Molecular modelling of epitopes recognized by neoplastic B lymphocytes in Chronic Lymphocytic Leukemia

Antonio Lupia[1,2,+,*], Selena Mimmi[3,+], Enrico Iaccino[3], Domenico Maisano[3], Federica Moraca[2,4], Carmine Talarico[5], Eleonora Vecchio[3], Giuseppe Fiume[3,], Francesco Ortuso[1,2], Giuseppe Scala[3], Stefano Alcaro[1,2,*,§] and Ileana Quinto[3,*,§]

[1] Università "Magna Græcia", Dipartimento di Scienze della Salute, Campus Salvatore Venuta, Viale Europa, Catanzaro, 88100, Italy.

[2] Net4Science srl, Università "Magna Græcia", Campus Salvatore Venuta, Viale Europa, Catanzaro, 88100, Italy.

[3] Università "Magna Græcia", Dipartimento di Medicina Sperimentale e Clinica, Campus Salvatore Venuta, Viale Europa, Catanzaro, 88100, Italy.

[4] University “Federico II” of Naples, Department of Pharmacy, Via D. Montesano 49, Naples, 80131, Italy.

[5] Dompé farmaceutici S.p.A.,Via Pietro Castellino 111, Napoli, 80131, Italy.

* corresponding author

[+] These authors equally contributed to this work

[§] Co-last authors

Abstract

Identification of epitopes recognized by tumour B cells could provide insights into the molecular mechanisms of B cell tumorigenesis through aberrant B cell receptor (BCR) signalling. Here, we analysed the structure of eleven peptides binders of BCRs expressed in Chronic Lymphocytic Leukemia (CLL) patients in order to identify the chemical features required for cross-reactive binding to different CLL clonotypes. Four cross-reactive (CR) and seven no-cross-reactive (NCR) peptides were analysed by means of GRID molecular interaction fields, ligand-based pharmacophore and 3D-QSAR approaches. Based on pharmacophore model, two peptides were generated by specific amino acids substitutions of the parental NCR peptides; these new peptides resumed the common chemical features of CR peptides and bound the CLL BCR clonotypes recognized by CR peptides and parental NCR peptides. Thus, our computational approach guided the pharmacophore modelling of CR peptides. In perspective, peptide binders of CLL BCR clonotypes could represent a powerful tool for computational modelling of epitopes recognized by tumour B cells clones.

Keywords: CLL, B cell receptor, phage display, peptides, GRID-pharmacophore, LB-3D-QSAR.

Introduction

Chronic Lymphocytic Leukemia (CLL) is the most common adult leukemia in Western countries[1-4], characterized by the presence of neoplastic CD5/CD19/CD20/CD23-positive B lymphocytes that accumulate in bone marrow, blood, lymph nodes and spleen[5,6]. As current hypothesis[7], the selection of CLL clones occurs through the antigen-binding to their immunoglobulin B cell receptor (IgBCR), promoting a deregulated proliferation and survival signalling[8-12]. The IgBCR contains two immunoglobulin-heavy (IgH) and two -light (IgL) chains bound by disulphide S-S bridges[13,14], with a unique and randomly determined antigen-binding site corresponding to the variable regions of IgH and IgL chains. The variable regions are codified by the gene segments V_H-D-J_H for IgH chain and V_L-J_L for IgL chain and are generated by DNA rearrangements and point mutations occurring in an antigen-dependent process of B cell differentiation and maturation in the bone marrow and lymph nodes[15,16]. The maturation of a functional IgBCR occurs during B cell development from pro-B to pre-B stage and results in the production of the primary repertoire of low-affinity IgM and IgD antibodies. Thereafter, the antigen stimulation of IgBCR in mature B lymphocytes promotes the production of antibodies with increased affinity for the antigen through somatic recombination and somatic hyper-mutation[17]. These molecular events determine the IgBCR uniqueness of each B cell and the unlimited number of IgBCRs in overall B cell population[18]. Thus, the variable regions of the IgBCR represent a tumour-B cell-specific marker and an optimal target for personalized cancer therapy in B-lymphoproliferative disorders[19,20].

In CLL cells, the IgBCR peculiarly assembles a limited set of heavy and light chain variable genes of immunoglobulin, with 30% CLL patients showing stereotyped IgBCRs[21-29]. These findings support the hypothesis that common environmental antigens or auto-antigens are responsible for CLL pathogenesis through their binding to the IgBCR of neoplastic B cells, thus promoting their positive selection[30-34].

In this regard, we previously demonstrated that the screening of phage-displayed peptide libraries (RPLs) is a powerful tool for investigating protein-protein interaction[35,36] and selecting peptide binders of the idiotypic region of IgBCRs, so-called Id-peptides[37-41]. Indeed, by this experimental approach, we selected Id-peptides of tumour B cell clones, which were used for detecting B cell populations *in vitro* in cell culture and *in vivo* in mouse models of B-lymphoma and multiple myeloma[19,42,43]. Recently, we selected a pool of Id-peptides of neoplastic B cells in six CLL patients, which were classified as no-cross-reactive (NCR) Id-peptides, when they exclusively bound the cognate IgBCR, or cross-reactive (CR) Id-peptides, when they bound the IgBCRs of different CLL patients in addition to their cognate IgBCR[43]. In this context, it is worthwhile to highlight that only a single study was previously reported for characterization of epitopic reactivity of CLL, where a cohort of 100 CLL patients was characterized for phage-binding to their IgBCRs.[44] To date, no study has reported chemical analysis of common chemical features of peptide-binders of CLL IgBCRs.

In the present study, we have analysed the structure of CR Id-peptides to pick up the common chemical features determining their binding to different CLL clonotypes. Due to the lack of crystallographic information on the Id-peptide/IgBCR complex, we adopted a *ligand-based* pharmacophore method using the Fingerprints for Ligands and Proteins (FLAP) software[45-47], which employs the molecular interaction fields (MIFs) calculated by the GRID software[46]. The advantage of the GRID method is due to the use of the MIFs as molecular descriptors of the interaction energy variation between a molecular target and an appropriate probe. A key feature of the GRID software is the transformation of the molecular field describing the interaction of a probe with part of the target into site points, where "target*"* means the ensemble of aligned structures or the binding site of a generic protein. The site points can be used to build all 3- or 4-point pharmacophores encoding a fingerprint[48]. Therefore, every single atom of the ligand corresponds to GRID probes features such as hydrophobicity, hydrogen bond donor and/or acceptor capabilities and charge. In addition, in ligand-based design, the pharmacophore model and MIFs can be used as

input filters to derive a ligand-based 3D-Quantitative structure-activity relationship (LB-3D-QSAR) model[45,47,49]. In this study, the statistical modelling LB-3D-QSAR guided the design of second-generation CR peptides, which incorporated the common chemical features of CR Id-peptides and were able to bind the IgBCR clonotypes of different CLL patients. Our study represents a new frontier in the fields of tumour B-cell monitoring to be further validated on a larger scale of patients.

