Mini review

P450s in Microbial Sterol Biosynthesis and Drug Targets

Colin John Jackson, David Christopher Lamb, Andrew Graham Samuel Warrilow, Josie Elizabeth Parker, Nadja Rodrigues de Melo, Diane Elizabeth Kelly and Steven Lewis Kelly*

Institute of Life Science, School of Medicine, Swansea University, Singleton Park, Swansea SA2 8PP, UK.; Tel: +44 1792 602207; Fax: +44 1792 513430

** Corresponding author: E-mail: s.l.kelly@swansea.ac.uk*

Received: 17-10-2007

Abstract

The diversity of cytochromes P450s (CYP) in species from various Kingdoms of Life is extensive, sometimes comprising more than 1% of coding genes. Some bacteria do not contain *CYP* genes while in others, such as the actinomycetes, they are numerous. Microbial eukaryotes usually contain at least CYP51, the ancestral activity needed for sterol 14-demethylation. This CYP exists in some bacteria such as *Methylococcus capsulatus* and sterol biosynthesis in this methane utilising bacterium is discussed together with CYP51-like proteins of bacteria that have other endogenous functions. In fungi CYP51 is a drug target for azole compounds and here evolution is also occurring to drug resistant forms. Evolutionary aspects of bacterial sterol biosynthesis are discussed.

1. Introduction

The involvement of the cytochromes P450 (CYP) superfamily in eukaryotic biological processes and toxicology has been widely recognised, but much less well understood is the importance and diversity of this family of haem-thiolate proteins in microbes. By the late 1990's, relatively few CYPs had been identified or characterised in prokaryotes. This situation has undergone a dramatic reversal with the advent of wide-scale whole genome sequencing of bacteria and other microorganisms. From a position of relative ignorance regarding microbial CYPs, we are now beginning to fully appreciate the true level of diversity of these enzymes within prokaryotes and simple eukaryotes. Most bacteria, such as *Escherichia coli*, have no *CYP*s present in their genome, whilst at the other end of the spectrum, members of the genus *Mycobacterium* have forty or more *CYP*s representing 1% of coding genes. A parallel situation is found in fungi, where a simple budding yeast such as *Saccharomyces cerevisiae* requires only 3 CYPs,[1] but the saprophytic wood-rot fungus *Phanaerochaete chrysosporium* has in excess of 100 *CYP*s.[2] Other unicellular eukaryotes, such as the primitive protozoan *Giardia lamblia*, appear to have no *CYP*s present at all.

The scale of microbial CYP biodiversity has stimulated enormous interest in applying functional genomics and proteomics to investigate the biological roles of these novel proteins. Such studies have to address two primary issues. First, to understand why many microbes can exist with no or very few *CYP*s,[1] whilst others have large and diverse suites of these enzymes. Secondly, to develop experimental approaches for investigating the often completely novel P450 families to which many of these microbial *CYP*s belong.

2. Abundant Cytochromes P450 in *Streptomyces* spp. and the Formation of Secondary Metabolites

Assigning form to function has been most successful with the bacterial CYPs in the genus *Streptomyces*. These organisms are collectively responsible for producing approximately two-thirds of microbially-derived antibiotics,[3] and a wide array of pharmacologically significant secondary metabolites.[4] In addition, they are able to catalyse numerous biotransformations of compounds of industrial and environmental importance, such as alkaloids, coumarins and other complex xenobiotics.[5] Clues to individual streptomycete CYP function have often been revealed by the genes that are associated with them; for example, gene clusters for macrolide antibiotics often

contain one or more *CYP* gene(s) which are involved in oxidative tailoring. Other *CYP*-containing operons, show strong conservation and gene organisation across members of the genus.[6] However, for the vast majority of microbial CYPs, no useful bioinformatic indicators exist, and they are effectively "orphans" with no known function. The situation is further complicated by the fact that not only do many bacterial CYPs belong to new families, but that within these families there is often evidence of gene duplication and diversification, as well as gene loss. In attempting to understand how CYP diversity has evolved, our group has been investigating the properties of ancestral CYPs: – CYPs that are present in different Kingdoms of Life. To date, only one CYP has been identified – sterol 14α demethylase (CYP51) – which is present in bacteria, lower eukaryotes, plants, fungi and animals.

