Консервативный участок длиной 20 аминокислот:
**AGHDTTSSSSGGAIIGLSRN**
**AGHETTSGLLSFALYFLVKN**
**AGFEASVSLIGIGTYLLLTH**
**GGLDTVVNFLSFSMEFLAKS**
**AGNATMVNMIALGVATLAQH**

Профиль profile45.prf построен по матрице blosum45.cmp:
**./pfmake -m ms_file.msf blosum45.cmp > profile45.prf**

По базе UniProtKB/SwissProt найдены белки, соответствующие профилю:
**./pfsearch -f -C 3.0 profile45.prf uniprot_sprot.fasta | sort -nr > scoresBaseD.txt**

Найдены все цитохромы P450 в базе:
**grep -i 'GN=cyp' uniprot_sprot.fasta > BaseCytochromeP450.txt**
**wc BaseCytochromeP450.txt**
**1020** 10248 99304 BaseCytochromeP450.txt

Проведен анализ полученного профиля (вычислены sensitivity, specificity и PPV):
TP = все P450 в найденных
TN = все не P450 в базе - все не P450 в найденных
FP = все не P450 в найденных
FN = все P450 в базе - все P450 в найденных

**grep -i 'GN=cyp' scoresBaseD.txt > CytochromeP450.txt**
**wc CytochromeP450.txt**
**427** 6674 56528 CytochromeP450.txt
TP = 427

**grep '>' uniprot_sprot.fasta > BaseAll.txt**
**wc BaseAll.txt**
**542782** 7014762 63438950 BaseAll.txt
**wc scoresBaseD.txt**
**469** 7415 62687 scoresBaseD.txt
TN = 542782 - 1020 - (469 - 427) = 541720

FP = 469 - 427 = 42

FN = 1020 - 427 = 593

**sensitivity** = TP / (TP + FN) = 427.0 / (427 + 593) = **0.42**

**specificity** = TN / (TN + FP) = 541720.0 / (541720 + 42) = **1**
**PPV** = TP / (TP + FP) = 427.0 / (427 + 42) = **0.91**

Еще одна итерация, в качестве входных данных найденные последовательности:

**./pfsearch -f -s -C 3.0 profile45.prf uniprot_sprot.fasta > scoresBaseMANS.txt**

Профиль profile1.prf построен по матрице blosum45.cmp:

**./pfmake -m scoresBaseMANS.txt blosum45.cmp > profile1.prf**

По базе UniProtKB/SwissProt найдены белки, соответствующие профилю:

**./pfsearch -f -C 3.0 profile1.prf uniprot_sprot.fasta > scoresBase1.txt**

Проведен анализ полученного профиля (вычислены sensitivity, specificity и PPV):

**grep -i 'GN=cyp' scoresBase1.txt > CytochromeP4501.txt**
**wc CytochromeP4501.txt**
**751** 11606 99064 CytochromeP4501.txt
TP = 751

**wc scoresBase1.txt**
**909** 14548 123717 scoresBase1.txt
TN = 542782 - 1020 - (909 - 751) = 541604

FP = 909 - 751 = 158

FN = 1020 - 751 = 269

**sensitivity** = TP / (TP + FN) = 751.0 / (751 + 269) = **0.74**
**specificity** = TN / (TN + FP) = 541604.0 / (541604 + 158) = **1**
**PPV** = TP / (TP + FP) = 751.0 / (751 + 158) = **0.83**

После второй итерации результаты улучшились, sensitivity заметно увеличилось:

**./pfsearch -f -s -C 3.0 profile1.prf uniprot_sprot.fasta > scoresBase1MA.txt**

С помощью PRATT для изначальных цитохромов был найден паттерн , также с помощью ScanProsite он был прогнан по базе UniProtKB/SwissProt:

**Hits for**

**USERPAT1{F-G-x-G-x-[HR]-x-C-[ILM]-[AG]-[EQR]-x-[FL]-A-x(2)-E-x(4)-[FL]} motif on all UniProtKB/Swiss-Prot**

В результате мы как и в предыдущем пункте получили Cytochrome P450. Можно сделать вывод, что последовательность является достаточно консервативной и встречается в схожем виде в различных организмах. Поиск по профилю, благодаря использованию матриц, является чувствительнее и выдает большее количество белков.