Results

Identification of common chemical features of CR Id-peptides.

We analysed the chemical features of the following eleven Id-peptides: four CR Id-peptides, showing cross-reactive binding to four different CLL clonotypes including their cognate CLL clonotype; seven NCR Id-peptides, showing exclusive binding to their cognate CLL clonotype (Supplementary Table S1). The pCLL1 peptide was classified as NCR Id-peptide as it bound only three CLL clonotypes including its cognate clonotype (Supplementary Table S1). In order to achieve a collection of information on the degree of flexibility/motion of the peptides, the conformational space of them was explored by using 500 ns of Molecular Dynamics simulations (MDs)[50]. The resulting trajectories of each peptide were used to build the Root Mean Square Deviations (RMSD) matrix. The graphical analysis of matrices highlighted the large conformational variability of the Id-peptides (see Supplementary Fig. S1, panels A-M). This observation excluded that CR Id-peptides shared a common structural preorganization degree. For identifying the common chemical features of the CR Id-peptides, a careful ligand-based approach was conducted by using the FLAPpharm algorithm[47]. According to the FLAPpharm methodology, the generation of pharmacophore model requires at least three molecules[47]. The pharmacophore extracted by FLAPpharm, named pharmacophore "pseudomolecule", consists of common pharmacophoric interaction fields (PIFs) and common atom-centered pharmacophoric pseudofields (pseudoPIFs). PIFs and pseudoPIFs correspond to the MIFs and pseudoMIFs in a ligand[49]. The best pharmacophore models were built upon the CR Id-peptides pCLL2-1, pCLL2-2, pCLL6-1 and

pCLL6-2, using the S-score value for defining the goodness of each pharmacophore[47]. A pharmacophore with the best S-score value must include the following features: (a) hydrogen bond donors (N1-type GRID); (b) hydrogen bond acceptor (O-type GRID); (c) positive charge centers (N+-type in GRID); (d) negative charge centers (O-type in GRID); (e) hydrophobic centers (DRY probe in GRID); (f) shape (H-type in GRID); (g) the distance (i.e. spatial resolution) between two GRID points (0.75 Å). The pharmacophore obtained for the CR Id-peptides was composed of a larger hydrogen donor surface area (HBD), a small part characterized by hydrogen acceptor surface area (HBA), and a hydrophobic core (Fig. 1; see Supplementary Fig. S2).

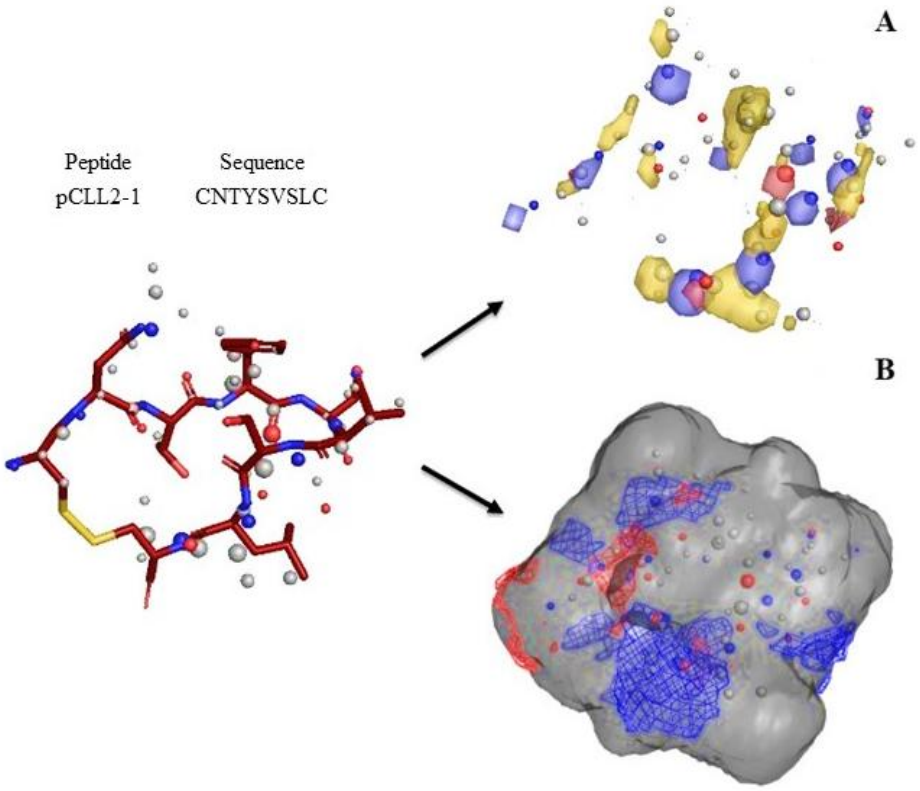

Figure 1. Pharmacophore model generated with the FLAPpharm algorithm. GRID molecular interaction fields (MIFs) are calculated around all CR Id-peptides and condensed into pharmacophoric point in terms of common pharmacophoric points at the centroid of pseudoPIFs (A) and in terms of common pharmacophoric interaction fields (PIFs) (B). The hydrogen bond donor areas (HBD, blue), the hydrogen acceptor areas (HBA, red), the hydrophobic areas (yellow) and the solid surface area (grey) are shown for both pharmacophore depictions (A, B). The cut-off values were -4.0 kcal, -2.5 kcal and -2.0 kcal, for N1, O and DRY probes, respectively. For clarity reasons, we report

the alignment of only one CR Id-peptide (pCLL2-1); the same approach was adopted for pCLL2-2, pCLL6-1 and pCLL6-2 (see Supplementary Fig. S2).

