Advanced Protein based Alignments Based on Sequence

Γενικές πληροφορίες από το κεφάλαιο 1 της βιολογίας προσανατολισμού 3ης λυκείου. Εξετάζουμε έννοιες που είτε χρειάζονται άμεσα στην κατανόηση του κειμένου είτε μπορούν να παρομοιαστούν με άλλες έννοιες πχ προσανατολισμός , αλληλουχίες βάσεων.

Το DNA και το RNA, είναι μακρομόρια, αποτελούμενα από νουκλεοτίδια. Κάθε νουκλεοτίδιο του DNA αποτελείται από μια πεντόζη, τη δεοξυριβόζη, ενωμένη με μια φωσφορική ομάδα και μια αζωτούχα βάση. Στα νουκλεοτίδια DNA η αζωτούχος βάση μπορεί να είναι :

a. αδενίνη (A)
b. γουανίνη (G)
c. κυτοσύνη (C)
d. θυμίνη (T)

Σε κάθε νουκλεοτίδιο η αζωτούχος βάση συνδέεται με τον 1ο άνθρακα δεοξυριβοζης και η φωσφορική ομάδα με τον 5ο. Μια πολυνουκλεοτιδική αλυσίδα σχηματίζεται από την ένωση πολλών νουκλεοτιδίων με ομοιοπολικό δεσμό. Ο δεσμός αυτός δημιουργείται μέσω του υδροξυλίου του 3ου άνθρακα της πεντόζης και του επόμενου νουκλεοτιδίου και της φωσφορικής ομάδας που είναι συνδεδεμένη με στον 5ο άνθρακα της πεντόζης του επόμενου νουκλεοτιδίου. Ο δεσμός αυτός ονομάζεται φωσφοδιεστερικός. Με τον τρόπο αυτό η πολυνουκλεοτιδική αλυσίδα που δημιουργείται έχει έναν σκελετό, που αποτελείται από επανάληψη των μορίων φωσφορική ομάδα-πεντόζη-φωσφορική ομάδα-πεντόζη.

Ανεξάρτητα από τον αριθμό των νουκλεοτιδίων από τα οποία αποτελείται η πολυνουκλεοτιδική αλυσίδα, το πρώτο της νουκλεοτίδιο έχει πάντα μια ελεύθερη φωσφορική ομάδα συνδεδεμένη στον 5ο άνθρακα της πεντόζης του και το τελευταίο νουκλεοτίδιό της, έχει ελεύθερο το υδροξύλιο του 3ου άνθρακα της πεντόζης του. Έτσι καθορίζεται ο προσανατολισμός της πολυνουκλεοτιδικής αλυσίδας 5->3.

- Το DNA αποτελείται από 2 πολυνουκλεοτιδικές αλυσίδες που σχηματίζουν στο χώρο μια δεξιόστροφη διπλή έλικα.
- Η διπλή έλικα έχει έναν σταθερό σκελετό, από επαναλαμβανόμενα μόρια φωσφορικής ομάδας-δεοξυριβόζης ενωμένων με φωσφοδιεστερικό δεσμό. Ο σκελετός αυτός είναι υδρόφιλος και βρίσκεται στο εξωτερικό του μορίου. Προς το εσωτερικό του σκελετού βρίσκονται οι αζωτούχες βάσεις που είναι υδρόφοβες.
- Οι αζωτούχες βάσεις της μιας αλυσίδας συνδέονται με δεσμούς υδρογόνου με τις αζωτούχες βάσεις της απέναντι αλυσίδας με βάση τον κανόνα της συμπληρωματικότητας. Η αδενίνη συνδέεται μόνο με θυμίνη και αντίστροφα, ενώ η κυτοσίνη μόνο με γουανίνη και αντίστροφα. Οι δεσμοί υδρογόνου που αναπτύσονται μεταξύ των βάσεων σταθεροποιούν την δευτεροταγή δομή του μορίου.
- Ανάμεσα στην αδεμίνη και τη θυμίνη σχηματίζονται 2 δεσμοί υδρογόνου, ενώ ανάμεσα στη γουανίνη και την κυτοσίνη σχηματίζονται 3 δεσμοί υδρογόνου.