3. CYP51 has a Central Role in Sterol Biosynthesis

CYP51 was one of the first CYPs to be fully characterised at both the genetic and biochemical level.[7] It was shown to be responsible for catalysing three successive monooxygenation reactions in a single enzymatic reaction step in a biosynthetic pathway leading to the synthesis of sterols in eukaryotic organisms. The presence of CYP51 in plants, animals and fungi led to the hypothesis that this was an ancestral CYP; an observation given weight by the later discovery of CYP51 in bacteria.[8] Some animals, such as insects, lack the ability to synthesize sterols and have no CYP51. However, the presence of CYP51 in wide-ranging life forms suggests that in this phylum, whose members rely on dietary assimilation of exogenous sterols as do nematodes, the entire sterol biosynthetic pathway may have been lost in some strands of animal evolution.

The discovery of a prokaryotic CYP51 in *Mycobacterium tuberculosis* (MTB) was both unexpected and exciting because of the pharmacological potential.[8] The structure of MTB CYP51 was substantially different to eukaryotic isoforms, but it was capable of sterol 14α demethylase activity. However, no sterol biosynthetic products were observed, and a role for this enzyme within the organism has yet to be determined. Interestingly, another CYP51-like protein was revealed in *Streptomyces coelicolor*,[9] and this too appears to be an orphan of the sterol biosynthetic pathway or else has arisen by convergent evolution. The *S. coelicolor CYP51* homologue is located in a different operon structure to that conserved in mycobacteria,[10] and appears functionally linked to a terpene synthase that points to a role in secondary metabolism (and named CYP170). The CYP51 protein is the target for the azole class of antifungals, and it was intriguing to discover whether these compounds might also be active against CYP51 in MTB. Experiments to date show that whilst current azoles do exert a toxic effect on mycobacteria, they are unlikely to be sufficiently active to be therapeutically useful.[11]

4. A Novel CYP51 Isoform is Involved in Sterol Biosynthesis in the Methanotrophic Bacterium, *Methylococcus Capsulatus*

The second bacterial CYP51 to be discovered came from the aerobic methanotroph, *Methylococcus capsulatus*. Production of sterols in this organism had been well characterised previously using classical biochemical analyses, including GC/MS and NMR. The discovery of 4-dimethyl, 4-methyl and 4-desmethyl sterols gave a strong indication that a functional CYP51 enzyme was indeed present.[12] However, several decades elapsed before these initial biochemical observations could be confirmed from genomic data. Our group was able to identify a sterol 14α demethylase ORF in *M. capsulatus* encoding a protein with close homology to that of MTB. Interestingly, this *M. capsulatus* CYP51 consisted of a CYP51 domain fused to a ferredoxin domain by an alanine-rich linker or hinge region (Figure 1). As such, this novel protein, termed MCCYP51FX, represented a completely new class of cytochromes P450, related to the Class I CYPs.[13]

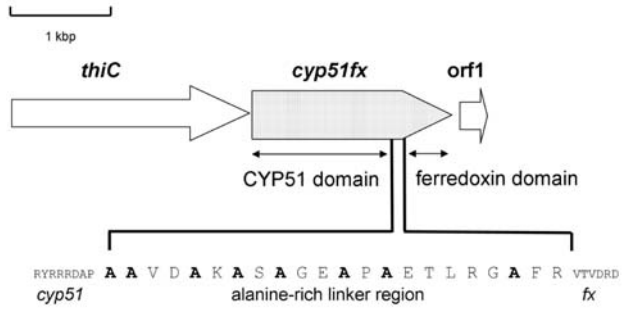

Figure 1: The alanine-rich linker region of MCCYP51FX. This linker may function as a flexible hinge, enabling interaction between the functional domains of the P450 and ferredoxin components of the protein. Location of the two ORFs (*thiC* and *orf1*) flanking *mc-cyp51fx* are indicated.

