

# Neurodegenerative diseases associated antibody repertoire signatures in mimotope arrays based on cyclic versus linear peptides\*

Shina Pashova-Dimova<sup>1</sup>, Peter Petrov<sup>2</sup>, Sena Karachanak-Yankova<sup>3</sup>, Anastas Pashov<sup>4</sup>

<sup>1</sup> *Institute of Biology and Immunology of Reproduction, Bulgarian Academy of Sciences, Sofia, Bulgaria*

<sup>2</sup> *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria*

<sup>3</sup> *Department of Medical Genetics, Medical University, Sofia, Bulgaria*

<sup>4</sup> *Institute of Microbiology, Bulgarian Academy of Sciences, Sofia, Bulgaria*

Corresponding author: Anastas Pashov ([a\\_pashov@microbio.bas.bg](mailto:a_pashov@microbio.bas.bg))

Received 3 November 2023 ♦ Accepted 16 November 2023 ♦ Published 30 November 2023

**Citation:** Pashova-Dimova S, Petrov P, Karachanak-Yankova S, Pashov A (2023) Neurodegenerative diseases associated antibody repertoire signatures in mimotope arrays based on cyclic versus linear peptides. *Pharmacia* 70(4): 1439–1447. <https://doi.org/10.3897/pharmacia.70.e115179>

## Abstract

The role of peptide probes' conformational flexibility in extracting immunosignatures has not been sufficiently studied. Immunosignatures profile the antibody diversity and prove promising for early cancer detection and multi-disease diagnostics. A novel tool for modeling antibody repertoires, the concept of antibody reactivity graphs, proved instrumental in this respect. Serum samples from patients with Alzheimer's disease (AD), frontotemporal dementia (FTD), dementia of unknown etiology (DUE), and healthy controls were probed using a set of 130 7-mer peptides relevant to neurodegenerative diseases. Results show that linear peptides probed with IgM yielded higher graph density compared to IgG, indicating different levels of polyspecificities. Additionally, the impact of peptide topology and antibody isotype on feature selection was studied using recursive feature elimination. Findings reveal that IgM assays on linear peptides offer superior diagnostic differentiation of neurodegenerative diseases and define the degree of agreement between IgG and IgM immunosignatures with linear or cyclic peptides.

## Keywords

igome, immunosignatures, cyclic peptides, mimotopes, antibody repertoires, graphs, peptide arrays

## Introduction

Cyclic and linear peptides have different applications in the field of vaccines and immunoassays. Cyclic peptides can be engineered to resist non-specific degradation in the body and can be activated upon exposure to target-specific

environmental factors, making them suitable for targeted drug delivery (Hamley 2022). On the other hand, linear peptides are used as antigens in the development of vaccines and immunotherapies for infectious diseases and cancer (Schoeniger and Anderson 2019; Touti and Kwong

\* This work was supported in part by the Bulgarian National Science Fund under Grants KP-06-China-10/2020 and KP 06 N 33/5 2019.

2019). Linear peptides can stimulate specific immune responses and are being used in advanced clinical trials as vaccines (Pashov and Kieber-Emmons 2019; Tardón et al. 2019). In immunoassays, linear peptides can be used to mimic secondary structures of proteins and assemble epitopes for antibody assays (Samuel and Nir 2016). Cyclic peptidomimetics, which are modified cyclic peptides, have advantages in binding assays such as conformational constraint and increased specificity.

Extracting immunosignatures is a technology that can be used to evaluate vaccines and predict vaccine effectiveness (Joseph Barten and Stephen Albert Johnston 2013; Luhui et al. 2014). It involves profiling the antibody diversity of an individual using high throughput binding assays on large peptide arrays. The peptides can be random as well as derived from existing self or non-self proteins (Stafford et al. 2014; Pashov et al. 2019). Immunosignatures have shown promise in detecting cancer at early stages. Additionally, immunosignatures can be used to provide a comprehensive diagnostic for multiple diseases simultaneously, as they are effectively disease-agnostic (Stafford et al. 2014). They have demonstrated high accuracy in classifying autoimmune diseases such as Systemic Lupus Erythematosus (SLE) and Rheumatoid Arthritis (RA) (Pashova et al. 2022). Recently, an optimization of the immunosignature technique has been proposed by replacing the random peptides with libraries of peptides (mimotopes) affinity selected by the complete IgM repertoire (IgM IgOme) (Pashov et al. 2019; Pashova et al. 2022). This allows better focusing on immunologically relevant peptide probes. Although the immunosignature approaches have been studied in a variety of aspects and the effect of cyclization on antibody binding is well known, so far very little is known about the changes in the immunosignatures after cyclisation of linear probes. To fill this gap, patterns of binding to linear or cyclic forms of a set of IgM mimotopes (IgOme (Ryvkin et al. 2012; Pashov et al. 2019)) was studied using sera from patients with Alzheimer's disease (AD), frontotemporal dementia (FTD), dementia of unknown etiology (DUE) and healthy controls. The IgOme library, on its part, was designed based on a preliminary analysis of the binding of IgM from patient with AD or FTD to an IgOme library (Pashov et al. 2019) and selection of targets of over- and under-expressed IgM reactivity.

## Materials and methods

### Serum samples

Patients were recruited at the Neurology Clinic of the Medical University, Sofia. Serum samples (0.5 ml) from patients with AD, FTD, DUE or no signs of dementia ( $n = 4$  for each group) were collected after informed consent. The collection and the following studies were approved by the Human Studies Ethics Committee of the Medical University, Sofia. The anonymized samples were kept frozen at  $-15^{\circ}\text{C}$  until processing.