- Οι 2 αλυσίδες ενός μορίου DNA είναι συμπληρωματικές, και αυτό υποδηλώνει ότι η αλληλουχία της μια καθορίζει την αλληλουχία της άλλης. Ησυμπληρωματικότητα έχει τεράστια σημασία για τον αυτοδιπλασιασμό του DNA, μια ιδιότητα που το καθιστά το καταλληλότερο μόριο για τη διατήρηση και μεταβίβαση της γενετικής πληροφορίας. Κάθε αλυσίδα DNA μπορεί να χρησιμεύσει ως καλούπι για τη σύνθεση μιας συμπληροωματικής αλυσίδας, ώστε τελικά να σχηματίζονται 2 δίκλωνα μόρια DNA πανομιότυπα με το αρχικό μόριο.

Μια λειτουργία του DNA είναι η έκφραση των γενετικών πληροφοριών, που επιτυγχάνεται με τον έλεγχο της σύνθεσης των πρωτεϊνών. Για την περιγραφή του μήκους ή της αλληλουχίας ενός νουκλεϊκού οξέος χρησιμοποιείται ο όρος αλληλουχία βάσεων. Στην πραγματικότητα εννοούμε την ακολουθία των νουκλεοτιδίων του νουκλεϊκού οξέος. Το 1953 υπήρξε η πρώτη υπόθεση ότι το DNA αυτοδιπλασιάζεται.

Τα κύτταρα, διαθέτουν εξειδικευμένα ένζυμα και άλλες πρωτεϊνες που δρουν ταυτόχρονα και καταλύουν τις χημικές αντιδράσεις της αντιγραφής DNA με ταχύτητα και ακρίβεια.

Η αντιγραφή ξεκινά από τις θέσεις έναρξης αντιγραφής. Στα ευκαριωτικάκύτταρα το DNA κάθε χρωμοσώματος είναι ένα μακρύ γραμμικό μόριο, με πολυάριθμες θέσεις έναρξης αντιγραφής. Έτσι η αντιγραφή γίνεται ταυτόχρονα σε εκατοντάδες σημεία σε όλο το μήκος και τα τμήματα που δημιουργούνται ενώνονται.

Για να αρχίσει η αντιγραφή, είναι απαραίτητο να ξετυλιχθούν στιςε θέσεις εκκίνησης οι δύο αλυσίδες. Αυτό επιτυγχάνεται με τη βοήθεια ειδικών ενζύμων που σπάζουν τους δεσμούς υδρογόνου μεταξύ των αλυσίδων. Τα ένζυμα αυτά ονομάζονται DNA ελικάσες.

Τα κύρια ένζυμα που συμμετέχουν στην αντιγραφή ονομάζονται DNA πολυμεράσες. Επειδή αυτά δεν μπορούν μόνα τους να αρχίσουν την αντιγραφή, το κύτταρο έχει ένα ειδικό σύμπλοκο, αποτελούμενο από πολλά ένζυμα, το πριμόσωμα, το οποίο συνθέτει στις θέσεις έναρξης αντιγραφλης μικρά τμήματα RNA, συμπληρωματικά προς τις μητρικές αλυσίδες. Οι πολυμεράσες επιμηκύνουν τα αρχικά τμήματα , τοποθετώντας συμπληρωματικά δεοξυριβονουκλεοτίδια απέναντι από τις μητρικές αλυσίδες. Επίσης οι πολυμεράσες επιδιορθώνουν λάθη κατά το στάδιο αντιγραφής.

