Στο paper με τίτλο “Antibody Clustering Using a Machine Learning Pipeline that Fuses Genetic, Structural, and Physicochemical Properties” παρουσιάζεται μια υβριδική προσέγγιση για την αναγνώριση και συσταδοποίηση ομάδων αντισωμάτων. Η συσταδοποίηση (clustering) του V domain των αντισωμάτων είναι σημαντική για τη βελτίωση τους ώστε να μπορούν να χρησιμοποιηθούν σε ανοσολογικές θεραπείες. Χάρη στην πολυπλοκότητα των αντισωμάτων, το πρόβλημα της κατηγοριοποίησής τους σε συστάδες φαίνεται να διαφέρει διαρκώς ανάλογα με το μέγεθος των διαθέσιμων δεδομένων. Δεδομένου του τεράστιου αριθμού αντισωμάτων από διαφορετικά είδη οργανισμών και τις νέες διαθέσιμες μέσω προσομοίωσης τρισδιάστατες κατασκευές των μεταβλητών περιοχών heavy-chain(VH) και light-chain(VL) των αντισωμάτων, η προσοχή έχει στραφεί στην υπερμεταβλητή περιοχή (hypervariable region(HVR)) του αντισώματος και συγκεκριμένα στις περιοχές συμπληρωματικότητας (complementarity-determining regions(CDRs)) που συμμετέχουν κυρίως στην αλληλεπίδραση αντισώματος-αντιγόνου.

/\*Επέκταση με γενικές πληροφορίες\*/

Στην ανοσολογία, αντιγόνο είναι μια μοριακή δομή που μπορεί να εισέλθει στο σώμα μέσω ενός παθογόνου μικροοργανισμού και προκαλεί υπό φυσιολογικές συνθήκες την αντίδραση του σώματος που είναι γνωστή ως ανοσολογική απόκριση. Τα αντισώματα είναι μοριακές δομές που παράγονται από το ανοσοποιητικό σύστημα. Κάθε αντίσωμα έχει την ικανότητα να αναγνωρίζει και να προσδένεται με συγκεκριμένο αντίσωμα, δεσμεύοντάς το και αναστέλλοντας έτσι την δράση του. ( <https://el.wikipedia.org/wiki/%CE%91%CE%BD%CF%84%CE%B9%CE%B3%CF%8C%CE%BD%CE%BF> )

Η προτεινόμενη προσέγγιση σύμφωνα με τους συγγραφείς διατηρεί τα πλεονεκτήματα της συσταδοποίησης αλλά αξιοποιεί παράλληλα την πληροφορία που προσφέρει η τρισδιάστατη αναπαράσταση. Οι συγγραφείς υποστηρίζουν ότι η τεχνική αυτή είναι χρήσιμη για την συσταδοποίηση αντισωμάτων που δεν μπορούν να κατηγοριοποιηθούν αποτελεσματικά με φυλογενετικές τεχνικές ή αποκλειστικά με αλγορίθμους συσταδοποίησης, και ότι μπορεί να εφαρμοστεί και για άλλες κατηγορίες πρωτεϊνών πέρα από τα αντισώματα.

Οι περισσότερες HVR στα αντισώματα έχουν έναν μικρό αριθμό διακριτών διατάξεων που ονομάζονται κανονικές δομές (canonical structures) και έναν μικρό αριθμό υπολειμμάτων-κλειδιών που θα μπορούσαν να χρησιμοποιηθούν για την πρόβλεψη της συστάδας στην οποία θα πρέπει να κατηγοριοποιηθεί κάποιο συγκεκριμένο αντίσωμα. Επιπλέον, έχουν ανακαλυφθεί με βάση το CDR νέες συστάδες κανονικών δομών σε συγκεκριμένες θέσεις συντήρησης που είναι υπεύθυνες για τις διαφοροποιήσεις ανάμεσα σε συστάδες αντισωμάτων. Σε μια παλαιότερη προσέγγιση για την κατηγοριοποίηση των παρατηρούμενων CDRs, έγινε μελέτη των συστάδων στον εσωτερικό χώρο συντεταγμένων (internal coordinate space), ακολουθούμενη από συγχώνευση ομάδων δομών στον Καρτεσιανό χώρο χρησιμοποιώντας την απόκλιση της μέσης τετραγωνικής τιμής (root-mean-square-deviation). Πρατηρήθηκε στο παρελθόν ότι παρότι δύο συστάδες μπορεί να ήταν πιο κοντά σε κάποια κανονική κλάση όσο αφορούσε την αλληλουχία τους, οι δομές τους μπορεί να ήταν πιο κοντά σε κάποια άλλη. Παρατηρήθηκε επίσης ότι όταν το heavy chain H3 χωρίστηκε σε δύο περιοχές, κεφαλή (head) και σώμα (torso) και μελετήθηκαν τα τρία CDRs σε αυτό, το σώμα συνήθως έπαιρνε μία από δύο μορφές, ενώ η δομή της κεφαλής εξαρτιόταν από τη δομή του σώματος. Μπορεί επομένως να γίνει ανάλυση και κατηγοριοποίηση με βάση το μήκος των CDRs ή δομικές πληροφορίες στην επιφάνεια των σημείων συνένωσης του V domain.