The 1656bp gene for MCCYP51FX was cloned and expressed in *E. coli*, and a soluble, active protein of 62.4 kDa was purified by Ni^{2+} nitrilotriacetic acid (NTA)-affinity chromatography. This protein was shown to have a typical CYP-reduced carbon monoxide difference spectrum, with a spectral maximum located at 448 nm. Quantification of the iron content showed that there were four atoms of iron per molecule of MCCYP51FX, confirming a 3Fe-4S ferredoxin component to the fusion protein, with

one atom of iron associated with the CYP haem. Enzymatic reconstitution experiments with spinach ferredoxin reductase and lanosterol as substrate demonstrated the production of a 14α demethylated reaction product (silylated 14α demethylated 4α-methyl-5α-ergosta-8,14,24(28)-trien-3β-ol).[13]

A postscript to the discovery and characterisation of this bacterial CYP51 was the observation that downstream and adjacent to the *CYP51* of MTB was an ORF encoding a ferredoxin-like protein (locus tag Rv0763c) with similarity to the fused ferredoxin domain of MCCYP51FX. In addition, the codon usage patterns used in MTB and *M. capsulatus* appear almost identical. These features suggest that the MCCYP51FX gene may have been horizontally transferred to an ancestor of MTB, with subsequent loss of the linker/hinge region. Attractive though this hypothesis may be, it still provides little insight into the possible role of CYP51 in mycobacteria given that it is not involved in sterol production.

5. Characterization of the Post-squalene Pathway of Sterol Biosynthesis in *M. Capsulatus*

The discovery and characterisation of CYP51 in *M. capsulatus* provided an impetus to search for other genes of the sterol biosynthetic pathway in this organism. Production of sterols is generally associated with eukaryotic organisms, with different end products in each kingdom of life: cholesterol in animals, ergosterol in fungi and phytosterols in plants. In *M. capsulatus*, a mixture of sterol intermediates were known to be present, including 4α-methyl-5α-cholest-8(14)-en-3β-ol, 4,4-dimethyl-5α -cholest-8(14)-en-3β-ol, 4α-methyl-5α-cholest-8(14)24-dien-3β-ol, and 4,4-dimethyl-5α-cholest-8(14),24-dien-3β-ol).[12,14] None of the eukaryote terminal sterols were identified, revealing that only a partial sterol biosynthetic pathway was likely to be present. Sterol biosynthesis originates from squalene, which is converted to 2,3 oxidosqualene and then cyclised to form the first sterol molecule, lanosterol. The presence of these intermediates required that, at the very least, genes encoding a squalene monooxygenase and an oxidosqualene cyclase had to be present alongside the already identified lanosterol 14α demethylase. Homology searches of the flanking regions upstream and downstream of *MCCYP51FX* revealed no evidence of predicted sterol pathway members. This confirmed that whatever genes of the sterol pathway were present, they were not organised as a contiguous operon or functional gene cluster.

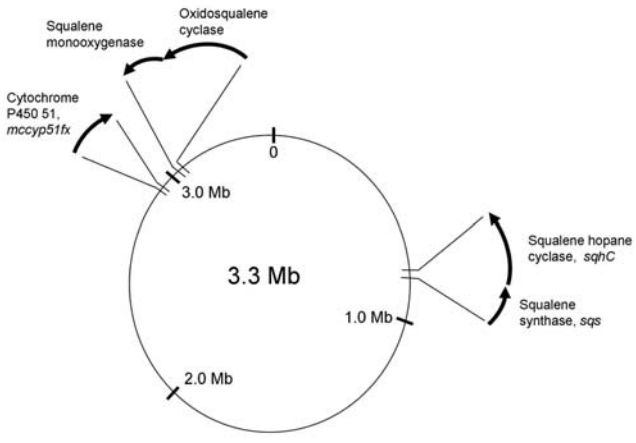

Figure 2: Location of sterol-associated genes on the 3.3 Mb chromosome of *Methylococcus capsulatus* str.Bath.

However, at a locus some 170kb distant from *mccyp51fx* (Figure 2), a pair of overlapping ORFs were discovered encoding predicted proteins with strong homologies to oxidosqualene cyclase (alternative name:ERG7) and squalene monooxygenase (alternative name:ERG1) (Figure 3A/B).[15] The first gene of the pair to be transcribed is the cyclase (locus tag MCA2873), which was predicted from the genome annotation (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genome&Cmd=Retrieve&dopt=Protein+Table&list_uids=507) to be 670 amino acids in length. However, close inspection revealed the absence of a candidate Shine-Dalgarno (S-D) sequence upstream of the initiation codon. In addition, codon usage analysis indicated that a sequence of twelve amino acids upstream of the assigned initiation codon were predicted to be part of the expressed protein. We therefore designated a new initiation codon for this protein, upstream of which were sequences encoding both a potential S-D and riboprotein S1 binding sites. The *M. capsulatus* oxidosqualene cyclase was therefore 682 amino acids in length, encoded by an ORF of 2049 bp. The end of the cyclase gene had a 4 bp overlap with the downstream gene, the predicted squalene monooxygenase (locus tag MCA2872). This ORF was 1350 bp in length, encoding a predicted protein of 449 amino acids.

