

High Fidelity Cyclic Peptide Microarray to Diagnose Rheumatoid Arthritis

Hari Krishnamurthy (✉ hari@vibrantsci.com)

Vibrant Sciences

Vasanth Jayaraman

Vibrant Sciences

Karthik Krishna

Vibrant Sciences

Karenah Rajasekaran

Vibrant America

Tianhao Wang

Vibrant America

Kang Bei

Vibrant Sciences

John Rajasekaran

Vibrant Sciences

Keith Elkon

University of Washington

Research Article

Keywords: ACPA, CMP, citrullinated, carbamylated peptides, cyclisation, linear peptides, carbamylated antigens

Posted Date: May 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-501179/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

High fidelity Cyclic Peptide Microarray to diagnose Rheumatoid Arthritis.

Hari Krishnan Krishnamurthy^{1*}, Vasanth Jayaraman¹, Karthik Krishna¹, Karenah E. Rajasekaran², Tianhao Wang², Kang Bei¹, John J. Rajasekaran¹, Keith B Elkon³

¹*Vibrant Sciences LLC., San Carlos, CA, USA*

²*Vibrant America LLC., San Carlos, CA, USA*

³*Division of Rheumatology, Department of Medicine, University of Washington, Seattle, WA, USA*

Correspondence to:

Hari K. Krishnamurthy: hari@vibrantsci.com

Keith B Elkon: elkon@uw.edu

Emails of Authors:

Hari K. Krishnamurthy: hari@vibrantsci.com

Vasanth Jayaraman: vasanth.jayaraman@vibrantsci.com

Karthik Krishna: karthik@vibrantsci.com

Karenah E. Rajasekaran: karenah@vibrant-america.com

Tianhao Wang: tianhao.wang@vibrantsci.com

Kang Bei: kang@vibrantsci.com

John J. Rajasekaran: jjrajasekaran@vibrantsci.com

Keith B Elkon: elkon@uw.edu

ABSTRACT

Introduction

Antibodies reactive with cyclic citrullinated peptides (ACPA) have been included as a part of the ACR criteria for rheumatoid arthritis (RA) since 2010ⁱⁱ. However, due to the inability to capture the polydispersity of antibodies in RA without losing specificity, the ELISA assays cannot be used as definitive markers for diagnosis or prognosis. Current ELISA methods employ at most 30 unique cyclic citrullinated peptides for detection of ACPA. Apart from antibodies to citrullinated peptides, the serum from RA patients also contains antibodies to carbamylated antigens. Citrullinated and Carbamylated peptides hence show great promise for the improved diagnosis of RA. Here we tested whether a novel silicon-based combinatorial high-fidelity and high throughput cyclic peptide microarray that includes both citrullination and carbamylation modifications identifies RA with improved sensitivity compared to standard ELISA assays.

Methods

Using a cyclic modified peptide (CMP) library of more than 2 million sequences we tested 121 clinically diagnosed RA patients and compared binding profiles to corresponding linear epitopes and to commercially available CCP kits. Disease controls and samples from healthy individuals were also used to determine the specificity of the assay. The tests were performed using high throughput liquid handlers integrated with biochip imagers enabling high automation and a rapid turnaround time of 2.5 hrs. Since the citrulline and homocitrulline are very difficult to distinguish except by mass spectrometry, we will use the term Vibrant ACPA to include antibodies that react with either or both of these modified peptides in our assay.

Results

Using our CMP library, the sensitivity of RA detection was 95.04% with a specificity of 95.27% compared to linear peptides (Sensitivity 43.80% and Specificity 96.21%) and to commercial CCP kits (Sensitivity 66.94% and Specificity 89.20%). Inclusion of both citrullinated and carbamylated sequences provided increased sensitivity of testing (from 67% to 95%) as shown in Table 5. Some samples were only positive for carbamylated peptides and some only positive for citrullinated peptide indicating the importance of both peptide modifications.

Conclusions

These novel cyclic modified peptide (CMP) sequences have a high degree of accuracy in differentiating RA from controls, compared with standard serologic ELISA tests. Both citrullinated and carbamylated peptides are necessary for increasing the current sensitivity of RA testing. We also established here that a rigid conformation of the peptide is necessary for improved capture of antibodies by comparing cyclic and linear versions of the same peptide showing improved sensitivity with cyclisation. This high throughput pillar plate platform along with the 2 million data points generated per sample enable immune profiling on an unprecedented scale. While improved diagnostics is the primary outcome presented here, future identification of antigenic peptides that enable better prediction of prognosis and therapy would potentially improve outcomes in the affected population.

INTRODUCTION

The National Arthritis Data Workgroup estimated that rheumatoid arthritis (RA) affects 1.3 million (0.6%) adults in the USⁱ. Women are two to three times more likely to be affected in comparison to men. The diagnosis of rheumatoid arthritis (RA) depends on a constellation of clinical signs and symptoms based on the 2010 American College of Rheumatology-European League Against Rheumatism (ACR-EULAR) classificationⁱⁱ. In this scoring definition, patients are diagnosed with definite RA if they achieve a score of $\geq 6/10$.

Based on 2005 US Medicare/Medicaid data, total annual societal costs of RA were \$39.2 billion. Intangible costs included (\$10.3 billion) quality-of-life deterioration and (\$9.6 billion) premature mortalityⁱⁱⁱ. The early recognition of rheumatoid arthritis (RA) has become a central issue in clinical practice since there is evidence that early treatment improves patient outcomes and retards disease progression by limiting joint destruction and functional disability^{iv}. Although antibodies to a citrulline containing protein, filaggrin was discovered more than 20 years ago in patients with RA^{v,vi} interest has increasingly focused on anti-citrullinated peptide/protein antibodies that serve as an important serological marker in the early diagnosis of RA as well as the progressive course of the disease^{vii}. The 2010 ACR/EULAR (European League Against

Rheumatism) criteria have been specifically prepared to classify and identify patients with early RA who might benefit from early DMARD therapy introducing ACPA as alternative criterion to rheumatoid factorⁱⁱ.

In addition to citrullination (enzymatic deimination of arginine), peptides can be carbamylated by the reaction of cyanate with lysine to form homocitrulline^{viii} which has also been found to be antigenic in RA patients. Although testing for ACPA has contributed significantly to diagnosis of RA, the sensitivity of ACPA as a biomarker for RA is suboptimal and diagnostic criteria require clinical as well as other serologies for a confirmatory diagnosis and guide to prognosis^{ix}. Interestingly, studies have shown that anti-carbamylated antibodies alone could increase sensitivity of detection by up to ~16%^x. Whether the suboptimal sensitivity of ACPA is due to limitations of the antigenic probe set, assay methodology, disease heterogeneity or variation in the antibody response is not clear. In order to improve the sensitivity of detection, attempts have been made to modify citrullinated peptides, with the cyclic forms showing an improved sensitivity compared to linear counterparts^{xi}. By obtaining fixed geometries, cyclic peptides usually bind more efficiently and with higher affinity to their respective receptors since these peptides are conformationally constrained. Moreover, cyclic peptides are often chosen over their linear analogues due to their enhanced enzymatic stability and enhanced membrane permeability, which result in improved bioavailability. They also possess entropic advantages in molecular recognition^{xii}.

Recently, we created a peptide microarray with 110,000 peptides to detect autoantibodies in another autoimmune condition, celiac disease^{xiii}. Patients with celiac disease develop antibodies not only to tissue transglutaminase (tTG), but they also produce antibodies to post translationally modified gliadin peptides (deamidated gliadin peptides or DGP). Subsequently, using *in silico* generated peptides as antigen targets, we identified autoantibodies to tTG/DGP neopeptides in 99% of celiac disease patients with 100% specificity^{xiv}. Employing similar technology, we devised a peptide microarray with more than 2 million citrullinated and carbamylated sequences aiming to improve sensitivity for the diagnosis of RA without loss of specificity.

MATERIALS AND METHODS

Study Cohort

To evaluate the diagnostic utility of the cyclic peptide library, sera were obtained from a cohort comprising a total of 1056 samples: 121 RA samples (with a clinical diagnosis), 407 CCP positive samples (without a clinical diagnosis), 260 disease controls (126 SLE, 54 Scleroderma, 20 Psoriatic arthritis, 20 osteoarthritis, 20 Sjogren's syndrome, and 20 infectious arthritis) and 268 healthy controls. Clinically diagnosed RA, disease controls and healthy volunteers were recruited from the University of Washington (UW, Seattle WA). All individuals signed an informed consent in respective IRB-approved protocols (University of Washington; HSD number 50655). All methods were carried out in accordance with relevant guidelines and regulations for human participants at the University of Washington. Remnant de-identified CCP positive samples were obtained under Western IRB #1-1098539-1 at Vibrant America. The demographic characteristics of the sample cohort is shown in Table 1.