### IgM isolation

Each serum sample was thawed and incubated at  $37^{\circ}\text{C}$  for 30 min for the dissolution of IgM complexes. Next, the sera were centrifuged and 30-fold diluted with PBS followed by ultrafiltration (Amicon-Ultra 100kDa membranes, Millipore) for initial fractionation of serum proteins above 100 kDa. The high-molecular fraction of serum proteins was applied in a series on columns (HiTrap Protein G High Performance and HiTrap IgM Purification, GE Healthcare) for affinity purification of IgM and IgG according to the manufacturer's instructions.

### Peptide microarray binding assay

Custom peptide microarray chips from containing 130 7 amino acid residues long linear or cyclic peptides produced by PEPperPRINT (Heidelberg, Germany) were used. The peptides were synthesized in situ in an oriented array, attached to the chip's surface through their C-terminus (linear) or C- and N-terminus (cyclic) with a spacer sequence GSGSG. The microarray layout consisted of the peptide spots duplicated in random positions. The sequences of the peptides were selected from a larger IgOme library (Pashov et al. 2019) after preliminary analysis of the binding of IgM from patients with AD or FTD as well as IgM from healthy donors (IVIgM) (Pashov et al. 2019, manuscript in preparation). The present study uses sera from different patients than those used to determine the relevant mimotopes. The sequences of the peptides used is given in Table 1. The microarrays were blocked for 60 min using PBS, pH 7.4, and 0.05% Tween 20 with 1% BSA on a rocker; washed  $3 \times 1$  min with PBS, pH 7.4, and 0.05% Tween 20; and incubated with sera in dilutions equivalent to 0.01 mg/mL IgM ( $\sim 1:100$  serum dilution) on a rocker overnight at  $4^{\circ}\text{C}$ . After  $3 \times 1$  min washing, the chips were incubated with secondary antibodies at room temperature, washed, rinsed with distilled water, and dried by spinning in a vertical position in empty 50 mL test tubes at  $100 \times g$  for 2 min. Two secondary antibodies with different fluorochromes and different wavelengths of detection were used to measure simultaneously the binding of IgG and IgM on the same array.

### Microarray data analysis

The microarray fluorescence images were acquired in a Innoscan 1100 (Innosys, Carbonne, France). The densitometry was performed using MAPIX software. All further analysis was performed using publicly available packages of the R statistical environment (Bioconductor, Biostrings, limma, pepStat, sva, e1071, uwot, clvalid, etc.) as well as in-house developed R scripts. The data underpinning the analysis reported in this paper and the scripts of the analysis are deposited at <https://github.com/ansts/cyclic>. The details of the analysis procedures are described elsewhere (Pashov et al. 2019). The reactivity graph approach was largely following the previously described

**Table 1.** Sequences of peptide mimotope probes with over/under expressed reactivity in AD and FTD used in this study.

Alzheimer's disease			Frontotemporal dementia		
Over expressed		Under expressed	Over expressed		Under expressed
ADDACPR	GTIPGQP	DAEGFTK	AATQLWW	MTDMSLL	AAYKGEE
AEECNIC	GYPGLWS	DAHVRLA	ADPGYHS	NLAPRPH	ARSVHPI
DAGPCRP	HDYENRG	DKAEIWH	ADVARTH	NPHHVTR	AWKWDFI
DGASNLP	HEIGSQL	DQPHVWN	AGVAPRL	NPVQAHY	DRCCVLD
DGGLIRI	HPLRHSG	DSGCGHQ	AHNWWFD	QDQICHC	EPVTSYL
DHCFARR	MEPQVII	EANSIAF	AQSMEFV	QFTMATF	EQSAWRE
DHRNSIR	MQCPNDC	EDCKWCR	ARPAEMS	QSSMLER	GHLPVWS
EAHYRGP	MYGVDQN	EEGLIRG	ATRADYF	QVIFPNH	HPDFWPI
EHVPRIL	NGEPLIP	EEVQIPV	AVDGTDR	RAADEYS	HPPAGIL
GAPKHWL	QDMPRLP	EHSLETE	AWARHES	RDVLDVY	HTRADVY
GATGSLP	QMGINLD	EPVIPRS	CCLAWDP	RSTDLYT	KPVEWRV
GHARLSP	QTVEWYR	ERLTCEF	GASLRPG	RTTPPHY	MDTDALT
GIVSYPG	RIAQNHP	ETVFWRM	GATGAYN	RWDFFPA	MGTPKED
GKHITMW	RWIDKVP	ETWIGPI	GATGSYP	SGWNEMV	MGVQTEV
GLLRPSP	SMHLGFI	GPAVTTS	GCCGADP	SPIDTWS	MIHDKRY
GMHLSNW	SPDDLRV	GPGSQAT	GDEARDG	SQGYSMH	MLRTADT
GMPTRTF	TLEEFFP	GPPGVSR	GEESYGW	TGVTRDS	MPHKNDF
GNRVAYV	TQEYWRG	GPPLTWK	GHCRRNM	TIWADF	MVKNYAD
GQAGGLI	VERMYTP	HPGWAWQ	GLENLSH	TKTVTER	RFPVDQH
GQIALSS	VWPQIIG	NPALWCC	GRWSDSY	TNPHGDT	RPFVYEV
GQIDKIP	WDRNIHL	RLPHPLP	GTPVLSH	TQGFQTM	TDEIHQM
GQNVTAP	WGTTTVA	RMEITNL	HDLMWHR	TTDARIH	TELKEMI
GQVFTYP	WHGVQNI	TGSSWLTV	HKVTDVF	TTDIPAR	THLAQDV
GSIIFHR	WPLMLMP	TQNYAAI	HMATHPW	TTDRMTM	TTELLVA
GTATTLT	WRDASMP		HWEPMRN	TTFRLPD	TTLPLPT
			IANRAEQ	VERTLSY	VQNMWPFV
			LDGPRPH	WERDCCT	
			MPIRGPM	WTKGEHF	

algorithms (Ferdinandov et al. 2023). Briefly, the densitometry data was cleaned based on the flags set during densitometry, the background was inferred based on the gradients of staining intensity between the randomly positioned duplicate peptide spots in the plane of the array. The background was subtracted by taking the residuals of the linear regression of the signal on the background followed by multiplying them by a factor compensating for the dependence of the standard deviation on the background (using a power law). This approach was adopted because the background was caused by insufficient mixing during the incubation.