Για τους σκοπούς της έρευνας συλλέχθηκαν 3.100 δείγματα δομών αντισωμάτων V domain από τη βάση δεδομένων Protein Data Bank (PDB). Χρησιμοποιήθηκαν μέθοδοι εξόρυξης δεδομένων, χαρακτηριστικά της αλληλουχίας και της δομής των αντισωμάτων, ένα υβριδικό μητρώο απόστασης (hybrid distance matrix) και αλγόριθμοι φυλογενετικής ανάλυσης. Για έλεγχο της ποιότητας τους, τα αποτελέσματα συγκρίθηκαν με τα 3D δεδομένα που υπάρχουν στην βάση δεδομένων Protein Data Bank. Το τελικό σύνολο δεδομένων περιείχε 2.022 μοναδικά V domains, το καθένα με τουλάχιστον ένα ολοκληρωμένο variable domain ενός heavy ή light chain. Kατασκευάστηκε μια βάση δεδομένων με τα δομικά τους στοιχεία και ανακαλύφθηκαν πάνω από 50 υπογραφές αντισωμάτων. Στη συνέχεια τα δείγματα αριθμήθηκαν και καθορίστηκαν οι CDR περιοχές μέσω κανονικών εκφράσεων για τα μοτίβα συντηρημένων ακολουθιών (conserved sequence).

Μελετήθηκαν τα διμερή αντιγόνου-αντισώματος που ανήκαν στο ίδιο PDB αρχείο και όλες οι αναγνωρίσιμες αλληλεπιδράσεις εντοπίστηκαν, σημειώθηκαν και μετατράπηκαν σε ψηφιακές υπογραφές βασισμένες στο V domain και στη συνάρτηση:

*I* (*absi* ) = 0, if all the elements of *R*(*absi ag m* ) > 4

*I* (*absi* ) = 1, if one of the elements of *R (abs ag i im* ) <= 4

, όπου R η ευκλείδια απόσταση μεταξύ δύο σημείων στον τρισδιάστατο χώρο, abs το αντίσωμα, ag το αντιγόνο, n ο αριθμός των ατόμων στο αρχείο PDB για το αντίσωμα και m ο αριθμός των ατόμων στο PDB αρχείο για το αντιγόνο. Τα V domains για τα αντισώματα και οι τρισδιάστατες κατασκευές για τα αντιγόνα θεωρήθηκαν τρισδιάστατα σημεία στο χώρο. Όλες οι εντοπιζόμενες αλληλεπιδράσεις αντικαταστάθηκαν στις πρωτεϊνικές αλληλουχίες με 0 ή 1 και κάθε περιοχή του V domain αντικαταστάθηκε με 0 ή 1 αν μία ή περισσότερες αλληλεπιδράσεις εντοπίστηκαν σε αυτήν. Το αποτέλεσμα ήταν ένα διάνυσμα «υπογραφής» 5 τιμών όπου κάθε τιμή αντιστοιχούσε σε μια περιοχή του V domain.