Sample Type	N	Age, mean (Range)	Male %	Female %
Rheumatoid Arthritis (RA)	121	42.8 (20.1 -72.5)	43%	57%
ELISA CCP Positive	407	50.4 (18.7-75.4)	42%	58%
Systemic Lupus Erythematosus (SLE)	126	35.7 (25.7-67.8)	35%	65%
Scleroderma	54	40.3 (20.9-61.5)	39%	61%
Psoriatic Arthritis	20	46.5 (25.2-64.3)	41%	59%
Osteoarthritis	20	50.4 (27.9-76.6)	38%	62%
Sjogren's Syndrome	20	48.1 (22.1-70.4)	38%	62%
Infectious Arthritis	20	39.5 (25.3-60.2)	36%	64%
Healthy Controls	268	35.3 (18.5-69.7)	34%	66%

Table 1. Study Population. The clinical or serologic characteristic, age and gender are shown.

All samples were handled according to standard laboratory procedures and stored at -80°C. All samples were probed using 1:100 primary antibody dilution and 1:2000 secondary antibody dilution (anti-human IgG and anti-human IgA) and scanned on a microarray fluorescence scanner.

Study Approval

The study was conducted under the ethical principles that have their origins in the Declaration of Helsinki. The serum samples used were obtained from the University of Washington and other providers following approval by the appropriate IRB panels as indicated.

Wafer Preparation and Peptide Synthesis

Silicon substrate preparation

Prime-grade 300-mm silicon wafers with p-type boron, (1 0 0) orientation, 1 to 5 $\Omega\cdot\text{cm}^{-1}$, and 725- μm thickness were obtained. The wafers were deposited with 100 nm thermal oxide by dry oxidation. The wafers were etched using a specific photomask design using standard lithography techniques creating a feature area of 100nm silicon dioxide depth and the gap area being bare silicon.

The surface derivatization of the wafers was done starting with an ethanol wash for 5 minutes followed by immersion in 1% by weight APTES in Ethanol for 20-30 minutes to grow the silane layer. The wafers were then cured in a 110°C nitrogen bake oven to grow a mono silane layer with a -NH₂ group to attach a linker molecule. Poly-L-glutamic acid (MW: 50,000-100,000) was activated and coupled to the silane layer to increase the surface density and poly-PEG layer was attached as a linker molecule. The surface derivatization was completed by adding a Fmoc-protected glycine layer as shown in Figure 1.

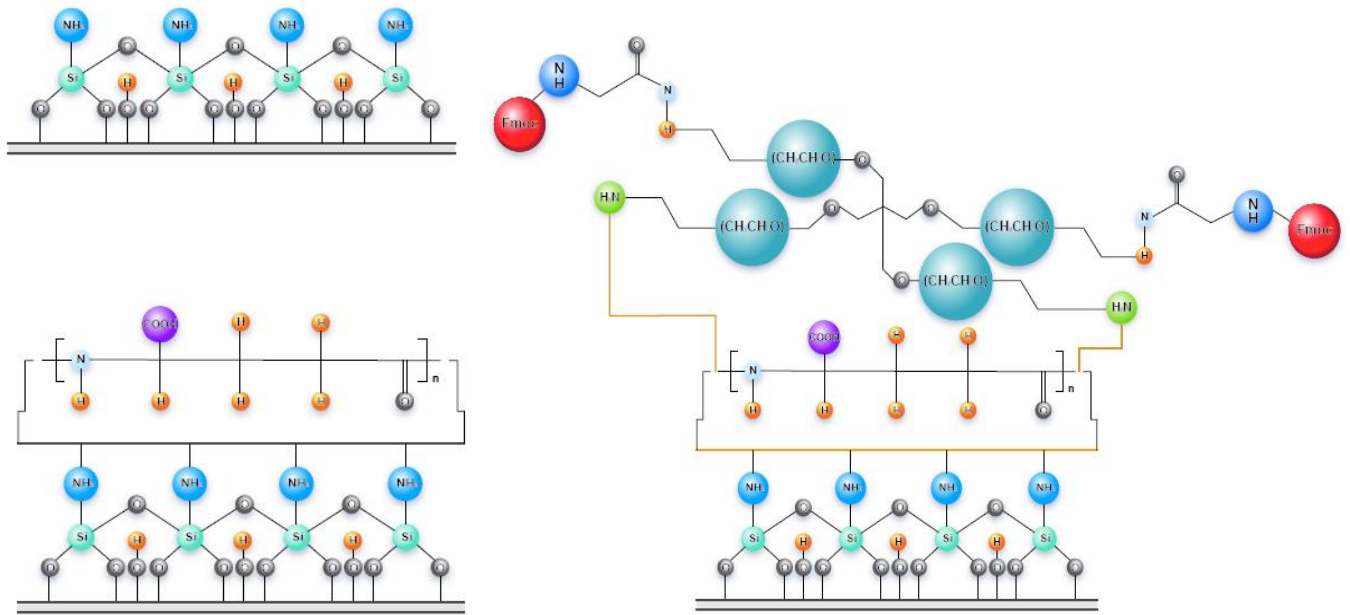


Figure 1 Method of Substrate Preparation. (A) Addition of APTES to grow silane layer. (B) Poly glutamic acid addition to silane layer. (C) Poly PEG layer addition along with Fmoc-glycine to complete surface derivatization. Colors used in the illustration are for molecular representative purposes only and do not hold any significance.

Peptide Array Synthesis

After the substrate preparation the peptides are synthesized as shown in Figure 2. A photoresist composition comprising of 1% by weight of Poly (vinylpyrrolidinone) (PVP) and 5% by weight of piperidine-glyoxylic acid (ionic photobase generator (PBG)) was dissolved in Dimethylformamide (DMF) and spin coated at 4000rpm for 1 minute and soft baked at 70°C for 1 minute in a hot plate. The wafer is then selectively exposed under 365nm UV at a dose of 50 mJ/cm² using a designed photomask wherein the PhotoBase Generator (PBG) converts into the piperidine base in exposed regions. The wafer is then hard baked at 75°C for 2 minutes selectively removing the Fmoc protection in the desired features. The incoming Fmoc-protected amino acid carboxylic acid is pre-activated using TBTU 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylammonium tetrafluoroborate, HOBt Hydroxybenzotriazole and DIEA N,N-Diisopropylethylamine. The mixture was spin coated onto the wafer at 3000rpm for 30 seconds and baked at 65°C for 2 minutes in a hot plate to enable coupling of the Fmoc-protected amino acid to the free amine of the peptide sequence. This was followed up by spin coating a solution of 50 wt% of DMF and 50 wt% of acetic anhydride

to cap any unprotected amines from the coupling reaction. The whole process was repeated for each individual layer of amino acid designed to be coupled to complete the synthesis of all layers of sequences as designed using individual reticles. The step yield of this process is greater than 99.9995% enabling high fidelity of synthesis. This is a key advantage of the photolithographic process which enables superior assay performance with respect to reproducibility in comparison to spotted peptide arrays.

Array Side Chain Deprotection

Following the completion of peptide array synthesis, the side group protection present for amino acids was removed to facilitate biological activity of the peptide sequences. A solution comprising of 95 weight % Trifluoroacetic Acid [TFA] and 5 weight % DI Water was reacted on the wafer for 2 hours to remove the side group protection for all amino acids as applicable.