Τέλος να προσθέσουμε ότι τα ένζυμα είναι πρωτεϊνες και οι πρωτεϊνες είναι μόρια αποτελούμενα από αλληλουχίες αμινοξέων. Οι αλληλουχίες νουκλεϊκών οξέων καθορίζουν τη σειρά των αμινοξέων στις πρωτεϊνες μέσω της μεταγραφής και της μετάφρασης. Οι πρωτεϊνες εκτελούν ποικίλες λειτουργίες στον οργανισμό, όπως η δομή και η λειτουργία των κυττάρων, η ανοσολογική αντίδραση, η μεταφορά οξυγόνου και η κατασκευή των ιστών. Η δευτεροταγής δομή των πρωτεϊνών αναφέρεται στην τρισδιάστατη μορφή των τοπικών τμημάτων της πρωτεΐνης. Η δευτεροταγής δομή ορίζεται επίσημα από το σχήμα των δεσμών υδρογόνου μεταξύ των ατόμων υδρογόνου της αμίνης και του καρβοξυλικού οξυγόνου στο πεπτίδιο σκελετού. Άλλοι τύποι βιοπολυμερών, όπως τα νουκλεϊκά οξέα, διαθέτουν επίσης χαρακτηριστικές δευτεροταγείς δομές. Συνολικά, η δευτεροταγής δομή παρέχει σημαντικές πληροφορίες για την τρισδιάστατη διάταξη των πρωτεϊνών και την εκτέλεση των βιολογικών τους λειτουργιών.
Τα πεπτίδια είναι δομικά στοιχεία πρωτεϊνών, ενζύμων, κυττάρων και ιστών του σώματος . Συνολικά, τα πεπτίδια αποτελούνται από δύο ή περισσότερα αμινοξέα που συνδέονται με

πεπτιδικό δεσμό. Ο αριθμός των αμινοξέων σε ένα πεπτίδιο υποδηλώνεται με αριθμητικό πρόθεμα, όπως διπεπτίδιο (2 αμινοξέα), οκταπεπτίδιο (με 8 αμινοξέα) ή ακόμα ολιγοπεπτίδιο ή πολυπεπτίδιο (περισσότερα από 50) . Η διάκριση μεταξύ πολυπεπτιδίου και πρωτεΐνης είναι μάλλον ασαφής, χωρίς καμία πρακτική σημασία.

Το κείμενο μας, αναφέρει τα μειονεκτήματα που έχουν οι αλγόριθμοι ομοιότητας αλληλουχιών αμινοξέων, όταν δεν βασίζονται στη δομή και τοπογραφία των μορίων. Αυτό διότι οι λειτουργίες των πρωτεϊνών βασίζονται πιο πολύ στη δομή. Συγκεκριμένα πολλές ομόλογες πρωτεΐνες, παρόλο την κοινή τους λειτουργία, το ποσοστό ομοιότητας των αλληλουχιών τους είναι μικρότερο από 10%. Τέτοια παραδείγματα είναι ένζυμα των οικογενειών helicases (ελικάσες), proteases και polymerases (πολυμεράσεις).

Ένα λογισμικό προτείνεται, το οποίο είναι ανοικτού κώδικα. Πιο συγκεκριμένα θα εξετάσουμε ένα εργαλείο του λογισμικού αυτού. Θα κάνει τις συγκρίσεις με μεγαλύτερη ακρίβεια επειδή εξετάζει μέσω της σειράς της αλληλουχίας αμινοξέον και τη δομή του μορίου. Έτσι προβλέπει καλύτερα τη λειτουργία των μορίων. Οι πληροφορίες για τη δομή θα είναι δευτεροταγείς, αλλά θα εξαιρούνται οι πλευρικές αλυσίδες. Ένα μητρώο περιέχει τις δευτεροταγής πληροφορίες πολλών μορίων και θα χρησιμοποιηθεί ώστε να μεταφράσει μια σειρά αμινοξέων σε σχηματική μορφή όταν είναι απαραίτητο. Αυτό το μητρώο βασίζεται στις μετρήσεις του χρησιμοποιώντας ιδιότητες υδροπάθειας των μορίων, καθώς αυτό συνδέεται άμεσα με τις φυσικοχημικές ιδιότητές τους και τη δομή τους.

Το PSSP είναι το εργαλείο πάνω σε αυτό το όγισμικό. Σε κάθε αναζήτησή μας μπορεί να ανακτά αποθηκευμένες δευτεροταγή πληροφορίες για πρωτεΐνες, όταν υπάρχουν αποθηκευμένες στη βάση PDB. Όταν δεν υπάρχουν για κάποιο συγκεκριμένο μοντέλο πρωτεϊνών, το λογισμικό τις προβλέπει με γρήγορους υπολογισμούς και μετά τις επαληθεύει χρησιμοποιώντας και άλλα εργαλεία. Εξετάζουμε τρισδιάστατες δομές ή τεταρτοταγείς. Το εργαλείο σε κάθε αναζήτηση ψάχνει στη βάση RCSB PDB για παρόμοιες δομές πρωτεϊνών.