In particular, the main amino acid residues involved in the alignment were: a) Tyr-4, Val-6 and Leu-8 common to pCLL2-1 and pCLL2.2; b) part of the backbone of pCLL6-1 with the alpha carbons (Cα) of Glu-3 and His-6, plus the Lys-8 residue; c) Val-4, Ile-5 and Arg-8 residues of pCLL6-2 (see Supplementary Fig. S2). To validate the discriminatory power of the pharmacophore model, we carried out a virtual screening using the CR Id-peptides as *sensor* ligands, the NCR Id-peptides as *test* ligands, and the best S-score pharmacophore as template. Furthermore, the pharmacophore was used in the PIFs mode to investigate the common GRID fields among the aligned peptides. The ROC curves of each one of the nineteen FLAP descriptors[47] were plotted to evaluate the screening results. The analysis of the area under the curve (AUC) plot showed that the best enrichment value (96%) was displayed by probe N1 (donor) descriptor, while the worst enrichment value (40%) was displayed by probe O (acceptor) descriptor (see Supplementary Fig. S3). The results indicated a good discrimination power between CR and NCR Id-peptides, supporting our pharmacophore model.

Pharmacophore-based design of mutating pAL1 and pAL2 peptides by amino acid substitutions of NCR Id-peptides.

The validated pharmacophore model was submitted to FLAP package to carry out a Ligand-Based 3D-Quantitative Structure-Activity Relationship (LB-3D-QSAR)[45] for the random design of new putative cyclic peptide binders. Quantitative Structure-Activity Relationship (QSAR) techniques depend on the assumption that every compound interacts with the same target in the same way[51]. This is a reasonable assumption when compounds are not too dissimilar. In FLAP package, given a set of aligned molecules, LB-3D-QSAR module allows performing the statistical analysis of the

GRID molecular interaction fields (MIFs) by considering the best pharmacophore as template[45,47]. The molecular descriptors were analysed by using the principal components analysis (PCA), which is a multivariate statistical technique used to analyze interrelationships among a large number of variables, with the aim to reduce a such number of variables into groups known as components or Principal Components (PCs). This method included the variables, thus providing the understanding of similarities and dissimilarities between different ligands. In this way, we extracted the best descriptors around the molecules for the optimization and design of new ligands. Thus, by using a script generate in the *Integrated Development Environment* (IDE) for R (RStudio)[52], we changed the primary sequence of the NCR Id-peptides pCLL3-1 and pCLL4-1 in order to generate an *in-house* library of mutant peptides. We assigned to "*aa*" parameter all amino acid residues (one-letter amino acid codes, e.g. "G = Gly"). In this specific case, we used 18 amino acids, excluding M = Met and C = Cys. Subsequently, we have coded a specific function with the following parameters: "*n*" = number of 2D sequences to be generated, "*l*" = length of the sequence. Then, the 2D sequences were filtered basing on the chemical-physical characteristics relating to the primary sequence and the resulting 53270 sequences were 3D built using a modified version of build_peptide.py script, provided by Schrödinger (Software: Maestro Schrödinger, LLC, New York, NY. 2017). Finally, the resultant library was further subjected to Protein Preparation Wizard tool of Schrödinger's Maestro software (charges and the Optimized Potentials for Liquid Simulations-all atom (OPLS-AA) force field 2005 parameters).[53] In FLAP package, the library was pre-filtered based on the shape-pharmacophore template and the best shot results were analysed in the PCA plot by visual inspection for their comparison with CR and NCR Id-peptides. Within the plot, the following considerations were assumed: (a) as the CR Id-peptides shared similar chemical characteristics and differ from the NCR Id-peptides, then the PCA plot should highlight putative differences; (b) as LB-3D-QSAR performs a statistical analysis based on MIF GRID maps, this method should provide the different chemical characteristics among all peptides. We analysed the evolution of mutant peptides by identifying the peptides with the most similar chemical

characteristics to CR peptides. Among several candidates, two peptides, namely pAL1 and pAL2, showed the most similar profile to CR Id-peptides. As the following step, we investigated the different chemical features of pAL1 and pAL2 respect to the parental peptides pCLL3-1 and pCLL4-1, respectively. Then, we highlighted the common and the different chemical features of pAL1, pAL2, NCR and CR Id-peptides based on the PIFs and pseudoPIFs. For clarity, we separated the alignments of all peptides according to the principal N1, O, C3 and C1= probes (Fig. 2; Fig. 3). In both PC1 and PC2 trends, the peptides pAL1 and pAL2 did not share the characteristics of native pCLL3-1 and pCLL4-1 Id-peptides (Fig. 2), and they were better grouped compared to them (Fig. 4). In particular, the presence of Lys-8 amino acid residue of pAL1 respect to Glu-8 amino acid residue of pCLL3-1 restored part of pharmacophoric characteristics shared by the CR Id-peptides (Fig. 3a-d). In fact, by comparing the surface areas drawn by O probe, a different mapping in this region was observed for pAL-1 compared to pCLL3-1 (Fig. 3e, f). Similarly, the Trp-5 and Trp-6 amino acid residues of pAL2, replacing the Pro-5 and Ala-6 amino acid residues of pCLL4-1, adapted to the chemical surfaces area mapped by C1= and C3 probes. Indeed, the substitution of these two amino acids led to a conformational change of the pAL2 peptide, which was highlighted by the different alignment of the N-terminal amino acid, Thr-3 and Asn-4 on the chemical surface mapped by O probe respect to pCLL4-1 (Fig. 3g, h). Interesting remarks came from the alignments of the NCR Id-peptides. In fact, the alignments of NCR Id-peptides did not superimpose the principal MIF GRID maps of CR Id-peptides (see Supplementary Fig. S4). This evidence suggested that pAL1 and pAL2 had the chemical features of CR Id-peptides, instead of NCR Id-peptides. Consistently, the RMSD matrix of both peptides, as analysed by 500 ns of MDs, showed a large conformational variability of pAL1 and pAL2 peptides, similarly to CR peptides (see Supplementary Fig. S1, panels N-O).