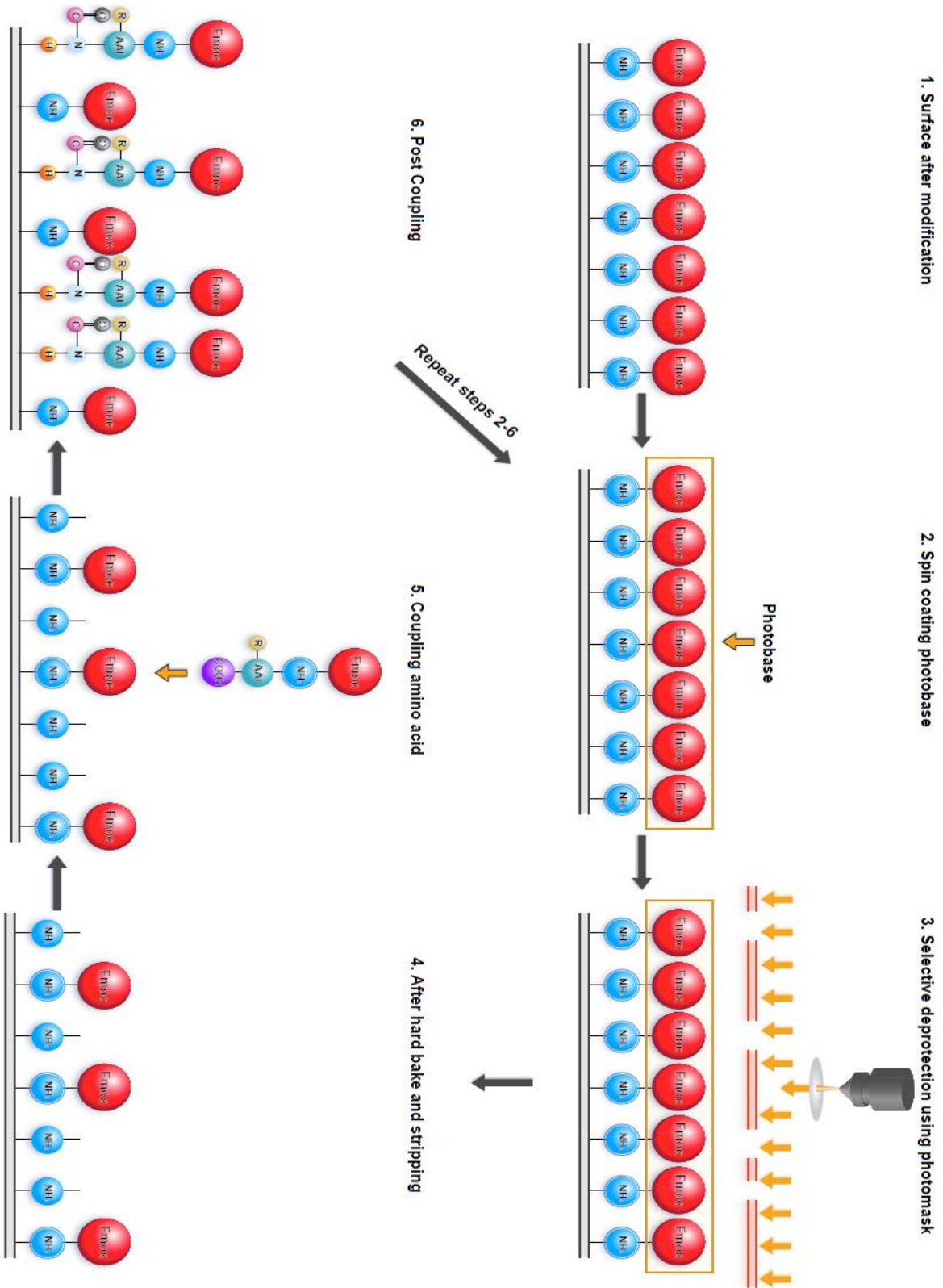


Figure 2 Peptide Synthesis. (1) Fmoc protected glycine base layer for peptide synthesis. (2) Photobase is spin coated to cover the entire wafer surface. (3) Photolithography tool is used to produce photobase at selective spots which will remove the Fmoc protection selectively. (4) Wash step exposes protected and unprotected Glycine layer. (5) Amino acid is coated on entire wafer, coupling to amine NH₂ in unprotected regions. (6) Coupled amino acid at end of single synthesis step.

Cyclization Methods

Various cyclization approaches were tested as shown in Figure 3 to identify the method which produces the highest cyclization yield and the approaches used were as detailed below:

- A) Cyclization was performed with a cysteine coupled as the first layer C-terminus and the last layer N-terminus during peptide synthesis. All peptides were cyclized using cysteine bridge formation under mild oxidative conditions (air oxidation).
- B) Peptides were cyclized using a glutamic acid at the C-terminus and coupled to the N-terminus amine in the peptide chain by reacting the wafer using an amino acid activation cocktail comprising of TBTU/DIEA facilitating the formation of the N->C cyclic peptide bond.
- C) Peptides were cyclized using the method of click chemistry using an azido-lysine in the C-terminus and a propargyl-glycine at the N-terminus. Cyclization was performed by Cu catalyzed alkyne azide cycloaddition yielding 1,4-disubstituted 1,2,3-triazoles covalently linking both termini to form a cyclized peptide.
- D) Cyclization was performed by bridging two primary amines between the N-terminus and a lysine sidechain (C-terminus) by reacting the wafer in a linker solution comprising of disuccinimidyl glutarate.
- E) Cyclization was performed with 2 cysteines at the N- and C-terminus by reacting the sulfhydryl group of two residues with dibromoxylene under mild basic conditions.
- F) Peptides were cyclized using a N-terminal cysteine and a C-terminal thioester using a 2-step N->C cyclization reaction. In the first step, the thiol group places a nucleophilic attack on C-terminal carbonyl moiety, and this is followed by an irreversible intramolecular S->N acyl shift on the newly formed thioester group yielding backbone cyclized peptides.

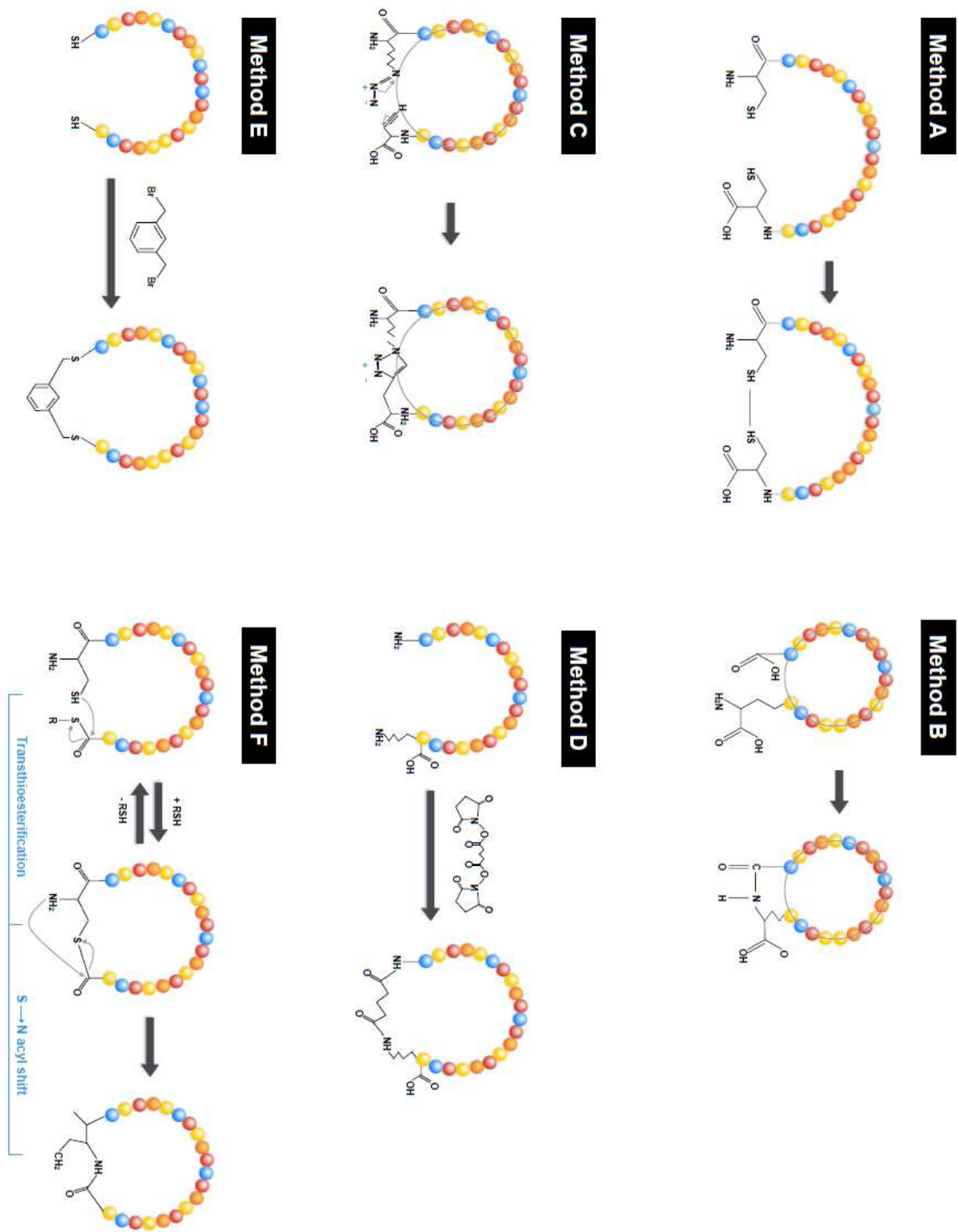


Figure 3 Cyclization Methods. Different chemistries that were used to achieve cyclisation are demonstrated from Method A to Method F as described under Cyclization Methods.

Cyclization yields were determined by testing the activity of C-terminus amino acid using fluorescence before cyclization and after the cyclization reaction to determine the cyclization yield as shown in Table 2. The results indicated that the Method E produced the highest cyclization yield and was utilized in the final peptide array cyclisation.