Further, the logarithms of the data were normalized with respect to amino acid residue composition (Pashov et al. 2019) to compensate for changes in the binding due to overall charge and hydrophobicity rather than the dependence on the sequence which is the main interest. In addition, the data was normalized between arrays using the `limma::normalizeCyclicLoess` function with the “affy” method applied in 3 iterations separately for the 4 sets of data grouped by isotype and peptide topology. The 16 sets of data separated by isotype, peptide topology and diagnosis were used further to generate separate reactivity graphs. The criterion for cross-reactivity between two peptides used to define the edges of the graphs used a similarity measure based on a function of the correlation and the coefficient of variation of the two profiles compared. If

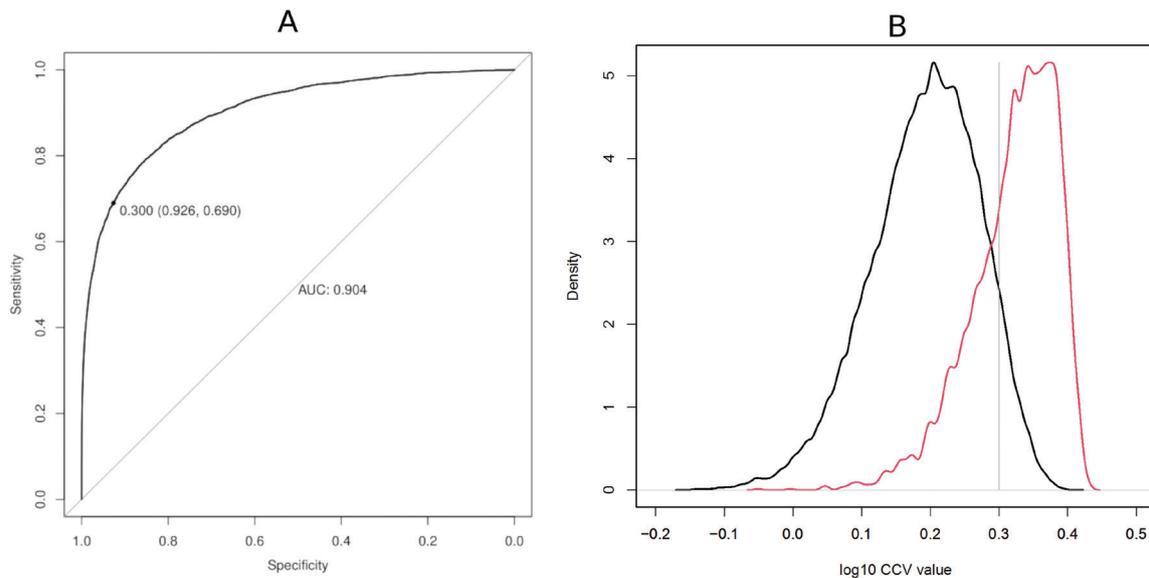
the two vectors of reactivity values (logarithm of the staining intensity in arbitrary units) for two peptides across a set of  $n$  patients' sera are  $v_i = x_{i,1} \dots x_{i,n}$  and  $v_j = x_{j,1} \dots x_{j,n}$ ,  $\rho = \text{corr}(v_i, v_j)$  is their correlation and

$$cv = \frac{\sqrt{\sum ((v_i, v_j) - \mu_{(v_i, v_j)})}}{2n - 1}$$

is the coefficient of variation of their concatenation, then:

$$CCV = \rho + 2 - \frac{cv}{k + cv}$$

is a function which tends to  $\rho + 1$  as  $cv \rightarrow 1$  and to  $\rho + 2$  as  $cv \rightarrow 0$ . The value for  $k = 0.3$  was found to maximize the area under the ROC curve of the CCV criterion when classifying cross-reactive peptides. The CCV criterion remains relatively high even for low correlation if the coefficient of variation is also very low. In this way, reactivity profiles which are flat and very similar in mean values are estimated cross-reactive. The criterion was tested using the algorithm and the sequence set form (Ferdinandov et al. 2023). The ROC curve and the selected threshold for connecting vertices of the graphs are illustrated in Fig. 1. Vertices representing peptides reactivities with CCV lower than the threshold remain disconnected while those above the threshold are connected with the value of CCV used as a weight of the respective edge.



**Figure 1.** (A) ROC curve illustrating the capacity of the CCV criterion to classify 4150 pairs of peptides overlapping in 11/15 positions as a model of cross-reactive peptide pairs compared to a set of 10,000 pairs of dissimilar peptides (sharing a longest common subsequence of fewer than 3/15). The analysis is done on the basis of the dataset from [1]. The logarithms of the values of CCV are used. (B) Distributions of the log CCV values for dissimilar sequences (black) vs. cross-reactive sequences (red). The optimal tradeoff between sensitivity and specificity was found for  $CCV = 2$  (specificity = 0.926, sensitivity = 0.69, AUC = 0.904)