Μια εικόνα για τη δευτεροταγή δομή των πρωτεϊνών μπορεί να δοθεί όταν αυτές βρίσκονται στην επιφάνεια του νερού κατά την μέτρηση.

Έχει αποδειχθεί ότι η περιοδικότητα πολικών και μη πολικών αμινοξέων, είναι καθοριστικός παράγοντας για τη δευτερογενή δομή των αυτο-συναρμολογούμενων ολιγομετρικών πεπτιδίων. Παρόλο που τα αμινοξέα μοιάζουν να διαφέρουν κάποιες φορές ενώ έχουν κοινές δομές, μέσα στο νερό, ο σχηματισμός υδρόφυλων δομών δείχνει πατέντες που μπορεί να είναι κοινές και αυτό έχει τεράστια σημασία σύμφωνα με έρευνες. Αυτές οι πατέντες είναι κοινές ανάμεσα σε οικογένειες, μπορούν εύκολα να διακριθούν όταν βάλουμε ένα μόριο μέσα στο νερό και θα ήταν πολύ δύσκολο να παρατηρηθούν διαφορετικά. Η υδροπάθεια μελετάει τις ιδιότητες του μορίου που συμβαίνουν λόγω της υδροφοβίας ή υδροφιλίας των αμινοξέων του. Αυτό μας έδωσε πολλές νέες δυνατότητες και έκανε δυνατή μέχρι και την ταυτοποίηση πιο μακρινών συγγενικών πρωτεϊνών.

Ομόλογες πρωτεΐνες μιας οικογένειας με παρόμοιες δομές, έχει βρεθεί να εμφανίζουν πολλές ομοιότητες όσον αφορά την υδροπάθειά τους, παρόλο που είχαν μηδαμηνές ομοιότητες στην αλληλουχία αμινοξέων.

Ομόλογα μοντέλα μορίων έχουν 30% και πάνω ομοιότητες. Για έναν συνηθισμένο αλγόριθμο που ψάχνει ομόλογα μόρια έχουμε τα εξής:

1. Ψάξε στη βάση για παρόμοια μοντέλα μορίων σχετικά με την είσοδό σου.
    1.1. Χρησιμοποίησε γρήγορες μεθόδους εξαντλητικής αναζήτησης και υπολόγισε ένα σκόρ μεταξύ αναζήτησης και μοντέλα της βάσης.
    1.2. Φτιάξε το μητρώο ομοιότητας και διέταξε τα πιθανά μοντέλα.
    1.3. Χρησιμοποίησε μια μέθοδο για μια ακόμη καλύτερη ταξινόμιση.
    1.4. Παρήγαγε μια πολλαπλή αντιστοίχιση με τα καλύτερα μοντέλα.

2. Ο αλγόριθμος θα ψάξει στη βάση, μέσω μιας ευρετικής συνάρτησης, τμήματα ίδιων δομών πρωτεϊνών, με την είσοδο(άγνωστη δομή). Αν η ομοιότητα είναι μεγάλη, η άγνωστη δομή θα αντιγράψει στοιχεία από το σκελετό και τις πλευρικές αλυσίδες. Αν η ομοιότητα είναι μικρή, θα αντιγράψει μόνο το σκελετό. Σε μηδέν ομοιότητα δεν αντιγράφει τίποτα. Τέλος για να τεστάρουμε την ποιότητα της αντιγραφής, ελέγχουμε αν τα τελικά στοιχεία έχουν ίδιες ιδιότητες υδροπάθειας.

Το φαινόμενο της υδροφοβίας, μεταφράζεται σε μη πολικές ουσίες (και στην τάση τους να αποφεύγουν το νερό). Η υδροπάθεια ενός αμινοξέος, που βασίζεται στις χημικές ιδιότητες των πλευρικών αλυσίδών του (αφού αυτές έχουν την ελευθερία κίνησης), βοηθάει να καθορίσουμε τον προσανατολισμό των πλευρικών αλυσίδων στις 3 διαστάσεις. Αν οι πλευρικές αλυσίδες είναι υδρόφοβες, μέσα στο νερό διπλώνουν προς τον πυρήνα του μορίου.