Cyclization	SNR Linear	SNR Cyclized	Cyclization yield
Method A	61.3	4.8	92.20%
Method B	61.8	7.5	87.90%
Method C	67.3	1.7	97.50%
Method D	71.6	3.1	95.70%
Method E	63.6	0.1	99.80%
Method F	64.9	0.9	98.60%

Table 2 Cyclization Efficiency of different chemistries Method A to Method F

Cyclic Peptide Design

A designed peptide library of RA related protein families including filaggrin, fibrinogen, vimentin, collagen II, enolase, histones, and 14-3-3 eta was synthesized with a lateral shift of 2 amino acids as shown in Figure 4 to bias the 2 million random peptide array sequences (data not shown). The library was constructed with native epitopes, citrullinated epitopes (conversion of the amino acid arginine to citrulline) and carbamylation epitopes (conversion of the amino acid lysine to homocitrulline). The library was designed with a photomask series with 3 replicates of each sequence and synthesized using photolithography.



Figure 4 Sliding library of peptides

A training cohort of RA positive and ELISA CCP positive samples were run on this design library to probe commonly binding subsequences and dominantly occurring amino acids. The learned biases from this library were used to design the final biased random library with 2 million cyclic modified peptides.

Conventional Assays

The following commercial kits were used to compare results of the Vibrant assays: For RF, Beckman Coulter (REF: 447070) and CCP Inova Diagnostics (REF: 704550).

Data Analysis

To identify the optimal peptide epitopes from the library, the sample cohort from Table 1 were run on the peptide microarray. Using the fluorescence microarray scanner, immunoassay binding activity was scanned. The scanned immunoassay binding data were mapped using a ROI analysis and converted into raw binding fluorescent intensity. Subsequently, the intensity was normalized from the spots where no epitopes were synthesized using a least variant set method of normalization algorithm.

Using historic data from previous microarrays, a random forest classifier was trained to detect spots which were not reproducible between replicates of each individual sequence. A random forest classifier is a machine learning procedure, which learns from existing data and applies them to new unclassified data sets. All epitopes not within the 95% linear regression confidence are removed from the data analysis.

The receiver operating characteristic (ROC) curve was then applied for each sequence to identify the optimal threshold value for obtaining the highest sensitivity and specificity. Overall sensitivity and specificity were calculated based on the condition that a sample is considered to be positive if at least 5 unique epitopes have normalized units greater than the threshold chosen for the corresponding sequence.

Statistical Analysis

McNemar's chi-square test was used to compare the sensitivity of paired data of Vibrant ACPA to combined RF and CCP assays. The complete sensitivity and specificity comparisons are shown in Table 5. The same set of samples were run on each individual method to obtain comparable values for sensitivity and specificity.

RESULTS

Analytical Studies

Analytical accuracy of the peptide array was validated using the spike/recovery method. Individual sets of 200 peptide sequences were synthesized commercially and compared with the corresponding microarray sequence. Five different dilutions of the commercial peptide labeled with a fluorescent dye was spiked on a sample with high antibody titers and competed with the corresponding array sequence. The results summarized in table 3 suggest a systematic pattern of lower recovery trends for all sequences at higher spiked concentrations.

SEQ ID	Spike 1	Spike 2	Spike 3	Spike 4	Spike 5
SEQ1-20	91%	66%	45%	20%	3%
SEQ21-40	90%	63%	35%	16%	5%
SEQ41-60	93%	61%	42%	15%	5%
SEQ61-80	88%	58%	29%	18%	4%
SEQ81-100	96%	70%	32%	14%	3%
SEQ101-120	93%	73%	32%	16%	4%
SEQ121-140	90%	75%	32%	8%	3%
SEQ141-160	85%	67%	33%	18%	2%
SEQ161-180	98%	70%	45%	18%	5%
SEQ181-200	90%	63%	41%	19%	1%

Table 3 Peptide Spike and Recovery. Results were calculated by spiking a known relative amount of commercial peptides (at 5 different concentrations) into the serum sample matrix which contains high titer of antibody to each individual analyzed peptide. The assay is run to measure the response (recovery) titer of the spiked sample matrix compared to the original titer which is run using the standard diluent.

Analytical precision of the peptide array was determined by running a panel of 10 analytical samples which includes samples across the analytical measuring range (AMR) over 12 days with 5 replicates run in each day. The coefficient of variation (CV) across replicates within the chip and across multiple days was analyzed and summarized in table 4.

Sample	Within chip CV	Day-to-day CV	Average Overall CV	Maximum Overall CV
Sample 1	2.3%	3.6%	3.4%	9.5%
Sample 2	1.6%	1.4%	5.4%	11.8%
Sample 3	1.9%	3.0%	4.0%	11.3%
Sample 4	2.8%	6.9%	4.7%	11.2%
Sample 5	3.0%	1.8%	4.2%	11.8%
Sample 6	2.1%	5.8%	5.2%	9.1%
Sample 7	1.0%	4.4%	3.3%	9.0%
Sample 8	2.1%	2.1%	4.4%	9.5%
Sample 9	2.5%	4.2%	5.9%	10.4%
Sample 10	2.5%	2.8%	3.7%	9.0%

Table 4 Extent of variability in repeated assays of the peptide array expressed as the coefficient of variation (CV)

Alanine scanning mutagenesis

Detailed characterization of antibody-antigen interactions is necessary for accurate diagnosis of autoimmune conditions. Antibody recognition of *in situ* synthesized peptides antigens at the amino acid level can be verified using a synthesis technique called alanine scanning. Briefly, peptide epitopes of a set of monoclonal antibodies were grown on the microarray platform along with mutants for each individual amino acid along the epitope sequence. The mutant library was synthesized by replacing each amino acid (AA) with alanine one AA at a time. Alanine has an inert methyl functional group and can help understand the binding contribution of the corresponding amino acid it replaces. The peptide mutants for each epitope were thus synthesized and reacted with the monoclonal to identify the key amino acids contributing to the binding interaction. This enabled validation of the antibody recognition of the in-situ synthesized peptides at the amino acid level and the data is presented under supplementary material I.

Clinical Sensitivity and Specificity

The study cohort of 1056 samples as shown in Table 1 were used to test the diagnostic utility of this cyclic modified peptide library (Vibrant ACPA) and the corresponding linear peptide library (ALPA). The samples were also tested on the commercial platform's RF (Beckman Coulter) and CCP (Inova Diagnostics). As summarized in Table 5 we observed that the clinical sensitivity was 95.04% and the clinical specificity was 95.27% for the Vibrant ACPA assay. Individual Citrullinated peptide library and Carbamylated peptide library sensitivities and specificities are also shown. The effect of this improved assay performance can also be seen reflected in the positive and negative predictive values as shown below.

Method	N Pos	N Neg	Sensitivity (95% confidence)	Specificity (95% confidence)	PPV	NPV
RF BECKMAN	121	528	67.8% (59%-75%)	92.1% (89%-94%)	66.1%	92.6%
CCP INOVA	121	528	66.9% (58%-75%)	89.2% (86%-92%)	58.7%	92.2%
RF + CCP	121	528	76.9% (69%-83%)	88.6% (86%-91%)	60.8%	94.4%
Vibrant ALPA	121	528	43.8% (35%-53%)	96.2% (94%-98%)	72.6%	88.2%
Vibrant CCP	121	528	90.1% (83%-94%)	96.0% (94%-97%)	83.9%	97.7%
Vibrant CarP	121	528	83.5% (76%-89%)	95.8% (94%-97%)	82.1%	96.2%
Vibrant ACPA	121	528	95.0% (90%-98%)	95.3% (93%-97%)	82.1%	98.8%

Method Comparison (Clinical Sensitivity)	N	McNemar's Chi-squared statistic	Degrees of freedom	p-value
Vibrant ACPA vs RF	121	31	1	<0.001
Vibrant ACPA vs CCP	121	32	1	<0.001
Vibrant ACPA vs RF+CCP	121	20	1	<0.001

Method Comparison (Clinical Specificity)	N	McNemar's Chi-squared statistic	Degrees of freedom	p-value
Vibrant ACPA vs RF	528	15	1	<0.001
Vibrant ACPA vs CCP	528	30	1	<0.001
Vibrant ACPA vs RF+CCP	528	33	1	<0.001

Table 5. Sensitivity and Specificity of each assay used in this study, method comparison with McNemar's Chi squared statistic for Clinical Sensitivity and Clinical Specificity with RF and CCP assays.