## Results

### Reactivity graphs

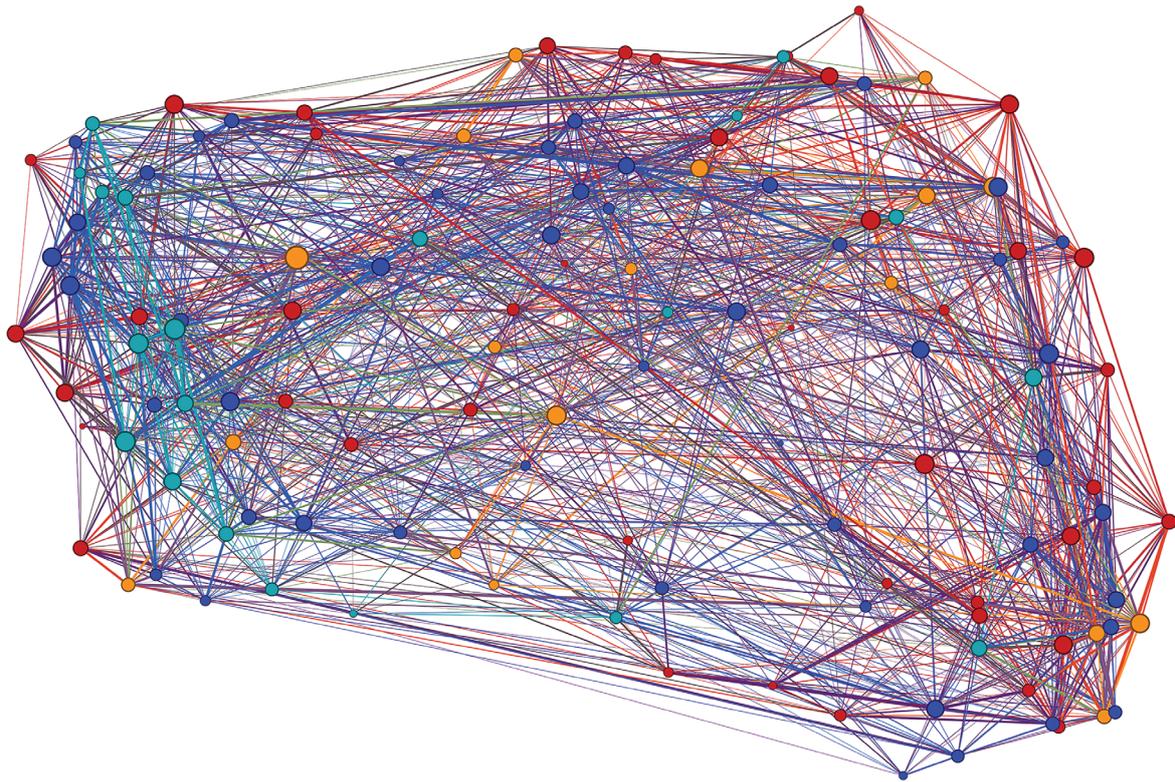
In our previous reports we demonstrated the utility of antibody reactivity graphs as a tool for system level studies of the repertoires of antibody specificities. The reactivity graphs represent the antibody cross-reactivity relations between the peptide probes used. They are weighted and undirected. Also, they are usually highly connected due to antibody cross-reactivity especially when the IgM repertoire is probed with short peptides. Their capacity to encode diagnostically relevant information depends on the degree of cross-reactivity and specificity, the public nature of the repertoire features addressed (to ensure generalization), the diversity of the probe array, the use of targeted arrays with known relevance to the problem at hand, etc.

Here we address the question of the relative utility in binding assays of an array of peptide probes in linear (free) or cyclic (constrained) topology (conformation) in IgG or IgM reactivity graphs. Serum IgG and IgM from patients with AD, FTD, DUE or controls without dementia ( $n = 4$  for each diagnosis) were tested using peptide arrays of 130 probes preselected for their significant over or under expression in AD and FTD. The binding of the antibodies was detected by a fluorochrome conjugated anti-IgG or anti-IgM secondary antibody and quantitated by scanning the fluorescence intensity. After acquisition, cleaning, background subtraction and normalization, the fluorescence data was used to construct 16 separate graphs using the binding profiles similarities (see Materials and methods).

The graphs represent data grouped by 3 factors: isotype (IgG and IgM), peptide topologies (cyclic or linear) and the diagnoses ( $n = 4$ ) having also 4 different sera for each diagnosis. The graphs were studied either separately or by combining them in multigraphs, e.g. – grouped by isotype and/or topology, as well as after simplifying the multigraphs by summing up the weights of the parallel edges so that there is only one edge between two vertices. The overall graph produced as the union of all 16 graphs is shown on Fig. 2. For this graph the edge weight threshold of 17.4 was used (range – 2–36) which made the graph sparser and easier to visualize its core structure. The layout (embedding) of the graph is calculated using the eigenvectors of the graph Laplacian corresponding to the 16 lowest non-zero eigenvalues (the number of eigenvectors is determined by a dimensionality reduction algorithm). The 16-dimensional embedding was further transformed to 2 dimensions using the UMAP algorithm (R functions `igraph::embed_laplacia_matrix` and `uwot::umap`). The modularity of the graph partition according to peptide library equaled the 0.949 quantile of the simulated modularity distribution for equal sized random partitions.

There was a weak but discernible separation between the peptide libraries with respect to their reactivities with the tested sera. When the modularity was calculated dichotomously for each library against the rest, the simulated distribution quantiles were respectively: 0.969 for AD high, 0.861 for FTD low, 0.647 for AD low, and 0.626 for FTD high. In all cases, the simulation was done by generating 1000 partitions of the same sizes as the tested.

