

IBS2 Assignment – 1

Akash Krishna M V

Bl.sc.u4aie24203@bl.students.amrita.edu

Department of Computer Science,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham Bengaluru, India

Question 1: Tool for Identifying NRP Regions

A simple and popular tool for finding nonribosomal peptide (NRP) regions in a genome is **antiSMASH**. It helps scientists locate special groups of genes called NRPS clusters, which are responsible for making important natural products like antibiotics and other bioactive compounds. To use it, you upload a genome sequence file (in FASTA or GenBank format) to the antiSMASH website or run it on your computer. The tool scans the DNA and automatically finds gene clusters related to NRPs, shows their positions, and predicts what kind of molecules they might produce. It also gives easy-to-read diagrams and compares the clusters with known ones, making it useful for quickly identifying potential NRP-producing regions for further research.

Question 2: Brute Force Algorithm for Parent Mass

The brute force method generates all possible peptide combinations and checks whether their total mass equals the given parent mass. If yes, then that combination is added to the result.

Python Code:

```
def brute_force(peptide, current_mass):  
  
    global count  
    if current_mass == parent_mass:  
        result.append(peptide)  
        count += 1  
        return  
    if current_mass > parent_mass:  
        return  
  
    for aa, mass in amino_acid_mass.items():  
        brute_force(peptide + aa, current_mass +  
mass)  
  
amino_acid_mass = {  
    'G':57, 'A':71, 'S':87, 'P':97, 'V':99,  
    'T':101, 'C':103, 'I':113, 'L':113, 'N':114,  
    'D':115, 'K':128, 'Q':128, 'E':129, 'M':131,  
    'H':137, 'F':147, 'R':156, 'Y':163, 'W':186  
}  
  
parent_mass = 457  
global count
```

Peptides with mass 457

['GGGGGD', 'GGGGGAT', 'GGGGGTA', 'GGGGGDG', 'GGGGAGT', 'GGGGAAS', 'GGGGASA', 'GGGGATG', 'GGGGSAA', 'GGGGTGA', 'GGGGTAG', 'GGGGTK', 'GGGGTQ', 'GGGGND', 'GGGGDGG', 'GGGGDN', 'GGGGKT', 'GGGGQT', 'GGGAGGT', 'GGGAGAS', 'GGGAGSA', 'GGGAGTG', 'GGGAAGS', 'GGGAASG', 'GGGASGA', 'GGGASAG', 'GGGASK', 'GGGASQ', 'GGGATGG', 'GGGATN', 'GGGANT', 'GGGAKS', 'GGGAQS', 'GGGSGAA', 'GGGSAGA', 'GGGSAAG', 'GGGSAK', 'GGGSAQ', 'GGGSKA', 'GGGSQA', 'GGGTGGA', 'GGGTGAG', 'GGGTGK', 'GGGTGQ', 'GGGTAGG', 'GGGTAN', 'GGGTNA', 'GGGTKG', 'GGGTQG', 'GGGNGD', 'GGGNAT', 'GGGNTA', 'GGGNDG', 'GGGDGGG', 'GGGDGN', 'GGGDNG', 'GGGKGT', 'GGGKAS', 'GGGKSA', 'GGGKTG', 'GGGQGT', 'GGGQAS', 'GGGQSA', 'GGGQTG', 'GGAGGGT', 'GGAGGAS', 'GGAGGSA', 'GGAGGTG', 'GGAGAGS', 'GGAGASG', 'GGAGSGA', 'GGAGSAG', 'GGAGSK', 'GGAGSQ', 'GGAGTGG', 'GGAGTN', 'GGAGNT', 'GGAGKS', 'GGAGQS', 'GGAAGGS', 'GGAAGSG', 'GGAASGG', 'GGAASN', 'GGAANS', 'GGASGGA', 'GGASGAG', 'GGASGK', 'GGASGQ', 'GGASAGG', 'GGASAN', 'GGASNA', 'GGASKG', 'GGASQG', 'GGATGGG', 'GGATGN', 'GGATNG', 'GGANGT', 'GGANAS', 'GGANSA', 'GGANTG', 'GGAKGS', 'GGAKSG', 'GGAQGS', 'GGAQSG', 'GGSGGAA', 'GGSAGGA', 'GGSAGAAG', 'GGSAGK', 'GGSAGQ', 'GGSQGA', 'GGSQAG', 'GGSQAG', 'GGSQK', 'GGSQQ', 'GGPVF', 'GGPDM', 'GGPMD', 'GGPFV', 'GGVPF', 'GGVIM', 'GGVLM', 'GGVDE', 'GGVED', 'GGVMI', 'GGVML', 'GGVFP', 'GGTGGGA', 'GGTGGAG', 'GGTGGK', 'GGTGGQ', 'GGTGAGG', 'GGTGAN', 'GGTGNA', 'GGTGKG', 'GGTGQG', 'GGTAGGG', 'GGTAGN', 'GGTANG',

...

['FGGPV', 'FGGVP', 'FGPGV', 'FGPVG', 'FGPR', 'FGVGP', 'FGVPG', 'FGRP', 'FAAAP', 'FAAPA', 'FAPAA', 'FPGGV', 'FPGVG', 'FPGR', 'FPAAA', 'FPVGG', 'FPVN', 'FPNV', 'FPRG', 'FVGPG', 'FVPGG', 'FVPN', 'FVNP', 'FNPV', 'FNVP', 'FFY', 'FRGP', 'FRPG', 'FYF', 'RGPF', 'RGIM', 'RGLM', 'RGDE', 'RGED', 'RGM', 'RGML', 'RGFP', 'RAVM', 'RATE', 'RADD', 'RAET', 'RAMV', 'RSVD', 'RSTI', 'RSTL', 'RSIT', 'RSLT', 'RSDV', 'RPGF', 'RPTC', 'RPCT', 'RPF', 'RVAM', 'RVSD', 'RVVC', 'RVTT', 'RVCV', 'RVDS', 'RVMA', 'RTAE', 'RTSI', 'RTSL', 'RTPC', 'RTVT', 'RTTV', 'RTCP', 'RTIS', 'RTLS', 'RTEA', 'RCPT', 'RCVV', 'RCTP', 'RIGM', 'RIST', 'RITS', 'RIMG', 'RLGM', 'RLST', 'RLTS', 'RLMG', 'RDGE', 'RDAD', 'RDSV', 'RDVS', 'RDDA', 'RDEG', 'RDW', 'REGD', 'REAT', 'RETA', 'REDG', 'RMGI', 'RMGL', 'RMAV', 'RMVA', 'RMIG', 'RMLG', 'RFGP', 'RFPG', 'RWD', 'YMY', 'YFF', 'YYM', 'WGV', 'WGTI', 'WGT', 'WGLT', 'WGL', 'WGDV', 'WAAE', 'WASI', 'WASL', 'WAPC', 'WAVT', 'WATV', 'WACP', 'WAIS', 'WALS', 'WAEA', 'WSAI', 'WSAL', 'WSSP', 'WSPS', 'WSIA', 'WSLA', 'WPAC', 'WPSS', 'WPCA', 'WVGD', 'WVAT', 'WVTA', 'WVDG', 'WTGI', 'WTGL', 'WTAV', 'WTVA', 'WTIG', 'WTLG', 'WCAP', 'WCPA', 'WIGT', 'WIAS', 'WISA', 'WITG', 'WLGT', 'WLAS', 'WLSA', 'WLTG', 'WDGV', 'WDVG', 'WDR', 'WEAA', 'WRD']

Total peptides found: 6508