Citrullination Vs Carbamylation

Both Citrulline and homocitrulline are nonstandard amino acids with the presence of an ureido group. While citrulline is formed by enzymatic deimination of arginine, peptides can be carbamylated by the reaction of cyanate with lysine to form homocitrulline. Since we essentially grew the same peptide sequence for both citrulline and carbamylation modifications it was interesting to note that for a majority of patients the antibodies bound to both the sequences. Both citrulline and carbamylation favored the amino acids E, Q, S, G, R and T in closer proximity than the other amino acids (data not shown).

DISCUSSION

Autoantibodies against citrullinated peptides have emerged as biomarkers for the diagnosis of rheumatoid arthritis and are now routinely utilized as a part of diagnosis and disease management^{xv,vi}. Though these markers are very specific, sensitivity of disease detection is suboptimal and there is conflicting literature comparing anti-CCP2 and anti-CCP3 assays^{xvi,xvii,xviii}. Knowing that diverse proteins are implicated as antigens^{xix} and cyclization of post-translationally modified peptides improves antibody recognition^{xx}, we have developed powerful tools to create a vast library of cyclic citrullinated and carbamylated peptides that markedly improve the sensitivity of RA diagnosis. These novel peptides demonstrated a significantly improved sensitivity for the diagnosis of RA while maintaining specificity compared to the standard assays for CCP and RF combined as shown in Table 5.

The CMP microarray presented here paves way for the generation of novel cyclic peptides that can be tested for diagnosis, prognosis and, potentially, for monitoring in response to treatment. To synthesize millions of such peptides in a cost-effective approach to make it applicable in a healthcare setting, we employed the silicon platform to grow a diverse group of cyclic molecules. We first confirmed that, compared to linear epitopes, efficient cyclization plays an important role in improving the identification of ACPA antibodies in RA due to conformational stability. These cyclic epitopes enable diagnosis of RA with a 95.04% sensitivity and 95.27% specificity, which is superior to existing assays. The diversity of peptides

and the use of both citrullination and carbamylation are also key to capture the polyclonal response among individuals diagnosed with RA.

In a previous study, ACPA has been shown to be a more sensitive and better predictive of erosive disease in comparison to rheumatoid factor with the combined sensitivity and specificity of 67% and 95% respectively^{xxi}. Moreover, a meta-analysis of 151 studies shows sensitivity and specificity of ACPA to be 57% and 96% respectively (the analysis refers only to 15 relevant cohort studies)^{xxii} and being the preferred biomarker for early RA detection. ACPA appears to be a reliable predictor of erosive RA, making it a potentially important prognostic tool that might be used to direct patient management decisions. Other studies have also found that ACPA can be detected prior to the onset of arthritis and also RA patients with recent onset arthritis^{vi,xxiii}. The time interval from the onset of the first symptoms to the fulfillment of the classification criteria seems to be directly related to the initial ACPA level^{xxiv}. In the future, it will be interesting to determine whether individual peptide specificities relate to early disease or to development of erosions. In addition to its potential for early diagnosis, prognosis and exploration of pathogenesis, the CMP microarray described here may be broadly useful for identifying lead compounds for drug discovery and for the development of new peptide antigens.

In summary we present here a novel cyclic modified peptide microarray for the diagnosis of RA. The development of this high fidelity and high throughput method will be an important tool for improved diagnosis of this disease with the potential for application to prognosis and monitoring of treatments in the future.

ACKNOWLEDGMENTS

Research reported in this publication was supported by Vibrant America LLC. The content is solely the responsibility of the authors and does not necessarily represent the official views of the FDA. We would like to thank Miaomiao Sun and Ashish Gopalakrishnan for generating the figures.

Author Contributions

Conceived and designed the experiments: HKK VJ KK. Performed the experiments: TW KK. Analyzed the data: KK KER VJ TW. Contributed reagents/materials/analysis tools: TW KB KE. Wrote the paper: HKK KE VJ. Revised the manuscript critically for important intellectual content: JJR VJ HKK KE

Funding Statement

Vibrant America provided funding for this study in the form of salaries for authors [HKK VJ KK TW KER KB]. The specific roles of these authors are articulated in the 'author contributions' section. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests Statement

The authors have read the journal's policy and the authors of this manuscript have the following competing interests. All of the authors listed in this paper are employees of Vibrant Sciences or Vibrant America, with the exception of KE who an academic research collaborator. Vibrant America is a clinical laboratory performing commercial diagnostic testing.

SUPPLEMENTARY MATERIAL I

Alanine Scan Mutagenesis

Ab7757 – Epitope QETFSDLWKL

Sequence			Q	E	T	F	S	D	L	W	K	L
QETFSDLWKL		A	64580.57	58661.3	62717.8	6181.63	2458.72	757.62	13237.89	44002.07	533.95	4490.52
	R e p l a c e d	C	56642.02	46366.77	55624.7	5772.76	419.02	588.77	765.2	64531.75	416.19	918.21
		D	57910.02	42323.18	43945.61	16700.98	825.08	62529.72	480.83	65520	743.16	1097.11
		E	63590.38	62529.72	63621.04	39843.06	563.35	818.36	462.28	65153.95	1116.79	1062.01
		F	59588.16	32905.11	27980.83	62529.72	596.44	958.43	856.81	65394.11	920.33	2748.04
		G	63892.13	60385.37	34629.65	4875.1	450.47	705.53	1442.2	65359.9	607.15	12584.61
		H	56641.28	49388.64	57169.07	7375.77	903.34	528.17	572.75	58398.53	536.93	1326.38
		I	62357.69	45170.68	61817.86	2221.83	1073.14	985.79	2721.55	64241.53	892.59	18615.34
		K	60070.2	45780.81	59306.78	5349.09	650.42	644.34	579.65	43408.68	62529.72	1250.36
		L	34599.79	27517.48	59321.75	12448.58	689.58	884.02	62529.72	57441.58	857.19	62529.72
		M	38293.72	48560.92	64751.34	17428.99	841.15	756.26	546.5	59368.24	1104.69	54964.96
	A m i n o	N	61265.81	60777.32	60405.38	5747.61	395.09	575	1013.23	65455.5	585.46	1403.89
		P	62556.88	55684.28	62724.84	8300.55	439.14	629.18	593.84	57250.85	716.59	697.93
		Q	62529.72	62596.75	64477.64	9572.96	567.47	575.39	527.45	65496.04	870.64	2639.33
		R	62754.04	42562.88	64090.17	3540.26	747.93	821.11	664.57	48134.59	2971.14	1167.23
		S	53954.97	60332.16	38741.22	8907.57	62529.72	665.55	860.4	65405.93	758.43	2778.37
	A c i d	T	64607.85	56468.01	62529.72	8647.04	2954	672.1	558.96	65326.07	494.74	1587.93
		V	60964.55	46130.32	62993.28	10159.47	914.97	793.15	1587.44	64053.43	739.49	1833.09
		W	55211.87	26206.49	57045.52	7019.09	567.61	561.93	672.61	62529.72	531.95	668.07
		Y	63588.12	30012.82	62796.06	55362.93	326.19	746.15	500.73	65053.4	447.83	567.46

Ab80547 – Epitope RKNERALPLPQSSTC

Sequence			K	N	E	R	A	L	P	L	P	Q	S	S
RKNERALPLPQSSTC		A	57713.67	8606.36	47678	25886.54	34213.53	2939.74	8489.2	888.36	1860.03	1185.38	692.7	44418.48
	R e p l a c e d	C	34064.78	28254.28	29911.08	49816.06	6997.22	10644.97	3364.27	5775	1259.56	562.5	939.46	2220.7
		D	43681.04	13864.52	27874.47	61426.53	42367.86	4634.82	1908.43	4281.06	991.79	980.25	590.03	1321.41
		E	47626.01	30659.11	34213.53	57136.31	58168.97	6571.58	5488.87	2578.53	918.96	1484.3	1400.48	29345.27
		F	51530.61	34656.67	56050.91	46059.6	26312.88	27804.58	10097.01	8563.95	1237.25	1278.34	2234.87	14909.65
		G	60111.54	30045.06	39810.17	27147.25	32844.72	4963.35	5787.24	1195.41	841	1061.62	1624.56	18523.65
		H	50875.1	12900.62	20366.83	52443.27	18338.89	1820.51	2386.87	772.73	731.75	686.58	1784.17	3274.72
		I	42939.04	2509.07	45150.4	43181.2	12528.05	27205.14	2925.48	5944.01	4129.92	1030.43	849.77	6879.41
		K	34213.53	10425.01	32761.42	37038.79	22029.85	3567.81	1777.66	936.8	1018.85	2314.37	1531.62	6880.08
		L	51535.38	18462.51	17788.74	45850.84	24580.63	34213.53	5468.6	34213.53	377.58	1500.28	1407.76	7082.56
		M	22996.53	24145.21	51734.19	33115.3	2017.26	2018.17	2785.81	756.95	1021.82	2587.04	4730.23	
	A m i n o	N	27406.5	34213.53	62923.91	43138.34	17856.4	7890.07	2164	2121.47	1081.09	369.64	1063.31	5445.49
		P	28270.41	34640.83	48982.83	59561.61	32564.26	7307.48	2260.96	798.5	1180.35	34213.53	1213.47	24794.86
		R	38615.85	45586.38	37341.32	34213.53	24644.2	9709.54	1270.66	569.62	484.47	904.68	823.83	9979.92
		S	29757.02	25396.51	29804.02	51676.77	22381.85	1836.92	1756.52	2957.3	768.07	967.07	34213.53	34213.53
		T	39879.88	31375.65	23477.02	7471.8	33395.46	2760.45	1643.12	634.33	599.69	1075.45	3676.83	2936.08
	A c i d	V	57316.46	60183.2	31608.32	54673.67	58837.08	4448.47	2237.5	4818.93	1118.73	1044.01	1227.49	14281.82
		W	52553.93	20797.58	17068.36	62207.3	6792.03	7480.19	2280.53	3148.06	1213.4	2033.93	1347.87	2075.16
		Y	63685.66	21109.57	37784.92	45281.56	32123.48	57728.55	2020.69	910.02	465.44	469.92	1008.08	30032.45