The identification of a separate squalene hopane cyclase (see below), and the adjacent, overlapping location of the predicted squalene monooxygenase and oxidosqualene cyclase gave a strong indication that the latter two genes were directly involved in the biosynthesis of sterol intermediates. To demonstrate that this was in fact the case, a complete molecular characterization of the first three postsqualene steps of the sterol biosynthetic pathway in *M. capsulatus* was carried out.[15] The monooxygenase and cyclase genes were cloned and expressed in *E. coli*, and the proteins purified by Ni^{2+} – NTA affinity chromatography. Epoxidation of squalene by the monooxygenase was demonstrated by reconstituting the enzyme activity following the addition of spinach ferredoxin and ferredoxin reductase as electron donor partners, with NADPH as a source of reducing equivalent. Similarly, catalytic acti-

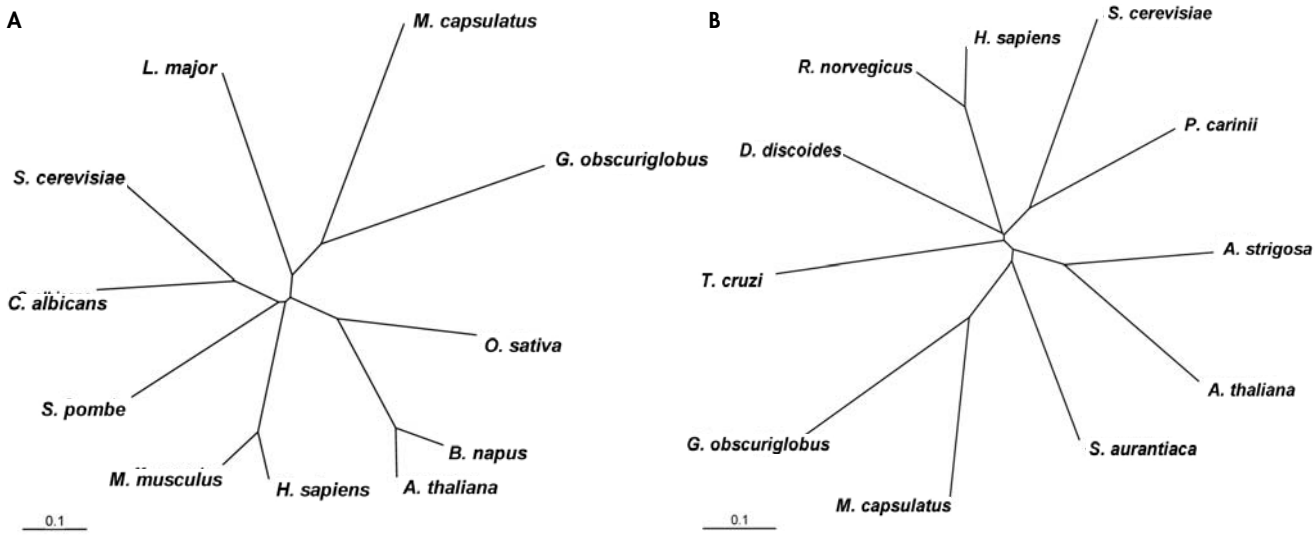

Figure3A/3B: Neighbour-Joining trees of selected squalene monooxygenase (3A) and oxidosqualene cyclase (3B) amino acid sequences from five Kingdoms of Life. **Bacteria** – *Methylococcus capsulatus; Gemmata obscuriglobus; Stigmatella aurantiaca*. **Plantae** – Land cress (*Arabidopsis thaliana*); Radish (*Brassica napus*), Rice (*Oryza sativa*), Black oat (*Avena strigosa*). **Fungi** – Bakers yeast (*Saccharomyces cerevisiae*), Fission yeast, (*Schizosaccharomyces pombe*), *Pneumocystis carinii*, *Candida albicans*. **Animalia** – Human (*Homo sapiens*), Mouse (*Mus musculus*), Rat (*Rattus norvegicus*), **Protista** – Amoeba (*Dictyostelium discoideum*), *Leishmania major*, *Trypanosoma cruzi.*

vity of oxidosqualene cyclase was demonstrated using oxidosqualene as substrate. Finally, in the MCCYP51FX reconstitution experiments described above, with spinach ferredoxin reductase and lanosterol as substrate, the production of a 14α demethylated reaction product was shown.