When the graphs were grouped by topology/isotype, the four resultant multigraphs had significantly different



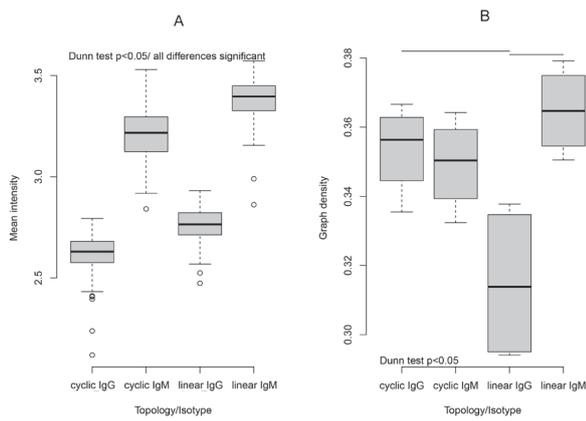
**Figure 2.** Overview of the general reactivity graph constructed as the union of the reactivity graphs under different conditions (IgG or IgM binding; cyclic or linear peptides; AD, FTD, DUE and Control – altogether 16 different graphs based on the same vertices). The edge weights of the separate original graphs were summed. To outline only the strongest similarities, the edges were kept if their weight exceeded 17.4 (range 2–36). The vertices (peptide sequences) are color coded according to the source mimotope library (red – over expressed in AD, orange – under expressed in AD, dark blue – over expressed in FTD, light blue – under expressed in FTD). The color of the edges is a mixture of the colors of their incident vertices. The thickness of the edges is proportional to their weight interpreted as strength of cross-reactivity. The layout of the graph is an embedding based on the the 16 eigenvectors corresponding to the lowest non zero eigenvalues further projected to 2 dimensions using the UMAP algorithm. The modularity of the graph with respect to the partition by mimotope libraries equaled the 0.949 quantile of the bootstrapped modularity. When the modularity was bootstrapped dichotomously for each library against the rest, the quantiles were: 0.969 for AD high, 0.861 for FTD low, 0.647 for AD low and 0.626 for FTD high.

mean intensities and graph densities (Fig. 3). The density of the reactivity graph can be interpreted as a correlate of the overall cross-reactivity of the antibodies tested with the peptide probes used. The graphs of linear peptide reactivities probed with IgM had a higher density compared to the IgG assay. For the cyclic peptides the difference between the IgG and IgM graphs was smaller than for the linear ones. These findings are interpretable in terms of difference in polyspecificities of IgG/IgM and in binding entropies as well as in the diversity of both the repertoire (Suppl. material 1: fig. S1) and the presented epitopes in constrained vs linear peptides.

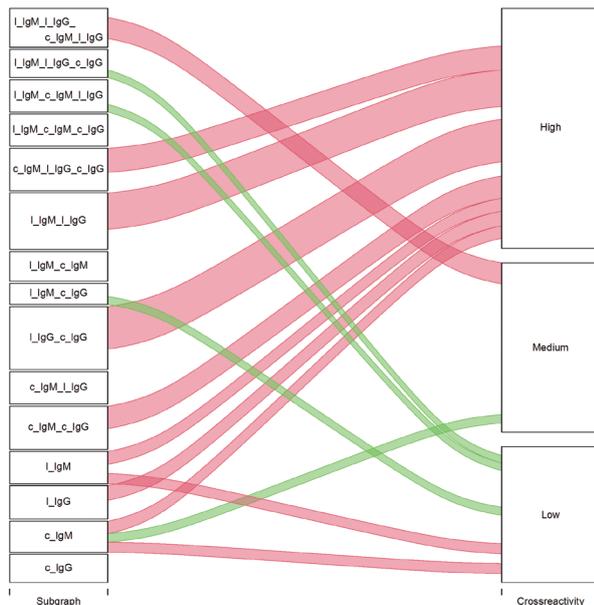
To study in more detail the commonality between the images of the IgG vs IgM repertoires probed with linear vs cyclic peptides, the sixteen original graphs were aggregated in 4 topology/isotype (T/I) graphs named linear\_IgM, cyclic\_IgM, linear\_IgG and cyclic\_IgG. These multigraphs were simplified by summing the parallel edges' weights. The significance of the overlap of the in-

dividual T/I graphs was estimated calculating the sum of the weights of the parallel edges in all multigraphs generated by uniting the different combinations of the four T/I graphs. These were compared to weight sums of 1000 random graphs generated by scrambling the existing edges and their weights. The scrambling is done among all edges existing in the overall multigraph some of which are not found in individual graphs and thus are assigned weight 0 initially. Each of the graphs obtained by uniting a combination of T/I graphs was further stratified into 3 subgraphs based on edge weight using the ranges [2, 2.3), [2.3, 2.6) and  $\geq 2.6$ .

The magnitudes of the weight sums which are outside the 0.05–0.95 quantile range of the simulated values were considered significant. The significant weight sums are shown on Fig. 4. The distribution is drawn towards high cross-reactivity among multiple subgraphs which indicates a considerable consensus between the different conditions including between arrays of peptides in linear



**Figure 3.** A. Mean intensity of the binding data for the graphs grouped by the topology of the peptides (cyclic or linear) and by the isotype of the tested antibodies; B. Graph density of graphs grouped by the topology and isotype. The graph density is the ratio of the number of edges to the theoretical maximum for each graph. Among the graphs of the data based on linear peptides and tested with patients IgG showed lower density while those tested with IgM – higher than the cyclic peptide graphs. This is interpreted as lower, resp.: higher, cross-reactivity.



**Figure 4.** Graph overlaps. The four T/I graphs and their various combinations had their edge weights categorized as low – [2, 2.3), medium – [2.3, 2.6) and, high – >2.6 indicating the respective levels of cross-reactivity (pattern similarity). The sums of the edge weights which were outside the 0.05–0.95 quantile range of the simulate distribution are illustrated. The thickness of the connecting strips corresponds to the sums of weights of the overlapping edges. For some of the graphs the number of overlapping edges is significantly increased (red) or decreased (green) relative to the simulated randomly connected graphs. The distribution is drawn towards high overlap among multiple subgraphs which indicates a considerable consensus between the different conditions including between arrays of peptides in linear vs cyclic topology.

vs cyclic topology. Furthermore, the high overlap of high cross-reactivities between all four T/I graphs ( $\cap(lIgM, lIgG, cIgM, cIgG)$ ) in contrast to no over representation but some under representation for three of the four 3 T/I graph combinations ( $\cap(lIgM, lIgG, cIgG)$ ),  $\cap(lIgM, cIgM, lIgG)$ ,  $\cap(lIgM, cIgM, cIgG)$ ) indicates also a non-random general agreement between the different assays for a substantial part of the peptide probes used. On the other hand, significant high cross-reactivity parts of the graphs remain which are not overlapping (the single graph edges of linear\_IgM, linear\_IgG and cyclic\_IgM). Interestingly, there was a high agreement between profiles in the IgG assay irrespective of the topology of the peptides which is in contrast with no non-random overlap for the two IgM graphs while there is a better agreement between cyclic\_IgM and the two IgG graphs ( $\cap(cIgM, lIgG, cIgG)$ ) as well as agreement between another subset of linear\_IgM and linear\_IgG.