Το υβριδικό μητρώο αποστάσεων υπολογίστηκε από τη συνάρτηση:

H = A\*B,

όπου Α Jukes-Cantor μητρώο και Β μητρώο για τις αποστάσεις μεταξύ ατόμων στα αντισώματα. Οι υβριδικές αποστάσεις που υπολογίστηκαν αναλύθηκαν ιεραρχικά και υποβλήθηκαν σε πολλαπλή στοίχιση μέσω του εργαλείου MATLAB. Έγινε φυλογενετική ανάλυση και τα αποτελέσματα οπτικοποιήθηκαν σε μορφή δέντρων με χρήση ακτινοβολίας MEGA και cycle option. Κάθε μοναδικό V domain αντισώματος προστέθηκε στη βάση με τις δομικές πληροφορίες και διαχωρίστηκε σε δύο μέρη, τα “heavy chains” και “light chains” και σημειώθηκαν τα αντισώματα που σχημάτιζαν διμερές σύμπλεγμα με κάποιο αντιγόνο. Τα δεδομένα της βάσης συμπληρώθηκαν με τις απαραίτητες πληροφορίες από τη βάση NCBI. Όλα τα δεδομένα της βάσης χωρίστηκαν σε συστοιχίες με βάση την υπογραφή τους. Τα φυλογενετικά δέντρα που κατασκευάστηκαν από το μοντέλο είχαν παρόμοια τοπολογία και μοτίβα διακλάδωσης. Τα αντισώματα με ασαφή συμπεριφορά ομαδοποιήθηκαν με αντισώματα που είχαν γνωστή συμπεριφορά για έναν αριθμό πιθανών αντιγόνων. Κάθε διαφορετική ομάδα-συστοιχία αντιστοιχεί σε συγκεκριμένο αριθμό μοτίβων αλληλεπίδρασης από όλα τα μοτίβα που βρέθηκαν συνολικά και οι πιο συχνά εμφανιζόμενες αλληλεπιδράσεις φάνηκε πως υπήρχαν μέσα στις συστάδες 1, 2, 3, 5 και 8.

Τα αποτελέσματα έδειξαν ότι δεν υπήρχε αλληλεπίδραση μόνο μεταξύ CDR και αντιγόνων, αλλά ότι διαφορετικά αμινοξέα αλληλεπιδρούν με τα αντιγόνα, και μάλιστα με διαφορετική συχνότητα από περιοχή σε περιοχή. Σε κάποιες περιπτώσεις υπήρχε αλληλεπίδραση μόνο με την περιοχή FR(fragment region) του αντισώματος μέσα στα VL και VH domains. Επιπλέον, πολλά αμινοξέα «κλειδιά» φαίνεται να παρουσιάζουν διαφορές στην κάθε εμφάνισή τους, ενώ κάποια άλλα φαίνεται να χάνονται. Αυτή η παρατήρηση είναι σημαντική για την πειραματική κατασκευή αντισωμάτων, πιθανών και νέων, που πιθανόν θα χρησιμοποιηθούν σε θεραπείες στο μέλλον. Στον μηχανισμό αλληλεπίδρασης του CDR καθοριστικές ήταν οι τυροσίνη (tyrosine) και σερίνη (serine), ενώ στον μηχανισμό αλληλεπίδρασης των περιοχών FR οι γλυκίνη (glycine), τρυπτοφάνη (tryptophan) και φαινυλαλανίνη (phenylalanine).

Ένα από τα θετικά του προτεινόμενου μοντέλου είναι ότι με τη χρήση του συνδυασμού των αποστάσεων στο υβριδικό μητρώο μπορούσε να γίνει συσταδοποίηση άγνωστων μορίων αντισωμάτων με γνωστά εύκολα και βέλτιστα χωρίς να χρειάζονται επιπλέον ευρετικές τεχνικές, οι οποίες θα απαιτούσαν και περισσότερο χρόνο. Το μοντέλο επίσης επέτρεψε να ληφθούν υπ’ όψη και οι τρισδιάστατες δομές, που έχουν μεγάλη σημασία στη συσταδοποίηση. Η συνάρτηση που χρησιμοποιήθηκε είναι ευέλικτη γιατί αντιστοιχεί κάθε συσχέτιση σε επίπεδο αλληλουχίας με την ομοιότητα σε επίπεδο τρισδιάστατης δομής.