6. The *M. Capsulatus* Genome Encodes Orthologues for Both Hopanoid and Sterol Biosynthesis

Many bacteria are known to synthesise hopanoids as sterol surrogates. The first post-squalene step in hopanoid biosynthesis requires squalene hopane cyclase, an enzyme which shows several similar structural motifs to oxidosqualene cyclase. The synthesis of hopanols has been reported in *M. capsulatus,*[16] and the genome sequence revealed an ORF encoding a predicted squalene hopane cyclase (locus tag MCA0812) adjacent to the gene for squalene synthase (locus tag MCA0813) (Figure 2). This organism therefore synthesizes both sterols and hopanols, although the biological significance of this dual capability in terms of membrane architecture or other processes remains unclear. The protein homology between *M. capsulatus* squalene hopane cyclase and oxidosqualene cyclase is low (26% identity, 41% similarity), suggesting that if these genes are in fact paralogues, then substantial divergence has occurred during the course of their evolutionary specialisation.

7. CYP51 and the Evolution of the Sterol Pathway

This series of experiments was able to demonstrate conclusively that in *M. capsulatus*, the early stages of prokaryote sterol biosynthesis parallel those seen in fungi and many animals, and that a cytochrome P450 orthologue of CYP51 was an essential component of this pathway. Sterol biosynthesis in prokaryotes remains a relatively rare phenomenon, and has been most well characterised in *M. capsulatus,*[12–15] and the planctomycete *Gemmata obscuriglobus.*[17] The latter has the most abbreviated sterol pathway known, terminating at lanosterol and parkeol. Recently, several species in eleven different genera of the fruiting, gliding bacteria, the *Myxococcales*, have also been shown to synthesis a variety of sterol intermediates.[18]

The existence of sterols as molecular fossils from a very early period of the Earth's history (circa 2.7 billion years ago) confirms that evolution of the sterol pathway was ancient.[19] Currently-known sterol-producing bacterial taxa occupy widely different branches of the prokaryote evolutionary tree. The scenario that the sterol biosynthetic pathway originally arose in bacteria, was then deleted during evolution from the vast majority of prokaryotes and yet was retained in a very restricted and diverse subset of microorganisms, has been questioned.[17] An alternative explanation is that there was limited lateral genetic exchange of sterol genes from ancestral eukaryotes into prokaryotes, (or *vice versa*), which were then retained only in certain specialised groups of bacteria, but almost all eukaryotic microbes. The lateral transfer of genes from eu-

karyotes into bacteria such as MTB has been inferred from genomic profiling.[20] In addition, biosynthetic pathway genes might have been taken up by microbes from plant material as a consequence of their saprophytic lifestyle, such as the cellulolytic pathway of the myxobacterium *Sorangium cellulosum*.[21] Indeed, a similar plant-derived genetic mechanism has been proposed to explain the presence of *CYP51* in the progenitor of *M. capsulatus*/MTB.[22]

A better understanding of the evolution of sterol pathways requires far greater knowledge of the biological roles for sterols in those bacteria in which they occur, and the coincident function of hopanoids alongside these sterols. On the available evidence, it remains possible that the prokaryotic ancestor of eukaryotes was also sterol producing. The answers to such complex evolutionary and functional aspects of sterol biosynthesis in prokaryotes may be revealed as more whole genome data from sterol producing bacteria become available.

8. Acknowledgements

This work was supported in part by National Institutes of Health (USA), BBSRC (UK) and the EU Framework Six projects STEROLTALK and EURESFUN.