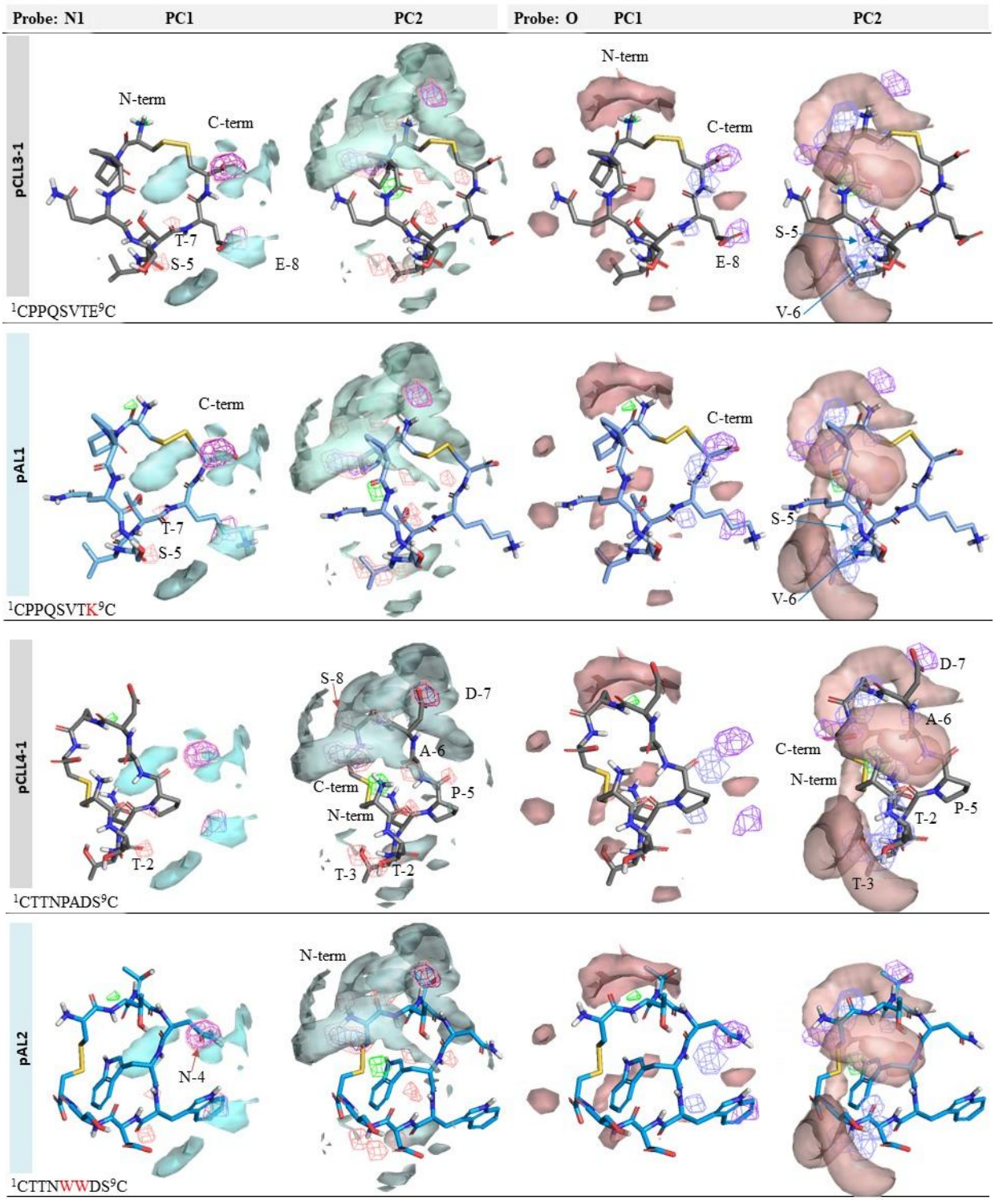

Figure 2. Different chemical features of pAL1 and pAL2 respect to the relative parental pCLL3-1 and pCLL4-1 peptides as identified by GRID MIFs maps alignments. The figure shows the different alignment of pAL1 and pAL2 generated by substitutions into the backbone of pCLL3-1 and pCLL4-1, respectively. The pharmacophoric features PIFs (solid view; light blue = hydrogen bond donor, red = hydrogen bond acceptor) and pseudoPIFs (wireframe view; cyan = hydrogen bond donor, red = hydrogen bond acceptor, violet = negative charged, green = positive charged) of pCLL3-1 and pCLL4-1 are fixed seals. The amino acid substitutions introduced in pAL1 and pAL2 respect to the parental

pCLL3-1 and pCLL4-1, respectively, are highlighted in red. The principal components (PCs) are indicated as the variables PC1 and PC2.