Συμπεράσματα της συγκεκριμένης μελέτης ήταν ότι η κατηγοριοποίηση των αντισωμάτων σε συστάδες αποτελεί πρόκληση για την ανοσολογία. Ακόμα και αν δύο αντισώματα έχουν ακριβώς την ίδια αλληλουχία μπορεί στον τρισδιάστατο χώρο να παρουσιάζουν εντελώς διαφορετική δομή και συμπεριφορά απέναντι στα αντιγόνα. Για να αντιμετωπιστεί το πρόβλημα, πρέπει να ληφθούν υπ’όψη όλες οι απαιτούμενες βιολογικές παράμετροι σε κάποιας μορφής υβριδικό μοντέλο, όπως το προτεινόμενο. Τα αποτελέσματα της χρήσης τέτοιων μοντέλων θα μπορούσαν να οδηγήσουν σε ριζοσπαστικές ανακαλύψεις που αφορούν την έρευνα γύρω από την ανοσολογία, την βιοτεχνολογία, τη διάγνωση, την ανοσοθεραπεία και άλλους τομείς.

/\*ΕΠΕΚΤΑΣΗ\*/

Μια εφαρμογή της συσταδοποίησης αντισωμάτων με ερευνητικό ενδιαφέρον είναι το antibody engineering, η κατασκευή δηλαδή, αντισωμάτων σε εργαστηριακό περιβάλλον. Προς το παρόν η βάση δεδομένων PDB περιέχει κυρίως θραύσματα (fragments) αντισωμάτων και κάποια ολόκληρα δείγματα. Η μοντελοποίηση των αντισωμάτων έχει μεγάλη σημασία γιατί ο τρόπος που γίνεται η πρόσδεση αντισώματος σε αντιγόνο είναι πολύ συγκεκριμένος και γίνεται σε συγκεκριμένα σημεία της επιφάνειας των μορίων υπό συγκεκριμένη γωνία (elbow bend). Στην πρόσδεση αυτή καθοριστικός είναι ο προσανατολισμός του V domain και οι κανονικές δομές των CDRs. Μάλιστα, δομικοί περιορισμοί στο σημείο πρόσδεσης είναι πιθανό να εμποδίσουν την αναγνώριση του αντιγόνου από το αντίσωμα. Η γνώση των κανονικών δομών των μορίων επιτρέπει την ευκολότερη μοντελοποίησή τους. Αφού όμως η μοντελοποίηση με προσομοίωση σε υπολογιστή είναι προς το παρόν περιορισμένη, η ύπαρξη πληροφορίας σε βάσεις δεδομένων για την συμπεριφορά συστάδων από αντιγόνα μπορεί να διευκολύνει τη σωστή πρόβλεψη της δομής, συνεπώς και την κατασκευή τους ή ακόμα και την βελτίωση της δομής τους στο εργαστήριο.

(πηγή:

Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody Structure and Function: The Basis for Engineering Therapeutics. Antibodies (Basel). 2019 Dec 3;8(4):55. doi: 10.3390/antib8040055. PMID: 31816964; PMCID: PMC6963682.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6963682/>

)

Το clustering αντισωμάτων έχει επίσης εφαρμοστεί στην επιλογή ασθενών για συμμετοχή σε δοκιμές θεραπειών με βάση την κατηγοριοποίηση των αντισωμάτων τους σε συστάδες. Έχει σημασία σε τέτοιες δοκιμές τα ανοσολογικά προφίλ των υποψηφίων να εμφανίζουν ποικιλομορφία, για να μπορεί να θεωρηθεί ότι τα συμπεράσματα της έρευνας αντιπροσωπεύουν το σύνολο του πληθυσμού. Τo clustering με βάση το CDR έχει αποδειχθεί χρήσιμο, αν και όχι με πλήρη επιτυχία, στον εντοπισμό ομοιοτήτων και διαφορών ανάμεσα σε σύνολα από δείγματα αντισωμάτων, ειδικά σε συνδυασμό με αλγορίθμους και εργαλεία τεχνητής νοημοσύνης.

(πηγή:

Chomicz D, Kończak J, Wróbel S, Satława T, Dudzic P, Janusz B, Tarkowski M, Deszyński P, Gawłowski T, Kostyn A, Orłowski M, Klaus T, Schulte L, Martin K, Comeau SR, Krawczyk K. Benchmarking antibody clustering methods using sequence, structural, and machine learning similarity measures for antibody discovery applications. Front Mol Biosci. 2024 Mar 28;11:1352508. doi: 10.3389/fmolb.2024.1352508. PMID: 38606289; PMCID: PMC11008471.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11008471/>

)