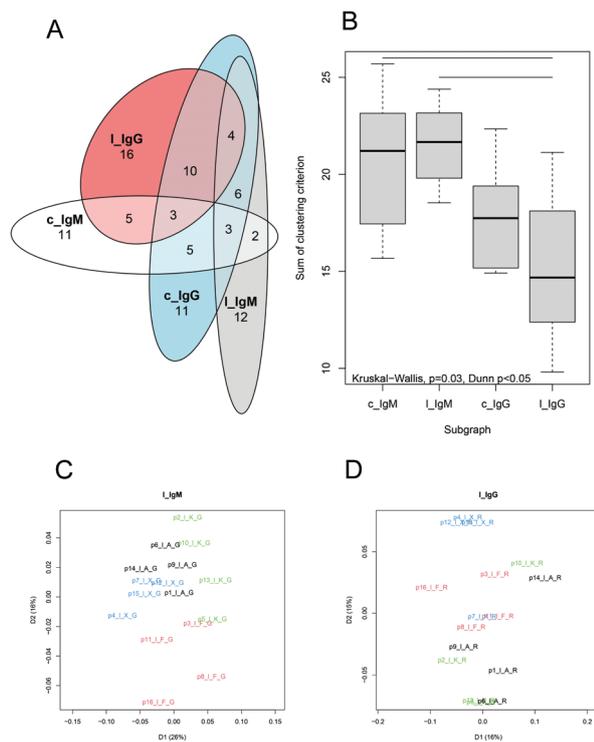
Thus, despite the differences both in the repertoire of reactivities and in the binding mode, IgG and IgM repertoires are partially comparable in their cross-reactivity with the tested array of peptide probes with the distinction of repertoire compartments with high and low overlap of reactivities.

### Effect of peptide topology and antibody isotype on feature selection

An efficient machine learning model based on repertoire patterns implies selecting a relevant subset among  $\sim 10^3$  peptide reactivities. Typically, less than a hundred peptides are selected which separate well the diagnostic groups of patients (Pashov et al. 2019; Ferdinandov et al. 2023). Previously, we have used a criterion which is a function of three clustering criteria: Dunn's criterion, Baker-Hubert Gamma index and connectivity (Dunn 1974; Baker and Hubert 1975; Handl et al. 2005). When used on data labeled according to diagnosis, the combined criterion measures the degree of separation of the predefined clusters. Maximizing it over subsets of features helps select a (locally) optimal feature set which typically generalizes well the classification (Ferdinandov et al. 2023). This was done using recursive feature elimination (RFE).

In the present study, the size of the groups does not allow building a generalizing model. Nevertheless, RFE can still be used to measure the performance of the different data sets in a possible classifier. The effect of the tested factors can be compared using the maximal value of the clustering criterion achieved by the optimal feature set since the different assays are performed on the same set of peptide sequences. The distribution of the different subsets of peptides selected in RFE using the four T/I graphs is shown in Fig. 5A.

The IgG and the cyclic peptide conditions led to selecting larger subsets of peptides (cyclic\_IgG – 42, linear\_IgG – 38, cyclic\_IgM – 29, and linear\_IgM – 28 sequences). This correlated inversely with the quality of the separation (Fig. 5B) with the linear peptides and IgM assay providing



**Figure 5.** **A.** Venn diagram of the overlap of peptide sequences selected by recursive feature elimination as a minimal set of peptide reactivities which separates optimally the 4 diagnoses; **B.** Comparison of the quality of separation based on the different graphs. The separation of the cases of the different diagnosis was estimated using the clustering criterion and the six values from these comparisons were used to further compare the different feature sets. IgM based immunosignature patterns were more efficient. For IgG based patterns, the topology seemed to have a greater (and opposite to those in the IgM assays) effect but it did not reach statistical significance. **(C)** and **(D)** multidimensional scaling projections of the different patients' sera profiles with the optimal feature sets (peptide sequences) for linear\_IgM (C – best separation) and linear\_IgG (D – worst separation).

the best separation. Interestingly, measured by the Jaccard distance, the subsets selected in the IgG assays overlapped more than those in the IgM assays and the subsets selected both in the linear and cyclic IgM conditions overlapped better with the cyclic\_IgG one than between themselves (not reaching statistical significance, though).

These findings indicate that IgM assays on linear peptides differentiate better the diagnoses especially with respect to the IgG conditions (Fig. 5C, D).

## Discussion

The present study explores the effects of peptide probe cyclisation on the performance of repertoire level binding assays. The primary tool of our functional repertoire studies are reactivity graphs. They are based on the concept of cross-reactivity of pairs of probes. The probability for  $n$

repertoires to contain each two different sets of antibodies which exhibit the same level of reactivity to the two probes is inversely proportional to  $n$  and very small. Thus, if two peptides' reactivities with a set of  $n$  repertoires correlate, most probably they are recognized by largely overlapping sets of antibodies in each repertoire, i.e. – they are cross-reactive (or isospecific). Using reactivity graphs, a tendency for a higher cross-reactivity of the IgM repertoire was found on the linear probes as compared to the cyclic ones ( $p = 0.11$ ) as well as a significant increase as compared to IgG on linear probes. An intriguing finding is the opposite effect of cyclisation of the cross-reactivity of IgG and IgM antibodies.