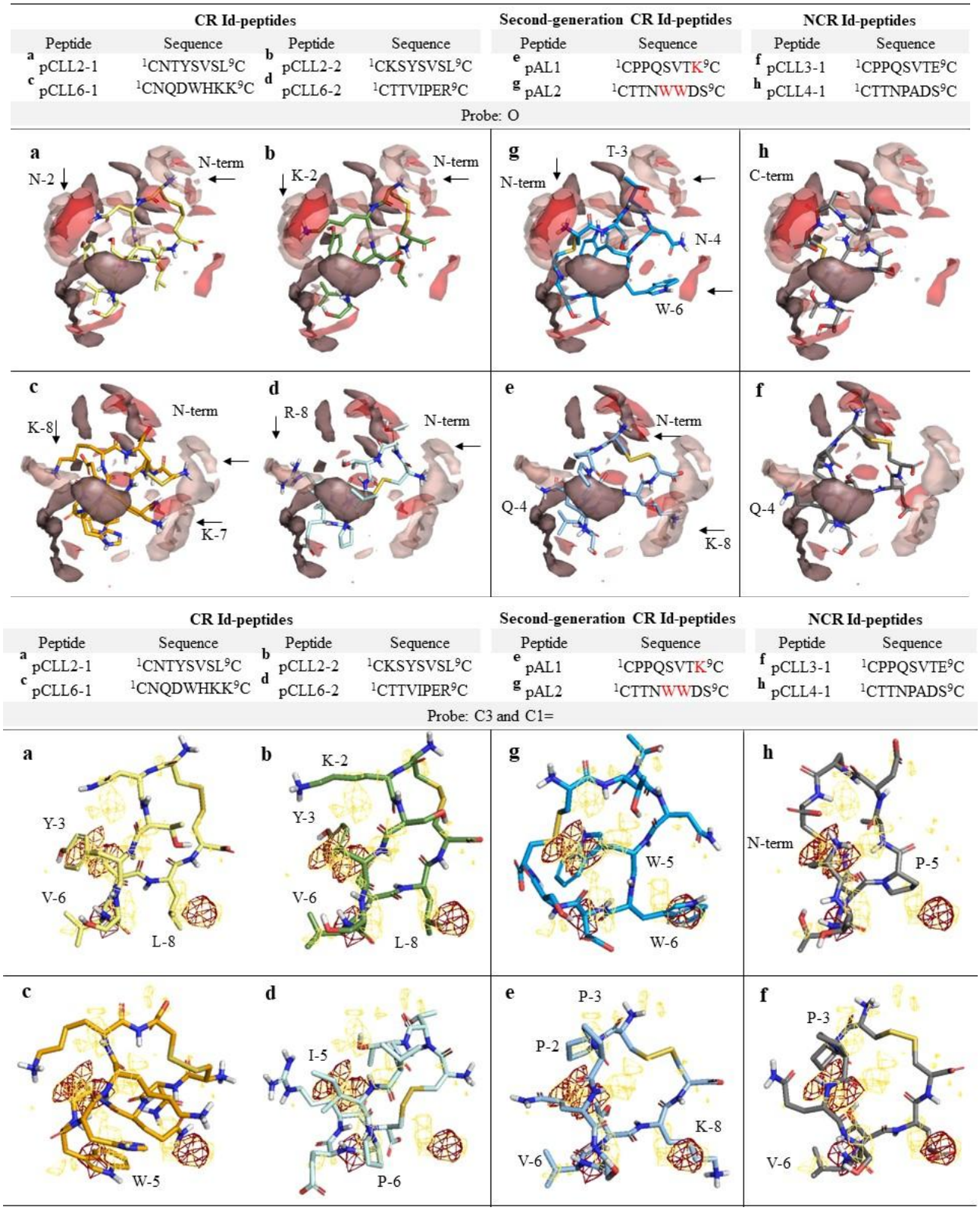

Figure 3. Common chemical features of pAL1 and pAL2 respect to CR Id-peptides. (Top panel) The alignments of the PIFs pharmacophoric features mapped by O probe (solid view) are shown for NCR Id-peptides (dark-red), pAL1 and pAL2 peptides (red) and CR Id-peptides (light-red). (Bottom panel) the alignments of the pseudoPIFs pharmacophoric features mapped by C3 (yellow) and C1= (brown) probes (wireframe view) are shown for NCR Id-peptides, pAL1 and pAL2 and CR Id-peptides. Only the common surface areas of CR Id-peptides (**a-b-c-d**), pAL1 (**e**) and pAL2 (**g**) are labelled and highlighted. As compared at the same cut-off of -2.50 kcal, the chemical surface area of NCR Id-peptides (**f-h**) does not fit the same chemical surface area.

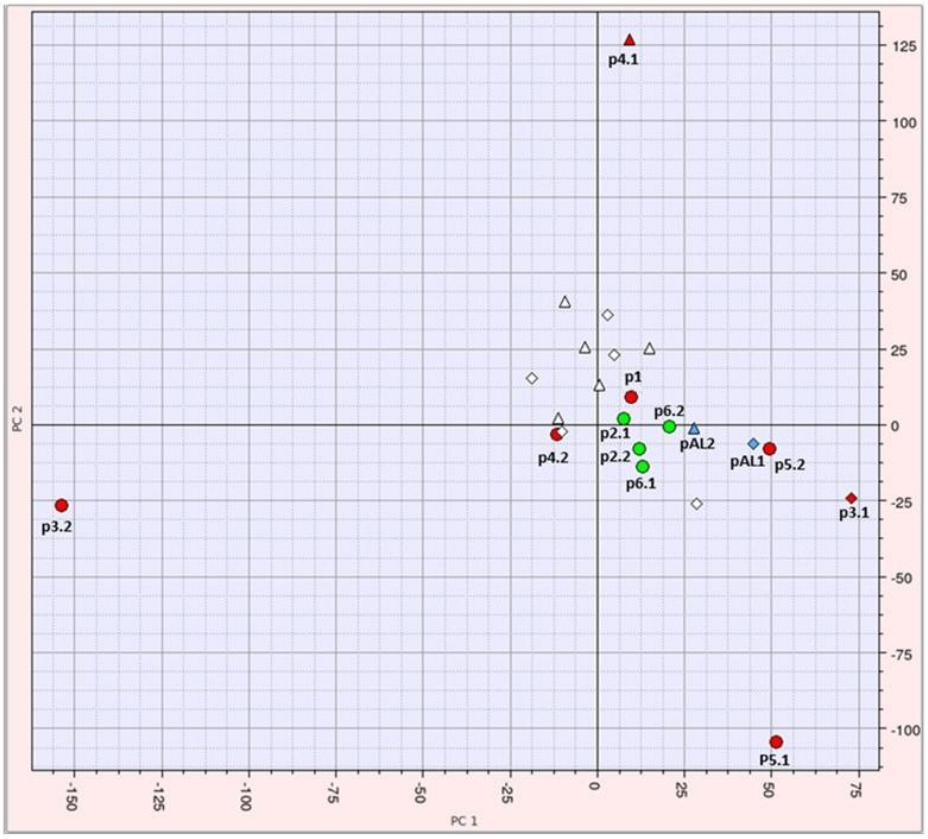

Figure 4. Principal component analysis of IgBCR peptide binders dataset. The plot of principal component analysis (PCA) shows the distribution of NCR Id-peptides and CR Id-peptides that are labelled in red and green, respectively. The second-generation peptides pAL1 and pAL2 (blue), which derived respectively from the NCR Id-peptides pCLL3-1 and pCLL4-1, are represented in the diamond and triangle shape.

pAL1 and pAL2 peptides bind the CLL IgBCR clonotypes that are commonly recognized by CR Id-peptides.