Ab35977 – Epitope KPLEEVLNL

Sequence			K	P	L	E	E	V	L	N	L
KPLEEVLNL			61837.44	3631.74	543.12	1949.12	1	1431.55	535.1	61446.31	62718.07
	R e p l a c e d	A	61837.44	3631.74	543.12	1949.12	1	1431.55	535.1	61446.31	62718.07
		C	62403.69	887.92	797.77	895.22	8391.57	4239.2	1067.76	827.61	64611.07
		D	63270.31	242.64	615.16	1458.28	26262.91	6359.06	1307.04	57050.31	62063.24
		E	63222.49	886.51	589.73	57890.97	57890.97	29206.17	1104.94	49722.94	64912.6
		F	63526.12	1004.34	26556.11	4573.29	1	22567.79	3524.05	59678.59	59136.58
		G	1	429.36	647.13	576.44	15712.98	2345.62	800.85	62397.47	62013.27
		H	65386.92	590.56	820.99	655.43	3241.71	3458.56	722.79	56893.69	42448.77
		I	60475.87	3349.9	15204.19	2066.74	42467.21	65151.21	7980.47	63613.2	61600.68
		K	57890.97	570.81	836.8	731.1	29090.12	16724.3	2083.83	53722.25	63342.33
		L	52950.98	994.39	57890.97	1138.85	33055.71	58491.1	57890.97	58916.4	57890.97
		M	62449.7	1168.84	1649.93	2861.71	45379.04	11111.53	3934.67	62501.39	63782.23
		N	63723.58	2024.68	721.84	1394.24	54738.85	1021.88	505.28	57890.97	59622.55
		P	1	57890.97	568.99	4353.06	13025.96	770.01	477.97	20729.6	1
		Q	64544.12	887.39	392.2	25482	45956.36	15869.67	551.49	59766.64	61089.67
		R	64864.86	520.11	556.11	594.02	22440.03	22886.58	522.3	63817.87	60425.2
		A c c i d	S	63598.8	1569.33	721.16	731.78	14268.29	6689.45	646.82	59904.36
	T		64724.96	3762.16	514.88	1153.16	29858.19	46188.89	659.98	61307.94	62195.17
	V		1	25639.26	1152.83	3145.46	61994.14	57890.97	6159.16	53043.26	61569.18
	W		63994.18	1081.45	1911.2	1866.68	5438.58	1106.56	961.33	1	59259.77
	Y		61522.54	2074	276.56	9242.78	37917.5	6792.33	788.6	40333.08	57441.88

Ab130275 – Epitope YPYDVPDYA

Sequence			Y	P	Y	D	V	P	D	Y	A	
YPYDVPDYA			64563.07	65109.85	63839.84	2278.55	63192.27	44494.04	12626.71	64432.46	51950.62	
	R e p l a c e d	A	64563.07	65109.85	63839.84	2278.55	63192.27	44494.04	12626.71	64432.46	51950.62	
		C	64937.28	62969.21	60871.74	8914	8586.4	2451.47	7174.33	60638.1	65311.16	
		D	64206.37	62786.94	50607.93	51950.62	12120.39	2686.91	51950.62	63568.73	1	
		E	63567.74	63560.63	49794.19	6056.68	11633.83	5862.77	58334.84	62755.45	64839.23	
		F	64624.4	62831.82	61115.88	2353.2	3672.84	1852.14	5371.72	65407.74	65383.71	
		G	63783.44	60645.84	23000.29	2581.84	30559.28	9786.91	7869.01	59155.72	64976.59	
		H	65385.95	62657.73	54385.27	2898.95	5700.38	4031.91	9670.87	59940.8	64560.1	
		I	65385.77	65163.64	58764.1	1603.59	61786.57	6176.73	6464.04	64200.46	64673.18	
		K	1	63853.08	60545.99	2367.9	1777.2	5371.06	3728.05	60231.61	64063.13	
		L	63803.94	64834.9	64525.08	2889.62	5251.42	3532.07	4772.88	63662.47	64452.2	
		A m i n o	M	1	63624.4	61892.84	3452.53	5439.97	6060.33	8973.41	61593.02	64738.51
			N	65319.02	64276	50652.08	1769.75	6617.61	8054.6	45183.46	64645.33	64931.82
			P	65227.57	51950.62	21980.48	1372.33	2985.69	51950.62	4779.25	62247.73	61999.02
			Q	65006.85	65007.06	38670.57	1723.02	12160.82	20475.61	13059.3	63731.99	64512.44
			R	64705.24	63445.57	61767.31	1380.97	2334.31	2877.83	4269.43	62519.18	64881.63
		A c c i d	S	64765.69	63907.83	54711.19	2521.41	29083.24	27349.23	7760.37	65173.65	62902.55
	T		64428.93	64452.15	31977.09	2178.57	51296.88	1	5668.63	1	64928.03	
	V		64814	63585.39	44519.84	2531.21	51950.62	7883.3	8060.59	64570.85	64780.31	
	W		65279.72	64373.27	57699.42	2531.22	6014.7	2720.05	7150.29	54846.13	1	
	Y		51950.62	63864.71	51950.62	2441.34	6144.55	2376.05	4457.4	51950.62	64890.18	

REFERENCES

- ⁱ Lawrence, R. C., Felson, D. T., Helmick, C. G., Arnold, L. M., Choi, H., Deyo, R. A., Gabriel, S., Hirsch, R., Hochberg, M. C., Hunder, G. G., Jordan, J. M., Katz, J. N., Kremers, H. M., Wolfe, F., & National Arthritis Data Workgroup (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis and rheumatism*, 58(1), 26–35. <https://doi.org/10.1002/art.23176>
- ⁱⁱ Aletaha, D., Neogi, T., Silman, A. J., Funovits, J., Felson, D. T., Bingham, C. O., 3rd, Birnbaum, N. S., Burmester, G. R., Bykerk, V. P., Cohen, M. D., Combe, B., Costenbader, K. H., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J. M., Hobbs, K., Huizinga, T. W., Kavanaugh, A., Kay, J., ... Hawker, G. (2010). 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis and rheumatism*, 62(9), 2569–2581. <https://doi.org/10.1002/art.27584>
- ⁱⁱⁱ Howard Birnbaum, Crystal Pike, Rebecca Kaufman, Maryna Maynchenko, Yohanne Kidolezi & Mary Cifaldi (2010) Societal cost of rheumatoid arthritis patients in the US, *Current Medical Research and Opinion*, 26:1, 77-90, DOI: [10.1185/03007990903422307](https://doi.org/10.1185/03007990903422307)
- ^{iv} Van der Helm-van Mil AH, le Cessie S, van Dongen H, Breedveld FC, Toes RE, Huizinga TW (2007) A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis. How to guide individual treatment decisions. *Arthritis Rheum* 56:433–440
- ^v Tan EM, Smolen JS. Historical observations contributing insights on etiopathogenesis of rheumatoid arthritis and role of rheumatoid factor. *J Exp Med*. 2016;213(10):1937-1950. doi:10.1084/jem.20160792
- ^{vi} Holers VM. Autoimmunity to citrullinated proteins and the initiation of rheumatoid arthritis. *Curr Opin Immunol*. 2013 Dec;25(6):728-35. doi: 10.1016/j.coi.2013.09.018. Epub 2013 Nov 8. PMID: 24215742; PMCID: PMC3895448.
- ^{vii} Wiik AS, van Venrooij WJ, Pruijn GJ (2010) All you wanted to know about anti-CCP but were afraid to ask. *Autoimmun Rev* 10:90–93
- ^{viii} Pruijn G. J. (2015). Citrullination and carbamylation in the pathophysiology of rheumatoid arthritis. *Frontiers in immunology*, 6, 192. <https://doi.org/10.3389/fimmu.2015.00192>
- ^{ix} Lee DM, Schur PH. Clinical utility of the anti-CCP assay in patients with rheumatic diseases *Annals of the Rheumatic Diseases* 2003;62:870-874.
- ^x Rigby, W., Skopelja-Gardner, S., & Jones, J. D. (2017). Editorial: Anti-Citrullinated Protein Antibody, Anti-Carbamylated Protein Antibody, and Rheumatoid Arthritis: Azurophilic Granules Sing the Blues. *Arthritis & rheumatology (Hoboken, N.J.)*, 69(12), 2251–2255. <https://doi.org/10.1002/art.40238>
- ^{xi} Schellekens, G. A., Visser, H., de Jong, B. A., van den Hoogen, F. H., Hazes, J. M., Breedveld, F. C., & van Venrooij, W. J. (2000). The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis and rheumatism*, 43(1), 155–163. [https://doi.org/10.1002/1529-0131\(200001\)43:1<155::AID-ANR20>3.0.CO;2-3](https://doi.org/10.1002/1529-0131(200001)43:1<155::AID-ANR20>3.0.CO;2-3)
- ^{xii} J. Vagner, H. Qu and V. J. Hruby, *Curr. Opin. Chem. Biol.*, 2008, 12, 292–296
- ^{xiii} Choung, R. S., Marietta, E. V., Van Dyke, C. T., Brantner, T. L., Rajasekaran, J., Pasricha, P. J., Wang, T., Bei, K., Krishna, K., Krishnamurthy, H. K., Snyder, M. R., Jayaraman, V., & Murray, J. A. (2016). Determination of B-Cell Epitopes in Patients with Celiac Disease: Peptide Microarrays. *PLoS one*, 11(1), e0147777. <https://doi.org/10.1371/journal.pone.0147777>

