeal genomes can be explained by the presence in some archaea of an alternate biosynthetic pathway that does not use E4P. The first phase of the classical aromatic amino acid biosynthesis pathway involves the seven-step synthesis of chorismate, beginning with the condensation of E4P and phosphoenolpyruvate (PEP) to form 2-dehydro-3-deoxy-D-arabinoheptulosonate-7-phosphate (DAHP) and the subsequent cyclization to 3-dehydroquinate (DHQ). Most of this pathway is present in 12 of the 13 archaeal genomes analyzed here, although it is entirely missing in *Pyrococcus horikoshii*, which requires at least one aromatic amino acid (tryptophan) for growth (Gonzalez et al. 1998). For the fifth enzymatic step, archaea possess a novel shikimate kinase (COG 1685) that has

Table 2. Phylogenetic distribution of the novel archaeal chorismate biosynthesis pathway genes (White 2004) compared with the first two steps of classical chorismate biosynthesis and transketolase. GenBank accession numbers are listed for each ortholog.

Genome	Novel chorismate pathway		Classical chorismate pathway		Transketolase (N- and C-terminal subunits) COG 3959/3958
	Novel aldolase COG 1830	Novel 3-dehydro- quininate synthase COG 1465	2-Dehydro-3-deoxy- phosphoheptonate aldolase COG 2876	3-Dehydroquininate synthase COG 0337	
<i>M. jannaschii</i>	NP_247374	NP_248244			NP_247665 NP_247663
<i>T. acidophilum</i>			NP_393727	NP_393765	NP_394090 NP_394089
<i>T. volcanium</i>			NP_111870	NP_111835	NP_111187 NP_111186
<i>P. abyssi</i>	NP_125781		NP_126138	NP_126139	NP_126136 NP_126137
<i>A. pernix</i>	NP_146901		NP_147338	NP_147337	NP_147340 NP_147339
<i>P. aerophilum</i>			NP_559644	NP_559640	NP_559642 NP_559641
<i>S. solfataricus</i>	NP_344531		NP_341851	NP_341852	NP_341848 NP_341849
<i>P. horikoshii</i>	NP_142099				
<i>M. kandleri</i>	NP_614692	NP_614691			
<i>M. thermoauto- trophicum</i>	NP_275722	NP_275723			
<i>A. fulgidus</i>	NP_068949	NP_069067			
<i>M. acetivorans</i>	NP_619450	NP_619451			
<i>Halobacterium</i>	NP_279409	NP_279411			

only recently been identified and characterized (Daugherty et al. 2001). Seven of the genomes are missing orthologs from COG 2876 and COG 0337, the first two biosynthetic steps responsible for the formation of DHQ from E4P and PEP (Table 2). Significantly, these are the same genomes (with the addition of *M. jannaschii*) that are also missing transketolase, and thus do not possess a recognizable means of making E4P.

The hypothesis that these archaeal species possess an alternate beginning phase of the aromatic amino acid biosynthesis pathway that does not involve E4P as a precursor (Tumbula et al. 1997) was recently confirmed (White 2004). In *M. jannaschii*, two genes were identified that catalyze the formation of DHQ from L-aspartate semialdehyde and 6-deoxy-5-ketohexulose, rather than from E4P and PEP. Although the novel aldolase identified in that study is widespread among archaea, including several species that have genes for the classical (E4P- and PEP-utilizing) chorismate synthesis pathway, the phylogenetic pattern of the novel DHQ synthase is highly complementary to the patterns observed both for the classical chorismate pathway genes and also for transketolase (Table 2).

It is notable that four of the 13 genomes (*Pyrococcus abyssi*,

*Aeropyrum pernix*, *Pyrobaculum aerophilum* and *Sulfolobus solfataricus*) have an incomplete NOPPP: orthologs of transketolase and R5P isomerase are present but not orthologs of epimerase or transaldolase (Table 1). Apparently, these four species use the RuMP pathway for Ru5P synthesis and thus do not require a complete NOPPP, but do require transketolase for generation of E4P from F6P and GAP. All four genomes contain orthologs for DAHP synthase, the E4P-utilizing step in aromatic amino acid biosynthesis. Furthermore, in three of these four genomes, the transketolase genes are located in operon-like clusters along with aromatic amino acid biosynthesis genes.

## Conclusions

The archaeal phylogeny of the PPP, when taken in isolation, is an enigmatic picture of partial pathways and seemingly inexplicable patterns of gene inheritance. When combined with genomic information about the RuMP pathway and the chorismate pathway of aromatic amino acid biosynthesis, however, a more coherent picture begins to emerge. In most archaea, the

ribose biosynthesis function of the OPPP is apparently carried out either by the enzymes of the NOPPP or by a reverse RuMP pathway. Those species that require E4P for aromatic amino acid biosynthesis produce it with transketolase, whereas those species with an alternate early phase of the chorismate pathway do not require transketolase. Finally, pentose synthesis in *Halobacterium* may occur via a modified OPPP, although the details of this putative pathway await experimental determination. The development of this picture of pentose and erythrose synthesis in archaea is an example of how the analysis of phylogenetic patterns within multiple genome sequences can yield valuable insights into diverse metabolic pathways. It will be interesting to see if the patterns observed in this study hold true as more complete genome sequences become available, and as the predicted biochemical functions of more genes are confirmed experimentally.

