Q1

score = 875

human:

M--------------------Q--NS--HSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEK-QQ

fly:

VEASTASHPHSTSSYFATTYYHLTDDECHSGVNQLGGVFVGGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATAEVVSKISQYKRECPSIFAWEIRDRLLQENVCTNDNIPSVSSINRVLRNLAAQKEQQ

Q2

fly and consensus PAX domain

score = 455

alignment:

fly (top): consensus PAX domain (bottom)

VEASTASHPHSTSSYFATTYYHLTDDECHSGVNQLGGVFVGGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATAEVVSKISQYKRECPSIFAWEIRDRLLQENVCTNDNIPSVSSINRVLRNLAAQKEQQ

----------------G-----------HGGVNQLGGVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIKPGVIGGSKPKVATPKVVEKIAEYKRQNPTMFAWEIRDRLLAERVCDNDTVPSVSSINRIIR---------

human and consensus PAX domain

score = 602

alignment:

human (top): consensus PAX domain (bottom)

MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQ

---GHGGVNQLGGVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIKPGVIGGSKPKVATPKVVEKIAEYKRQNPTMFAWEIRDRLLAERVCDNDTVPSVSSINRIIR--------

Q3

The similarity of human and fly Eyeless protein are high. For example, both shared conserved PAX domain, and each contains ~135 identical (or very similar) amino acid in this region. The probability of this happened by chance is around 1/(23^135).

Q5

mean: 51.478  
standard deviation: 6.67469220264  
z-score: 123.379771681

Q6

The local alignment score for human and fly eyeless protein is more than 3 standard derivation from the mean of score of randomly generated sequence. This is reflected in the high z-score. Hence, it is very unlikely that similarity between human and fly eyeless protein is due to chance alone.

Q7

diag\_score = 2

off\_diag\_score = 1

dash\_score = 0

Q8

bumble

fumble

humble

humbled

humbler

humbles

humbly

jumble

mumble

rumble

tumble

direly

finely

fireclay

firefly

firmly

firstly

fixedly

freely

liefly

refly

tiredly

Q9

Organize the list of words into a ‘tree structure’, in accordance of the sequence of letters of each word. For example:

root: c

second layer: ca, ce, ch, ci…

third layer: under ‘ca’: cab, cac, cad, cae,….

Such data structure will allow finding words and the neighboring words faster.