The peptides pAL1 and pAL2 were *in vitro* tested by enzyme-linked immunosorbent assay (ELISA) for the binding to eleven CLL clonotypes recognized by CR or NCR Id-peptides[43] (Supplementary Table S1). To this end, N-biotinylated peptides (2 μg/ml) were incubated with purified CLL IgBCRs (1,5 μg/ml) and their binding was revealed by alkaline phosphatase-conjugated anti-human IgG. Similarly, to their parental NCR Id-peptides, pAL1 bound the CLL3-1 and CLL3-2 clonotypes (Fig. 5A), and pAL2 bound the CLL4-1 and CLL4-2 clonotypes (Fig. 5B). Further, they acquired the ability to bind the CLL clonotypes recognized by the CR Id-peptides (CLL1, CLL2-1, CLL2-2, CLL4-2, CLL6-1 and CLL6-2 clonotypes) (Fig. 5 A, B). No binding of pAL1 and pAL2 was detected to the control Human IgG1 and the CLL clonotypes recognized by NCR Id-peptides (CLL4-1, CLL5-1 and CLL5-2 for pAL1; CLL3-1, CLL3-2, CLL5-1 and CLL5-2 for pAL2) (Fig. 5 A, B). To measure the affinity binding, N-biotinylated pAL1 and pAL2 peptides were incubated at increasing concentrations (0.01 nM up to 300 nM) in microtiter plates coated with the purified CLL IgBCRs (2 μg/ml) (see Supplementary Fig. S5). The K_D values are reported in Table S2. Thus, critical amino acid substitutions of pCLL3-1 (Glu-8 to Lys-8) and pCLL4-1 (Pro-5 to Trp-5; Ala-6 to Trp-6) switched their behaviour from NCR to CR peptides.

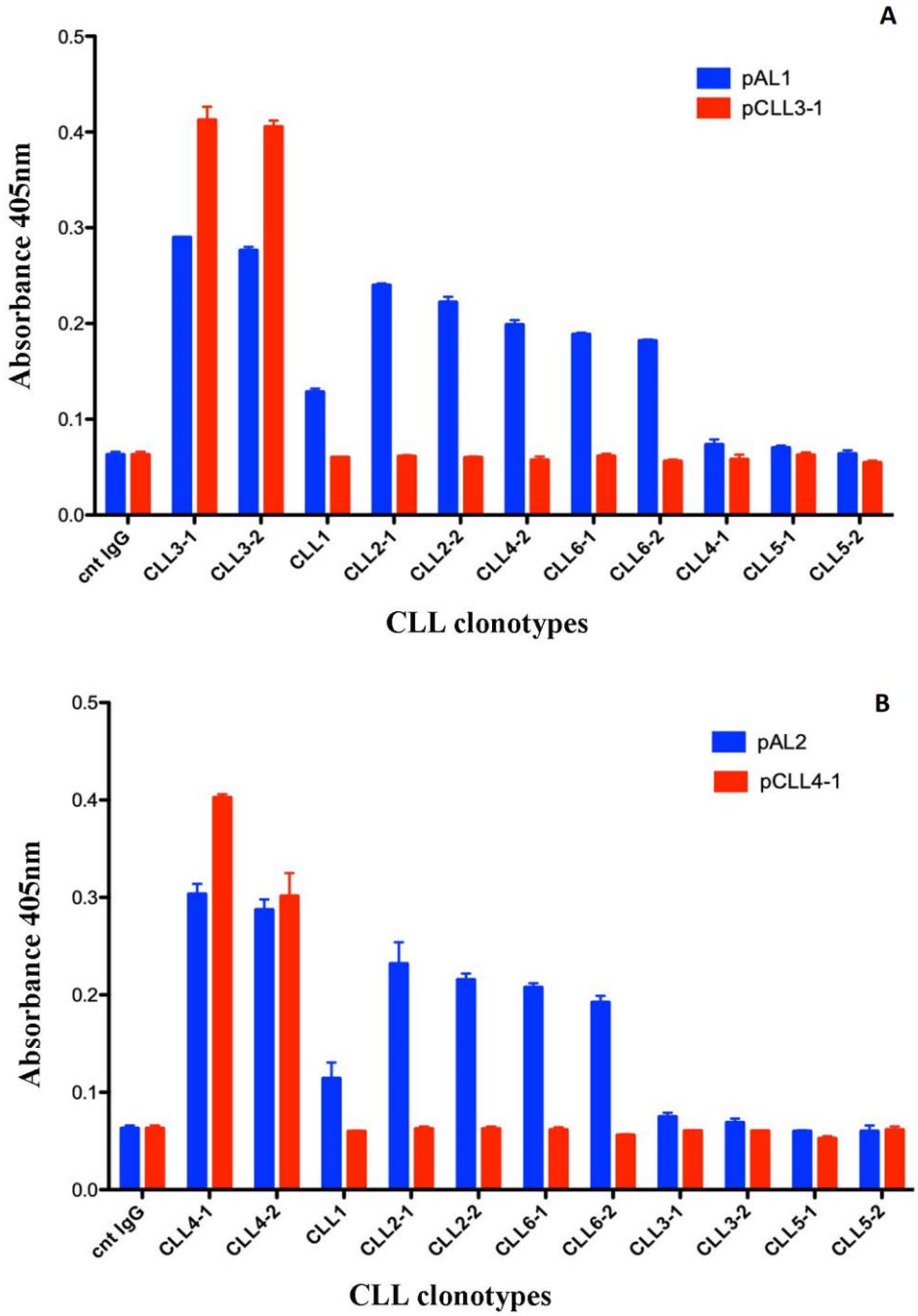

Figure 5. Binding affinity of pAL1 and pAL2 to IgBCR CLL clonotypes.

Microtiter plates coated with pAL1 (A) and pAL2 (B) were incubated with purified CLL IgBCRs and their binding was analysed as absorbance values by ELISA[41]. pCLL3-1 (parental peptide of pAL1) and pCLL4-1 (parental peptide of pAL2) were included as controls. Absorbance values (mean ± SEM) of 3 independent experiments were measured at 405nm and normalized to the blank sample.

Discussion

The IgBCR is B cell specific and can be considered a tumour marker in case of B-lymphoproliferative disorders. The hypothesis of antigen-driven B-tumorigenesis considers the epitope recognized by the variable regions of the IgBCR as a relevant stimulus promoting self-activation and proliferation of a tumour B cell clone. A protein surface exposes several epitopic regions; thus, the conformational epitopes that may be relevant for IgBCR signalling are difficult to map in the absence of high-resolution structural information of the antigen-antibody complexes. In this regard, the structural analysis of peptide binders of the IgBCR could give a clue on the epitopic conformation recognized by the tumour B cell.