It was tempting to interpret these findings in terms of diversity of the tested repertoires since the simulation (Suppl. material 1: fig. S1) indicates that a more diverse repertoire produces a higher mean intensity of the assay and lower reactivity graph density compared to sparser repertoires. The explanation can be found also in the targeted reactivities of the IgM repertoires vs more random reactivities in the IgG repertoires because the set of probes used were selected on the basis of IgM reactivities in AD and FTD. Differences in cross-reactivity of two repertoires/assays towards the same set of antigenic probes can be due also to the existence of public antibodies (Setliff et al. 2018; Shrock et al. 2023) towards a large proportion of the probes vs binding by totally diverse repertoires.

Previously, cyclic peptide mimotopes have been found to bind with higher affinity even compared to the nominal antigen (Chen et al. 2023). As expected, the antibody binding to a cyclic conformation of a mimotope is found stronger compared to the linear conformation (Cabezas et al. 2000; Heo et al. 2020). Nevertheless, cyclic peptides still have some flexibility and can engage a binding pocket through divergent modes (Patel et al. 2020) so differences are mostly quantitative. The repertoire of structures seen by the antibodies also changes with cyclisation. Cyclic peptides can function as mimetics of conformational epitopes, providing targets for antibody binding that cannot be identified using synthetic linear peptides (Denisova et al. 2010). Indeed, some antibodies exhibit exclusive preference to cyclic peptides (Brett et al. 2002). Here we find that overlaps between the graphs for the different conditions are greatest for pairs of graphs, lower for all four graphs and even lower for single graphs. Thus, subsets of the cross-reactivities are defined, some of which are characteristic for all graphs some are unique for a graph but are found in different diagnoses, but the largest compartment is the one found in pairs of graphs. In the latter case it is interesting that there is a considerable overlap between graphs on the same topology but from different isotypes as well as between cyclic and linear for IgG but not for cyclic and linear for IgM. This indicated that in the case of the IgM repertoire the set of epitopes seen in the cyclic and in the linear library are much more disparate than under the other conditions studied.

A major difference between the two isotypes is the valency of the antibodies. Under the conditions of the

peptide array, IgM antibodies can and IgG mostly cannot bind the peptide molecules in multivalent manner. As a rule, IgG antibodies have higher intrinsic affinity to their nominal epitopes than IgM of similar specificity but the IgM compensate by avidity (multiple binding sites). In the case of igome mimotope arrays (Pashov et al. 2019), where the binding is by definition dependent on polyspecificity and broad cross-reactivity, IgG have negligible probability of finding their nominal epitope in random peptides and their affinity advantage is lost to a great extent. Adding the increased entropic component in the binding energy in the case of linear peptides due to their conformational mobility, it is understandable that IgG antibodies will have both the lowest mean intensity and graph density on linear arrays compared to the cyclic ones or to IgM.

The slight increase in cross-reactivity in linear IgM vs cyclic IgM may be due to the higher flexibility of most of the IgM paratopes (Manivel et al. 2000). Thus, combined with the flexibility of the probe, may be this system explores a larger conformational space to the extent that it benefits sufficiently from a greatly increased polyspecificity to overcome the entropic penalty of the flexible binding.

## References

- Baker FB, Hubert LJ (1975) Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association* 70: 31–38. <https://doi.org/10.1080/01621459.1975.10480256>
- Brett PJ, Tiwana H, Feavers IM, Charalambous BM (2002) Characterization of oligopeptides that cross-react with carbohydrate-specific antibodies by real time kinetics, in-solution competition enzyme-linked immunosorbent assay, and immunological analyses. *Journal of Biological Chemistry* 277: 20468–20476. <https://doi.org/10.1074/jbc.M200387200>
- Cabezas E, Wang M, Parren PW, Stanfield RL, Satterthwait AC (2000) A structure-based approach to a synthetic vaccine for HIV-1. *Biochemistry* 39: 14377–14391. <https://doi.org/10.1021/bi0003691>
- Chen K, Ge L, Liu G (2023) Integrated in silico–in vitro rational design of oncogenic EGFR-derived specific monoclonal antibody-binding peptide mimotopes. *Journal of Bioinformatics and Computational Biology* 21: 2350007. <https://doi.org/10.1142/S0219720023500075>
- Denisova GF, Denisov DA, Bramson JL (2010) Applying bioinformatics for antibody epitope prediction using affinity-selected mimotopes – relevance for vaccine design. *Immunome Research* 6(suppl 2): S6. <https://doi.org/10.1186/1745-7580-6-S2-S6>
- Dunn JC (1974) Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics* 4: 95–104. <https://doi.org/10.1080/01969727408546059>
- Ferdinandov D, Kostov V, Hadzhieva M, Shivarov V, Petrov P, Bussarsky A, Pashov AD (2023) Reactivity graph yields interpretable IgM repertoire signatures as potential tumor biomarkers. *International Journal of Molecular Sciences* 24: 2597. <https://doi.org/10.3390/ijms24032597>
- Hamley IW (2022) Peptides for vaccine development. *ACS Applied Bio Materials* 5: 905–944. <https://doi.org/10.1021/acsbm.1c01238>
- Handl J, Knowles J, Kell DB (2005) Computational cluster validation in post-genomic data analysis. *Bioinformatics* 21: 3201–3212. <https://doi.org/10.1093/bioinformatics/bti517>
- Heo CK, Hwang HM, Lim WH, Lee HJ, Yoo JS, Lim KJ, Cho EW (2020) Cyclic peptide mimotopes for the detection of serum anti-ATIC autoantibody biomarker in hepato-cellular carcinoma. *International Journal of Molecular Sciences* 21(24): 9718. <https://doi.org/10.3390/ijms21249718>
- Joseph Barten L, Stephen Albert J (2013) Immunosignatures can predict vaccine efficacy. *Proceedings of the National Academy of Sciences of the United States Of America* 110(46): 18614–18619. <https://doi.org/10.1073/pnas.1309390110>
- Luhui S, Debra TH, Stephen Albert J, Joseph Barten L (2014) Could immunosignatures technology enable the development of a preventative cancer vaccine. *Expert Review of Vaccines* 13(5): 577–579. <https://doi.org/10.1586/14760584.2014.897616>
- Manivel V, Sahoo NC, Salunke DM, Rao KV (2000) Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* 13: 611–620. [https://doi.org/10.1016/S1074-7613\(00\)00061-3](https://doi.org/10.1016/S1074-7613(00)00061-3)
- Tardón MC, Allard M, Dutoit V, Dietrich P-Y, Walker PR (2019) Peptides as cancer vaccines. *Current Opinion in Pharmacology* 47: 20–26. <https://doi.org/10.1016/j.coph.2019.01.007>
- Pashov A, Shivarov V, Hadzhieva M, Kostov V, Ferdinandov D, Heintz K-M, Pashova S, Todorova M, Vassilev T, Kieber-Emmons T, Meza-Zepeda LA, Hovig E (2019) Diagnostic profiling of the human public IgM repertoire with scalable mimotope libraries. *Frontiers in Immunology* 10: 1–14. <https://doi.org/10.3389/fimmu.2019.02796>
- Pashov A, Kieber-Emmons T (2019) Mimetic Vaccines in Immuno-Oncology. In: Arnouk H (Ed.) *Cancer Immunotherapy and Biological Cancer Treatments*. IntechOpen, 112 pp. <https://doi.org/10.5772/intechopen.85593>
- Pashova S, Balabanski L, Elmadjian G, Savov A, Stoyanova E, Shivarov V, Petrov P, Pashov A (2022) Restriction of the global IgM repertoire in antiphospholipid syndrome. *Frontiers in Immunology* 13: 1–17. <https://doi.org/10.3389/fimmu.2022.865232>

