Description of the dataset m52orfs.txt

The dataset *m52orfs.txt* contains all predicted genes of bacterium Escherichia coli together with some of their calculated properties. Each row represents one gene. Columns contain the following information:

Col 1: Gene name
Col 3: Alternate gene name
Col 4: Start coordinate of the gene in the genome (first nucleotide)
Col 5: End coordinate of the gene in the genome (last nucleotide)
Col 6: Direction of the gene on the chromosome
Col 7: Genetic location of the gene in centisomes
Col 8: First codon (amino-acid coding triplet) of the gene
Col 9: Last codon (amino-acid coding triplet) of the gene
Col 10: Length of the coded protein
Col 11: Predicted molecular weight of the coded protein (kD)
Col 12: Predicted isoelectric point (pI) of the coded protein
This is pH value at which given protein is neutral, without positive or negative charges. Molecules with low pI are negatively charged (acidic) in normal cells and molecules with high pI are positively charged (basic) in normal cells.
Col 13: CAI - Codon Adaptation Index. This is calculated from gene sequence and is known to be correlated with expression level of the gene (the number of protein molecules produced per second from this gene)
Col 14: Known or predicted cellular function (functional class) of given gene