-
- ^{xiv} Choung, R. S., Khaleghi Rostamkolaei, S., Ju, J. M., Marietta, E. V., Van Dyke, C. T., Rajasekaran, J. J., Jayaraman, V., Wang, T., Bei, K., Rajasekaran, K. E., Krishna, K., Krishnamurthy, H. K., & Murray, J. A. (2019). Synthetic Neopeptides of the Transglutaminase-Deamidated Gliadin Complex as Biomarkers for Diagnosing and Monitoring Celiac Disease. *Gastroenterology*, *156*(3), 582–591.e1. <https://doi.org/10.1053/j.gastro.2018.10.025>
- ^{xv} Van Venrooij WJ, van Beers JJ, Pruijn GJ. Anti-CCP antibodies: the past, the present and the future *Nat Rev Rheumatol*. 2011;7:391–398
- ^{xvi} Szekanecz Z, Szabó Z, Zeher M, et al. Superior performance of the CCP3.1 test compared to CCP2 and MCV in the rheumatoid factor-negative RA population. *Immunol Res*. 2013;56(2-3):439-443. doi:10.1007/s12026-013-8425-8
- ^{xvii} Malher M, Bentow C, Albesa R, et al. Comparison of CCP2 and CCP3 assays in a large cohort of established rheumatoid arthritis and controls. *Annals of the Rheumatic Diseases* 2018;77:A24
- ^{xviii} Sieghart D, Konrad C, Swiniarski S, Haslacher H, Aletaha D, Steiner G. Diagnostic Performance of Anti-cyclic Citrullinated Peptide (CCP) 2 and CCP3.1 Assays in Early Rheumatoid Arthritis [abstract]. *Arthritis Rheumatol*. 2019; 71 (suppl 10)
- ^{xix} Demoruelle MK, Deane K. Antibodies to citrullinated protein antigens (ACPAs): clinical and pathophysiologic significance. *Curr Rheumatol Rep*. 2011 Oct;13(5):421-30. doi: 10.1007/s11926-011-0193-7. PMID: 21713412; PMCID: PMC4095867.
- ^{xx} Dorow DS, Shi PT, Carbone FR, Minasian E, Todd PE, Leach SJ. Two large immunogenic and antigenic myoglobin peptides and the effects of cyclisation. *Mol Immunol*. 1985; 22: 255–264
- ^{xxi} Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S et al (2007) Meta-analysis: diagnostic accuracy of anticyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Ann Intern Med* 146:797–808
- ^{xxii} Whiting PF, Smidt N, Sterne JA, Harbord R, Burton A, Burke M et al (2010) Systematic review: accuracy of anticitrullinated peptide antibodies for diagnosing rheumatoid arthritis. *Ann Intern Med* 152:456–464
- ^{xxiii} Tan, E. M., & Smolen, J. S. (2016). Historical observations contributing insights on etiopathogenesis of rheumatoid arthritis and role of rheumatoid factor. *The Journal of experimental medicine*, *213*(10), 1937–1950. <https://doi.org/10.1084/jem.20160792>
- ^{xxiv} Bizzaro N, Bartoloni E, Morozzi G, Manganelli S, Riccieri V, Sabatini P et al (2013) Anti-cyclic citrullinated peptide antibody titer predicts time to rheumatoid arthritis onset in patients with undifferentiated arthritis: results from a 2-year prospective study. *Arthritis Res Ther* 15:R16

Figures

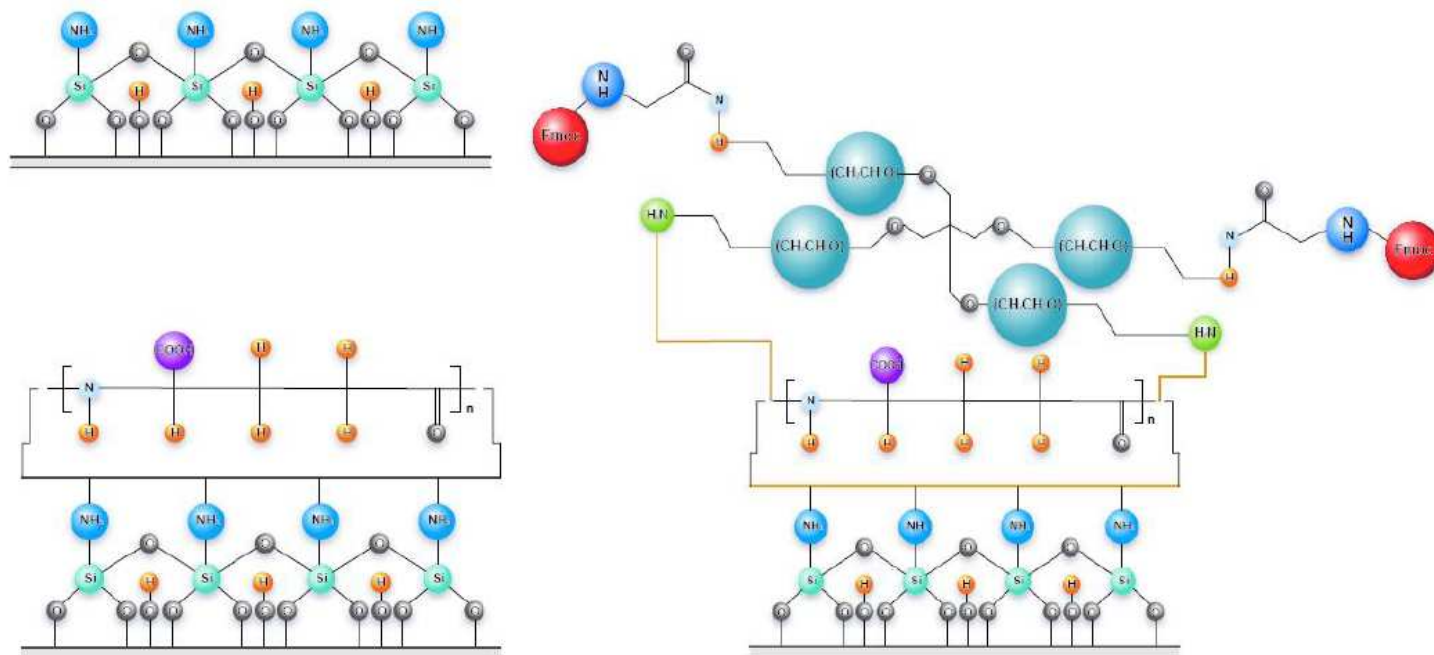


Figure 1

Method of Substrate Preparation. (A) Addition of APTES to grow silane layer. (B) Poly glutamic acid addition to silane layer. (C) Poly PEG layer addition along with Fmoc-glycine to complete surface derivatization. Colors used in the illustration are for molecular representative purposes only and do not hold any significance.