9. References

1. A. Goffeau, B.G. Barrell, H. Bussey et al., *Science* **1996**, *274*, 563–567.
2. D. Martinez, L.F. Larrondo, N. Putnam et al., *Nature Biotech*. **2004**, *22*, 695–700.
3. J. F. Andersen, C. R. Hutchinson, *J. Bacteriol*. **1992,** *174*, 725–735.
4. N. Serizawa, T. Matsuoka, *Biochem. Biophys. Acta* **1991,** *1084,* 35–40.
5. M. Taylor, D.C. Lamb, R. Cannell, M. Dawson, S. L. Kelly, *Biochem. Biophys. Res. Commun.* **1999**, *263*, 838–842.
6. S. D. Bentley, K. F. Chater, A. M. Cerdeno-Tarraga et al., *Nature* **2002,** *417*, 141–147.
7. Y. Yoshida, Y. Aoyama, *J. Biol Chem.***1984**, *259*, 1655–1660.
8. A. Bellamine, A. T. Mangla, W. D. Nes, M.R. Waterman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8937–8942.
9. D. C. Lamb, K. Fowler, T. Kieser, N. J. Manning, L. M. Podust, M. R. Waterman, D. E. Kelly, S. L. Kelly, *Biochem. J*. **2002**, *364,* 555–562.
10. C. J. Jackson, D. C. Lamb, T. H. Marczylo, J. E. Parker, N. L. Manning, D. E. Kelly, S. L. Kelly, *Biochem. Biophys. Res. Commun.* **2003,** *301,* 558–563.
11. C. J. Jackson, D. C. Lamb, D. E. Kelly, S. L. Kelly, *FEMS Lett*. **2000**, *192*, 159–162.
12. C. W. Bird, J. M. Lynch, S. J. Pirt, W. W. Reid, C. J. W. Brooks, B. S. Middleditch, *Nature* **1971**, *230*, 473–474.
13. C. J. Jackson, D. C. Lamb, T. H. Marczylo, A. G. S. Warrilow, N. J. Manning, D. J. Lowe, D. E. Kelly, S. L. Kelly, *J. Biol Chem.* **2002**, *277*, 46959–46965.
14. P. Bouvier, M. Rohmer, P. Benveniste, G. Ourisson, *Biochem. J.* **1976**, *159,* 267–271.
15. D. C. Lamb, C. J. Jackson, A. G. S. Warrilow, N. J. Manning, D. E. Kelly, S. L. Kelly, *Mol. Biol. Evol.* **2007,** *24,* 1714–1721.
16. S. Neunlist, M. Rohmer, *Biochem. J.* **1985**, *231,* 635–639.
17. A. Pearson, M. Budin, J. J. Brocks, *Proc. Natl. Acad. Sci. USA* **2003***, 100,* 15352–15357.
18. H. B. Bode, B. Zeggel, B. Silakowski, S. C. Wenzel, H. Reichenbach, R. Müller, *Mol. Micro.* **2003** *47*, 471–481.
19. J. J. Brocks, G. A. Logan, R. Buick, R. E. Summons, *Science* **1999,** *285*, 1033–1036.
20. J. Gamieldien, A. Ptitsyn, W. Hide, *Trends in Genet*. **2002**, *18*, 5–8.
21. H. B. Bode, R. Müller, *Plant Physio*. **2003,** *132,* 1153–1161.
22. T. Rezen, N. Debeljak, D. Kordis, D. Rozman, *J. Mol. Evol.* **2004,** *59*, 51–58.

Povzetek

Raznolikost citokromov P450 (CYP) med vrstami iz različnih kraljestev življenja je velika in včasih predstavlja več kot 1 % kodirajočih genov. Nekatere bakterije ne vsebujejo genov *CYP*, medtem ko druge, na primer aktinomicete, vsebujejo mnoge. Mikrobni evkarionti povečini vsebujejo vsaj CYP51, starodavno aktivnost, ki je potrebna za 14-demetilacijo sterolov. Ta CYP obstaja v nekaterih bakterijah, na primer v *Methylococcus capsulatus.* Sinteza sterolov v tej bakteriji, ki izkorišča metan, je opisana skupaj z drugimi CYP51-podobnimi proteini, ki imajo drugačne nedogene funkcije. Pri glivah je CYP51 tarča za azolne spojine, evolucija pa je privedla do na zdravila odpornih sevov. Članek opisuje tudi evolucijske aspekte sinteze sterolov pri bakterijah.