

Chapter 9 SELF-TEST Protein families and databases

1. In PROSITE, the term PATTERN indicates that the entry describes:
 - A. a block
 - B. a profile
 - C. a regular expression
 - D. a fuzzy regular expression
2. In PROSITE, the NR lines indicate:
 - A. the comment field
 - B. the list of true-positive sequences matched by the signature
 - C. the statistics of the diagnostic performance of the signature
 - D. the non-redundant list of sequences whose 3D structure are known
3. When searching the Blocks and PRINTS databases, a match is judged significant if:
 - A. a single motif is matched
 - B. two motifs are matched
 - C. the E-value is above e^{-4}
 - D. a combined E-value above a given threshold is reported for a multiple-motif match
4. TrEMBL is:
 - A. an automatically annotated composite protein sequence database
 - B. an automatically annotated supplement to the EMBL database
 - C. an automatically annotated supplement to the InterPro database
 - D. a translation of coding sequences in the EMBL database
5. UniProt is:
 - A. the universal protein sequence database derived from SWISS-PROT and TrEMBL
 - B. the universal protein resource derived from SWISS-PROT, TrEMBL and PIR-PSD
 - C. the universal protein family resource
 - D. the universal protein structure database
6. InterPro is:
 - A. an integrated protein family database
 - B. an integrated protein sequence database
 - C. an integrated protein structure database
 - D. an integrated protein interaction database
7. In a sequence database of a given size, which of the following expressions is likely to retrieve more matches:
 - A. D-A-V-I-D
 - B. [DE]-A-V-I-[DE]

- C. [DE]-[AVILM]-X-E
- D. D-A-V-E

8. Which of the regexs below is not compatible with the following motif:

PIFMIPAFYFTWIEMQCS
 PIFMIPAFYFSWIELQGS
 PIFMVPAFYFSWIQMAAS
 PLMALPAFYFSWWSLVTS
 PLMALPAYYFSWWHLKTS
 PLVTIGAFFFSWIDLAYS

- A. P-[IL]-[FMV]-X-[IVL]-[PG]-A-[FY]2-F-[TS]-W-[IW]-X-[ML]-X2-S
- B. P-[IVL]-[FMV]-[MAT]-[IVL]-[PG]-A-[FY]2-F-[TS]-W-[IW]-X-[ML]-X2-S
- C. P-[IL]-[FMV]-X-[IVL]-[PG]-A-[FY]4-[TS]-W-[IW]-X-[ML]-X2-S
- D. P-[IVL]-X2-[IVL]-[PG]-A-[FY]2-F-[TS]-W-[IW]-X-[ML]-X2-S

9. Which of the following groupings is not compatible with the traditional Venn diagram of overlapping amino acid physicochemical properties:

- A. FYWH, AVILMP, DEQN, KR, STCG
- B. FM, YWH, AVILP, DE, KR, STQN, CG
- C. FYW, AVILM, DE, KRH, STNQ, C, PG
- D. FYW, AVILPST, DEQN, KRH, MCG

10. With knowledge of the physicochemical properties of the amino acids, which of the following hydrophatic rankings is unlikely to be correct?:

- A. FILVWAMGTSYQCNPHEKDR
- B. IFVLWMAGCYPTSHENQDKR
- C. IVLRCMAGTSWYPHDNEQFR
- D. FYILMVWCTAPSRHGKQNE

11. Which of the sequences below is not compatible with the following regex:

H-[VILM]-G(2)-S-[EDQN]-T-A-[VILM]

- A. HMGGSQATAM
- B. HVGGETAV
- C. HIGGSSTAL
- D. HIGGETAL

12. The Midnight Zone is the region of sequence similarity:

- A. above 20% identity
- B. where sequence alignments are not statistically significant, as the same alignment may have arisen by chance
- C. below 20% identity

- D. where sequences fail to be detected by even the most sensitive sequence-based search algorithms

13. The Twilight Zone is the region of sequence similarity:

- A. above 50% identity
- B. where sequence alignments are not statistically significant, as the same alignment may have arisen by chance
- C. below 50% identity
- D. where sequences fail to be detected by even the most sensitive sequence-based search algorithms