The present study builds on the previous selection of eleven cyclic peptide binders of the idiotypic region of IgBCRs, so-called Id-peptides, expressed by six CLL patients[43]. These Id-peptides were distinguished in four CR Id-peptides (pCLL2-1, pCLL2-2, pCLL6-1, pCLL6-2) and seven NCR Id-peptides (pCLL1, pCLL3-1, pCLL3-2, pCLL4-1, pCLL4-2, pCLL5-1, pCLL5-2), depending on their ability to bind the same six CLL IgBCR clonotypes or not, respectively. Here, we aimed to identify the chemical features required for cross-reactive binding by computational analysis of CR and NCR Id-peptides structure.

Based on the concept that a pharmacophore is a well accurate three-dimensional abstraction of intermolecular interactions, including an ensemble of steric and electronic features[54], we used a computational approach for identifying the common chemical features of CR Id-peptides, not present in NCR Id-peptides, that mediated the binding to different CLL IgBCRs. As the first step, we generated a ligand-based pharmacophore model that was used for peptide-optimization workflows. By using the LB-3D-QSAR methodology, we identified the amino acids with similar chemical properties, which were present in CR Id-peptides, while absent in NCR Id-peptides, and thus they could be likely responsible for cross-reactive binding. Then, the pharmacophore model guided the specific amino acid substitutions of the NCR Id-peptides pCLL3-1 and pCLL4-1. Substitution of Glu-8 of pCLL3-1 into Lys-8 generated pAL-1. Substitution of Pro-5 and Ala-6 of

pCLL4-1 into Trp-5 and Trp-6, respectively, generated pAL-2. The amino acid substitutions introduced in the backbone of the NCR Id-peptides facilitated a better alignment to the chemical surface areas of the CR Id-peptides. In particular, the amino acid substitutions introduced in pAL-1 and pAL-2 fitted in the hydrophobic and aromatic regions of the pseudoPIFs pharmacophore features mapped by C3 and C1= probes of the CR Id-peptides, and shared the hydrogen bond acceptor areas of the PIFs pharmacophore features mapped by O probe of the CR peptides (Fig. 3). By comparing the alignments of all peptides based on the PIFs and pseudoPIFs pharmacophore features at the same energetic cut-off value, the chemical areas of the CR Id-peptides and pAL1 and pAL2 peptides were not superimposable to the NCR Id-peptides (Fig. 2; see Supplementary Fig. S4). Thus, the pharmacophore-guided amino acid substitution of NCR Id-peptides converted them to the configuration of CR Id-peptides.

As functional test, we analysed the binding of pAL1 and pAL2 to different CLL IgBCR clonotypes. pAL1 and pAL2 kept the binding activity to the IgBCR recognized by their parental NCR peptides, and also acquired the binding to the IgBCR recognized by the CR Id-peptides. These observations agree with the chemical features alignment of pAL1 and pAL2 to CR Id-peptides, as shown by MIFs GRID maps. Thus, the pharmacophore model derived from CR Id-peptides correctly guided the design of new cross-reactive peptides starting from the backbone of NCR Id-peptides.

Altogether, these results indicate that by computational means we designed a pharmacophore resuming the common chemical features of CR Id-peptides. This was possible by taking into account a pool of Id-peptides that had been selected by screening phage display peptide libraries using as bait the IgBCR of neoplastic B cells derived from CLL patients[43]. The pharmacophore model correctly identified the crucial amino acids within the sequence of the CR Id-peptides that were responsible for cross-reactive binding. This was proved by the pharmacophore-guided design of two peptides that showed the same binding ability of CR Id-peptides.

Recent studies have shown the occurrence of highly restricted and quasi-identical variable VH-CDR3 sequences of the CLL IgBCRs, so-called CLL BCR stereotyped subsets[29]. Here, we have improved this analysis by identifying the chemical features of peptides recognized as epitopes by eleven CLL clonotypes[43]. It is worthwhile to remind that the CLL IgBCRs can have different primary structures even though they can bind the same epitope. Thus, the CLL IgBCR classification based on the epitope recognition - and not exclusively on the primary structure of the IgBCR- should represent an advantage in terms of functional analysis of tumour IgBCRs. The IgBCRs of tumour B cells have been shown to recognize cellular and microbial antigens that were suspected to act as antigenic stimuli sustaining the pro-tumorigenic signalling in B cells[10,55-57]. Our pharmacophoric model supports this hypothesis as it successfully identified the chemical features of peptides commonly recognized by a set of CLL clonotypes, which likely mimic the epitopes harboured by putative antigenic stimuli. Thus, the computational method here described could be useful to identify the structure of putative super-antigens driving the selection of neoplastic B cells. Further, it could provide pharmacophore models as new tools for CLL patient stratification based on epitope recognition. In perspective, the pharmacophore modelling of CLL peptide-binders is a powerful methodology for designing peptide-based arrays for monitoring tumour B-cells and it could allow the development of new therapeutic peptide-based strategies for tumour specific targeting.

Materials and Methods

Conformational analysis and pharmacophore search. The 3D structures of pCLL1, pCLL2-1, pCLL2-2, pCLL3-1, pCLL3-2, pCLL4-1, pCLL4-2, pCLL5-1, pCLL5-2, pCLL6-1 and pCLL6-2 were built by using the Maestro GUI (Software: Maestro Schrödinger, LLC, New York, NY. 2017) and submitted to Schrödinger's Protein Preparation Wizard (charges and OPLS2005[53] force field parameters). The conformational search of each peptide was explored through *500 ns* of Molecular Dynamics simulations (MDs) using the Desmond package[50], with the OPLS2005 force

field.[53] Transferable intermolecular potential with 3 points (TIP3P)[58] explicit solvation model was adopted for considering the water solvent effects on peptides conformational properties. MDs were conducted in the isothermal-isobaric ensemble at 300 K under 1 atm pressure. At the end of MDs, the trajectory coordinates of the eleven Id-peptides were submitted to a clustering analysis and 10 representative peptide conformations (10 clusters) were obtained for each peptide, corresponding to 10 different conformers. The RMSD cut-off [59] value was set at 0.5 Å to avoid retrieving redundant conformations. According to the FLAPpharm algorithm[47] on the basis of fields and pseudo-fields superposition method, the CR Id-peptides conformers were aligned to each other in order to find the optimal MIF similarity, thus deriving the best pharmacophore model. The alignment was carried out using quadruplets formed by the association of the atoms in peptides and the molecular alignments were then scored using the MIF similarity. No constraints were applied in pharmacophore generation.