## References

- Choquet, C.G., J.C Richards and G.D. Sprott. 1994. Ribose synthesis in methanogenic archaea. *Arch. Microbiol.* 161:481–488.
- Daugherty, M., V. Vonstein, R. Overbeek and A. Osterman. 2001. Archaeal shikimate kinase, a new member of the GHMP-kinase family. *J. Bacteriol.* 183:292–300.
- Eisenreich, W., B. Schwarzkopf and A. Bacher. 1991. Biosynthesis of nucleotides, flavins, and deazaflavins in *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* 266:9622–9631.
- Ekiel, I., I.C.P. Smith and G.D. Sprott. 1983. Biosynthetic pathways in *Methanospirillum hungatei* as determined by <sup>13</sup>C nuclear magnetic resonance. *J. Bacteriol.* 156:316–326.
- Gonzalez, J.M., Y. Masuchi, F.T. Robb, J.W. Ammerman, D.L. Maeder, J. Yanagibayashi, J. Tamaoka and C. Kato. 1998. *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles* 2: 123–130.
- Ishikawa, K., I. Matsui, F. Payan, C. Cambillau, H. Ishida, Y. Kawabayashi, H. Kikuchi and A. Roussel. 2002. A hyperthermostable D-ribose-5-phosphate isomerase from *Pyrococcus horikoshii*: characterization and three-dimensional structure. *Structure* 10: 877–886.
- Martinez-Cruz, L.A., M.K. Dreyer, D.C. Boisvert, H. Yokota, M.L. Martinez-Chanter, R. Kim and S.H. Kim. 2002. Crystal structure of MJ1247 protein from *M. jannaschii* at 2.0 Å resolution infers a molecular function of 3-hexulose-6-phosphate isomerase. *Structure* 10:195–204.
- Ogushi, S., M. Ando and D. Tsuru. 1984. Formaldehyde dehydrogenase from *Pseudomonas putida*: a zinc metalloenzyme. *J. Biochem.* 96:1587–1591.
- Reizer, J., A. Reizer and M.H. Saier, Jr. 1997. Is the ribulose monophosphate pathway widely distributed in bacteria? *Microbiology* 143:2519–2520.
- Roy, R., M. Swarnalatha, G.J. Schut, D.M. Dunn, R. Weiss and M.W.W. Adams. 1999. Purification and characterization of the tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*: the third of a putative five-member tungstoenzyme family. *J. Bacteriol.* 181: 1171–1180.
- Selkov, E., N. Maltsev, G.J. Olsen, R. Overbeek and W.B. Whitman. 1997. A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. *Gene* 197:GC11–26.
- Snel, B., G. Lehmann, P. Bork and M.A. Huynen. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 28:3442–3444.
- Soderberg, T. and R.C. Alver. 2004. Transaldolase of *Methanocaldococcus jannaschii*. *Archaea* 1:255–262.
- Tanaka, N., Y. Kusakabe, K. Ito, T. Yoshimoto and K.T. Nakamura. 2002. Crystal structure of formaldehyde dehydrogenase from *Pseudomonas putida*: the structural origin of the tightly bound cofactor in nicotinoprotein dehydrogenases. *J. Mol. Biol.* 324: 519–533.
- Tatusov, R.L., E.V. Koonin and D.J. Lipman. 1997. A genomic perspective on protein families. *Science* 278:631–637.
- Tatusov, R.L., D.A. Natale, I.V. Garkavtsev et al. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29:22–28.
- Tumbula, D.L., Q. Teng, M.G. Bartlett and W.B. Whitman. 1997. Ribose biosynthesis and evidence for an alternative first step in the common aromatic amino acid pathway in *Methanococcus maripaludis*. *J. Bacteriol.* 179:6010–6013.
- von Mering, C., M. Huynen, D. Jaeggi, S. Schmidt, P. Bork and B. Snel. 2003. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31:258–261.
- White, R.H. 2004. L-Aspartate semialdehyde and a 6-deoxy-5-ketohexose-1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*. *Biochemistry* 43:7618–7627.
- Vorholt, J.A., C.J. Marx, M.E. Lidstrom and R.K. Thauer. 2000. Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J. Bacteriol.* 182: 6645–6650.
- Yasueda, H., Y. Kawahara and S. Sugimoto. 1999. *Bacillus subtilis* *yckG* and *yckF* encode two key enzymes of the ribulose monophosphate pathway used by methylotrophs, and *yckH* is required for their expression. *J. Bacteriol.* 181:7154–7160.
- Yu, J.P., J. Ladapo and W.B. Whitman. 1994. Pathway of glycogen metabolism in *Methanococcus maripaludis*. *J. Bacteriol.* 176: 325–332.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