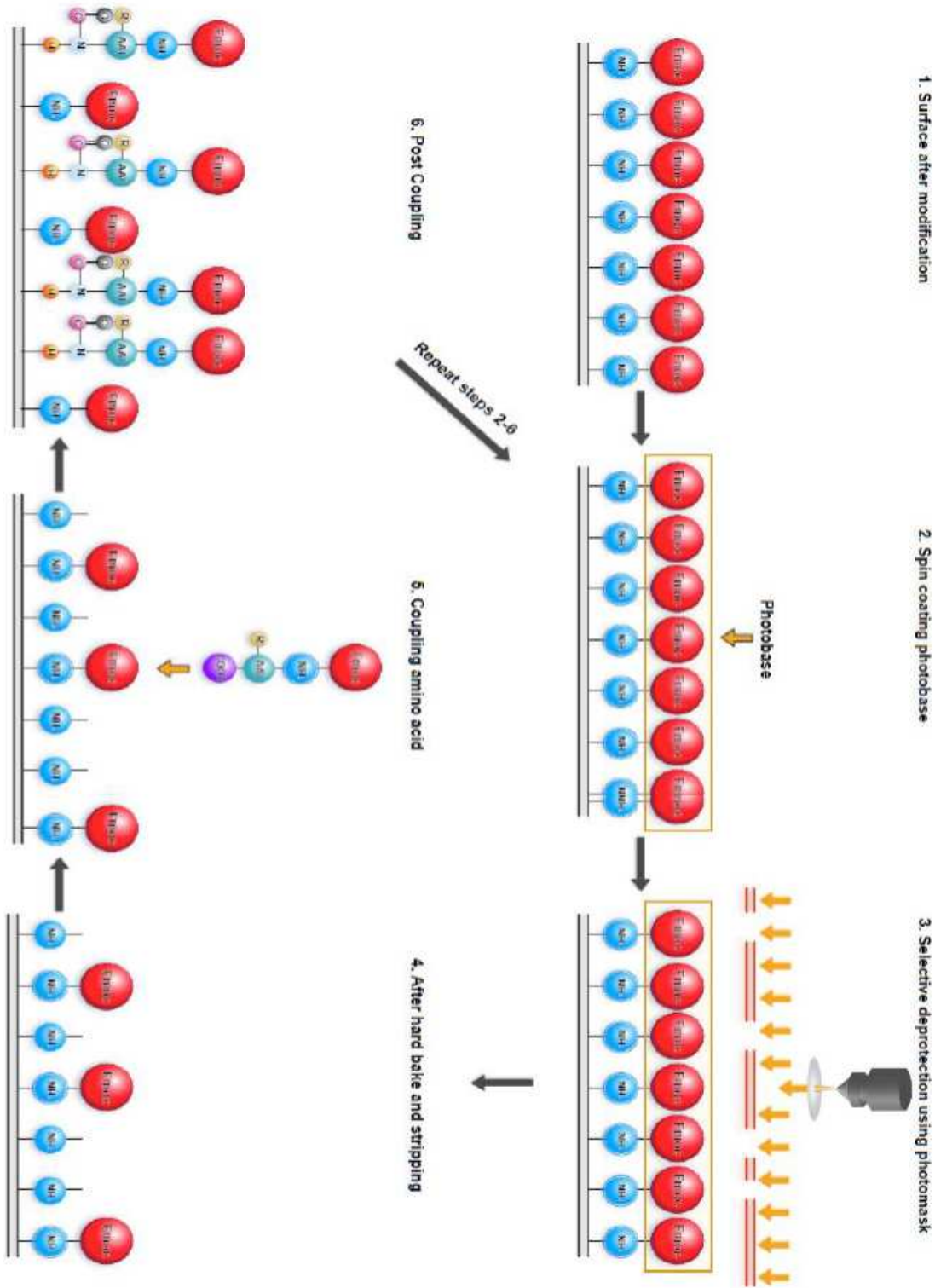


Figure 2

Peptide Synthesis. (1) Fmoc protected glycine base layer for peptide synthesis. (2) Photobase is spin coated to cover the entire wafer surface. (3) Photolithography tool is used to produce photobase at selective spots which will remove the Fmoc protection selectively. (4) Wash step exposes protected and unprotected Glycine layer. (5) Amino acid is coated on entire wafer, coupling to amine NH_2 in unprotected regions. (6) Coupled amino acid at end of single synthesis step.

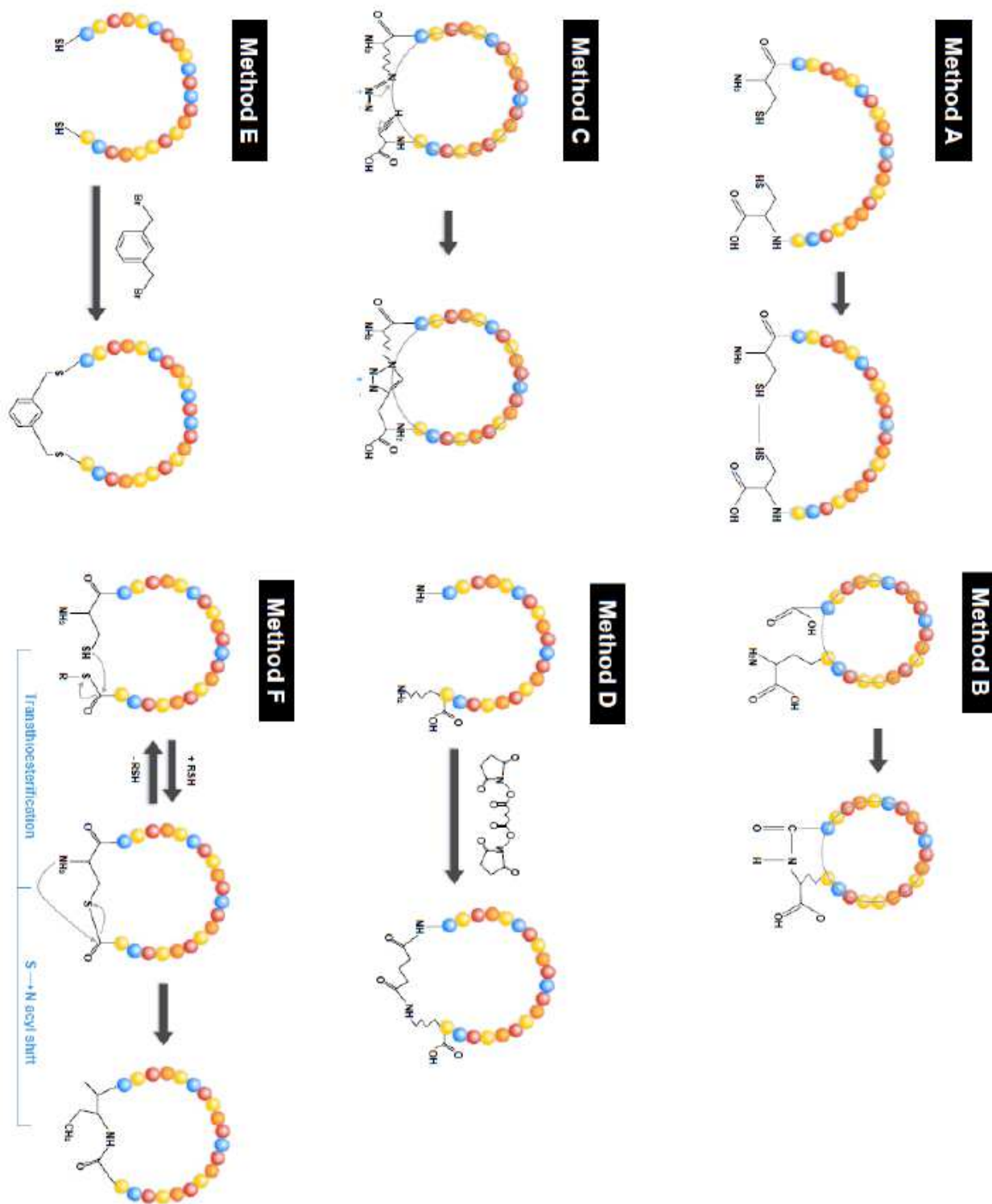


Figure 3

Cyclization Methods. Different chemistries that were used to achieve cyclisation are demonstrated from Method A to Method F as described under Cyclization Methods.

A C D E F G H I K L M N P Q R S T V W Y

Length: 8
Offset: 2

A C D E F G H I

D E F G H I K L

F G H I K L M N

H I K L M N P Q

K L M N P Q R S

M N P Q R S T V

P Q R S T V W Y

Figure 4

Sliding library of peptides