Ο προσδιορισμός των κλίσεων των υδρόφοβων ή υδρόφυλων πλευρικών αλυσίδων γίνεται με πολλούς τρόπους. Ένας είναι να μετρηθούν μέσα στο νερό και μετά σε ένα μη αντιδραστικό περιβάλλον ως ισοτροπικές, αφαιρώντας την διαφορά ενέργειας σε κάθε κατάσταση. Ένας άλλος τρόπος βασίζεται στον υπολογισμό του μέσου όρου της θέσης των ατομικών συντεταγμένων 12 πρωτεϊνών. Ένας τρίτος τρόπος είναι ο συνδυασμός των παραπάνω.

Τα 20 γνωστά αμινοξέα έχουν ήδη υπολογισμένες υδροπάθητικές ιδιότητες.

Κλάση υδρόφιλων αμινοξέων: D,N,E,Q,K,R

Κλάση υδρόφοβων αμινοξέων: I,V,F,C,M,A,W

Κλάση ουδέτερων αμινοξέων: G,T,S,Y,P,H

Τα αμινοξέα με μεγάλες μη πολικές πλευρικές αλυσίδες τείνουν να είναι υδροφοβικά, ενώ τα πολύ πολικά τείνουν να είναι υδροφιλικά. Πολλά αμινοξέα έχουν και υδροφιλικά και υδροφοβικά τμήματα. Για τη μέτρησή τους χωρίζουμε τις περιόδους μέτρησης. Οι δομές των πρωτεϊνών σχετίζονται με πολλά πράγματα όπως τη δυναμική ενέργεια των μορίων, την εντροπία και τις ηλεκτροστατικές δυνάμεις. Όμως η υδροπάθεια θεωρείται η πιο σημαντική - δυνατή.

Λίγα πράγματα για τη διεπαφή λογισμικού

Αρχικά, το λογισμικό μας δίνει επιλογές εισαγωγής τύπου-ερωτήματος. Επιλέγουμε σειρά αμινοξέων ή DSSP-δευτεροταγών στοιχείων σειρών πρωτεϊνών.

Η αναζήτηση για όμοια μοντέλα γίνεται είτε με τις κλασσικές σειρές αμινοξέων είτε με δευτερογενή στοιχεία με το STRAP μοντέλο. Το μοντέλο STRAP θα προβλέψει μια πρώτη εικόνα για την δομή. Έπειτα θα κάνει αναζήτηση για δευτερογενή τιαριάσματα με άλλα μόρια. Μετά θα γίνει αναζήτησηστη βάση δεδομένων πρωτεϊνών.

Η αναζήτηση ομοιότητας ακολουθεί δύο βήματα,

1. αναζήτηση για αλληλουχίες με ομοιότητα > 30%
2. μέσω του μητρόου ομοιότητας διαλέγονται από τις παραπάνω μόνο όσες αλληλουχίες ικανοποιούν το προφιλ υδροπάθειας.

Στα αποτελέσματα χρησιμοποιούνται χρώματα για να περιγράψουν κατά πόσο τομοντέλο είναι ακριβές.

Οι δευτεροταγείς πληροφορίες κωδικοποιούνται με 8 σύμβολα.

Το λογισμικό χρησιμοποιεί το πακέτο MOE για αναπαράσταση πρωτεϊνών 3-διαστάσεων. Ένας έλεγχος που έγινε με συναρτήσεις του MOE έδειξε ότι τα λάθη ήταν ασήμαντα και οτι το μοντέλο λειτουργεί κανονικά.

Οι υπολογισμοί μοριακών δυναμικών έγιναν με το GROMACS, λαμβάνωντας υπόψην τοπολογία και σύστημα. Το εργαλείο αυτό περιέχει και βάση με τις τοπολογίες νουκλεοτιδίων και αμινοξέων. Κάποιες δυσκολίες εντοπίζονται λόγω της πολυπλοκότητάς του.

Δυστυχώς όταν μπαίνουμε στη σελίδα του λογισμικού να πειραματιστούμε με αυτό οδηγούμαστε σε μια αρχική οθόνη που λέει enter. Αφού πατάμε enter οδηγούμαστε εδώ: http://ww25.bioinfoteam.com/?subid1=20240406-2046-3316-8900-afcc8ae4214a όπου είναι απλά μια κενή σελίδα.