Thus, with regards to repertoire immunosignature assays, interrogating the IgM repertoire with linear probes seems to have some limited advantage over probing IgG and the use of cyclic peptides. This conclusion is limited by circumstances in which IgG repertoire is of a particular interest as predominantly immune memory associated and pathogen selected. The advantage of using IgM was also confirmed in the efficiency of a feature selection algorithm. The superiority of IgM assays may be due to a difference between the IgG and the IgM repertoires with respect to the specificities which differentiate the neurodegenerative diseases. For an in-depth analysis, it would be better to test on two set probes selected by IgG and IgM disease specific repertoires.

## Conclusion

These results indicate that linear peptide based immunosignature probes provide more information and a more efficient extraction of features for a subsequent machine learning based design of biomarkers than their cyclic version, especially in terms of testing the IgM repertoire.

- Patel K, Walport LJ, Walshe JL, Solomon PD, Low JKK, Tran DH, Mouradian KS, Silva APG, Wilkinson-White L, Norman A, Franck C, Matthews JM, Guss JM, Payne RJ, Passioura T, Suga H, Mackay JP (2020) Cyclic peptides can engage a single binding pocket through highly divergent modes. *Proceedings of the National Academy of Sciences* 117: 26728. <https://doi.org/10.1073/pnas.2003086117>
- Ryvkin A, Ashkenazy H, Smelyanski L, Kaplan G, Penn O, Weiss-Ottolenghi Y, Privman E, Ngam PB, Woodward JE, May GD, Bell C, Pupko T, Gershoni JM (2012) Deep Panning: steps towards probing the IgOme. *PLoS ONE* 7: e41469. <https://doi.org/10.1371/journal.pone.0041469>
- Samuel JSR, Nir Q (2016) Cyclic Peptides for Protein – Protein Interaction Targets: Applications to Human Disease. *Critical Reviews in Eukaryotic Gene Expression* 26(3): 199–221. <https://doi.org/10.1615/CritRevEukaryotGeneExpr.2016016525>
- Schoeniger JSB, Anderson PC (2019) Linear Peptides That Mimic Structural Epitopes for Antibody Assays. *National Technology & Engineering Solutions of Sandia, LLC (Albuquerque, NM), USA*.
- Setliff I, McDonnell WJ, Raju N, Bombardi RG, Murji AA, Scheepers C, Ziki R, Mynhardt C, Shepherd BE, Mamchak AA, Garrett N, Karim SA, Mallal SA, Crowe Jr. JE, Morris L, Georgiev IS (2018) Multi-donor longitudinal antibody repertoire sequencing reveals the existence of public antibody clonotypes in HIV-1 infection. *Cell Host & Microbe* 23(6): 845–854.E6. <https://doi.org/10.1016/j.chom.2018.05.001>
- Shrock EL, Timms RT, Kula T, Mena EL, West AP, Guo R, Lee I-H, Cohen AA, McKay LGA, Bi C, K, Leng Y, Fujimura E, Horns F, Li M, Wesemann DR, Griffiths A, Gewurz BE, Bjorkman PJ, Elledge SJ (2023) Germline-encoded amino acid-binding motifs drive immunodominant public antibody responses. *Science* 380: eadc9498. <https://doi.org/10.1126/science.adc9498>
- Stafford P, Cichacz ZA, Woodbury NW, Johnston SA (2014) Immuno-signature system for diagnosis of cancer. *Proceedings of the National Academy of Sciences of America*: 111(30): E3072–E3080. <https://doi.org/10.1073/pnas.1409432111>
- Touti F, Kwong G (2019) Engineered Cyclic Peptides. *U.S. Patent Application* 16(435): 350.

## Supplementary material 1

### Supplementary method

Authors: Shina Pashova-Dimova, Peter Petrov, Sena Karachanak-Yankova, Anastas Pashov

Data type: docx

Explanation note: Description of the simulation of repertoire binding and reactivity graph.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/pharmacia.70.e115179.suppl1>