LB-3DQSAR model. Each peptide was superimposed to the validated pharmacophore model and was conformationally directed to assume the shape obligatory for its features. Then, the PCA plot was used to analyse the PCs, which described most of the information contained in the independent variables.

Enzyme-linked immunosorbent assay. The binding of pAL1 and pAL2 to CLL IgBCR clonotypes was tested by enzyme-linked immunosorbent assay (ELISA) as previously reported[43]. N-biotinylated-peptides were purchased from Caslo Laboratory ApS, Denmark. Briefly, streptavidin-coated multi-well plates (Thermo Scientific™ Micropiastre Nunc™ MicroWell™, MA, USA) were coated with N-biotinylated peptides (2 μg/ml) overnight at 4°C in coating buffer (50 mM NaHCO3, pH 9.6). The unbound peptides were washed out with washing buffer (0.05% $Tween_{20}$, Phosphate Buffered Saline pH7.4 1X), and the uncoated sites were blocked with a 5% Bovine Serum Albumin (BSA) solution. Purified CLL IgBCRs were produced as previously described[43], and they were 16 hr-incubated at the concentration of 1.5 μg/ml in the peptides-coated

microwells at 4°C in blocking buffer (0.05% $Tween_{20}$, 5% BSA, Phosphate Buffered Saline pH7.4 1X). A monoclonal human IgG1 (Sigma-Aldrich I 5154) was included as control. The unbound IgBCRs were washed out with washing buffer (0.1% $Tween_{20}$, Phosphate Buffered Saline pH7.4 1X). The peptide binding to IgBCRs was revealed by 1 hr-incubation with an anti-human IgG (Fc-specific) alkaline phosphatase-conjugated (Sigma-Aldrich A9544) at room temperature and measured by ELISA reader (Labsystems multiscan MS) at 405nm. For K_D calculation, maxi-sorp multi-well plates (Thermo Scientific™ Micropiastre Nunc™ MicroWell™, MA, USA) were coated with purified CLL IgBCRs (2 µg/ml) overnight at 4°C in coating buffer (50 mM NaHCO3, pH 9.6). The unbound IgBCRs were washed out with washing buffer (0.05% $Tween_{20}$, Phosphate Buffered Saline pH7.4 1X), and the uncoated sites were blocked with 5% BSA solution. N-biotinylated peptides were 16 hr-incubated at different concentrations (1 µg/ml up to 1 mg/ml) in blocking buffer (0.05% $Tween_{20}$, 5% BSA, Phosphate Buffered Saline pH7.4 1X) at 4°C. A scrambled peptide (sequence CGGNGPGLC) was included as control. The unbound peptides were washed out with washing buffer (0.1% $Tween_{20}$, Phosphate Buffered Saline pH7.4 1X). The peptide binding to IgBCRs was revealed by 1 hr-incubation with streptavidine-alkaline phosphatase-conjugated (Sigma-Aldrich S2890) at room temperature and measured by ELISA reader (Labsystems multiscan MS) at 405nm. K_D values were calculated by Scatchard plot of 3 independent experiments.

Acknowledgements

This work was supported by the following grants: Associazione Italiana per la Ricerca sul Cancro IG-2012-13388 to G. S.; Ministero della Salute RF-2010-2306943 to G. S.; PRIN project 2012CK5RPF to G. S.; PRIN project 2012CK5RPF_002 to I. Q. ; PRIN project 201744BN5T to S.A.; POR FES/FESR 2014-20-ATS ALCMEONE cup J18C17000610006 to I.Q.; S.M. was supported by Fondazione Italiana per la Ricerca sul Cancro (FIRC) fellowship. D.M. was supported by funds from the EU project 'PON-RI2014-2020. The authors thank Prof. Camillo Palmieri,

Department of Experimental and Clinical Medicine, University Magna Graecia of Catanzaro, for helpful discussions.

Abbreviations:

AUC, area under the curve; BSA, Bovine Serum Albumin; CLL, Chronic Lymphocytic Leukemia; CR, cross-reactive Id-peptides; ELISA, enzyme-linked immunosorbent assay; FLAP, Fingerprints for Ligands and Proteins; HBD, hydrogen bond donor; HBA, hydrogen bond acceptor; IDE, Integrated Development Environment; Id-peptides, peptide binders of the idiotypic region of IgBCRs; IgBCR, Immunoglobulin B cell receptor; IgH, Immunoglobulin-heavy; IgL, Immunoglobulin-light; LB-3D-QSAR, Ligand-based 3D-Quantitative structure-activity relationship; MDs, Molecular Dynamics simulations; MIFs, Molecular Interaction Fields; NCR, no-cross-reactive Id-peptides; OPLS, optimized potentials for liquid simulations; PCA, principal components analysis; PCs, Principal Components; PIFs, pharmacophoric interaction fields; pseudoPIFs, atom-centered pharmacophoric pseudofields; QSAR, Quantitative Structure-Activity Relationship; RMSD, Root Mean Square Deviations; RPLs, phage-displayed peptide libraries; TIP3P, Transferable intermolecular potential with 3 points.

Author contributions

A.L. performed and interpreted the computational data; S.M and E.I. performed and interpreted the biological data; A.L. and S. M. wrote the manuscript; D.M., F.M., C.T., E.V., G.F. and F.O. contributed to data analysis and discussion; S.A., G.S. and I.Q. conceived the research plan, provided funding and resources, and revised the manuscript.

References

1 Pekarsky, Y., Zanesi, N. & Croce, C. M. Molecular basis of CLL. *Semin Cancer Biol* **20**, 370-376, doi:10.1016/j.semcancer.2010.09.003 (2010).

2 Esencay, M. & Hamad, B. The chronic lymphocytic leukaemia market. *Nat Rev Drug Discov* **14**, 381-382, doi:10.1038/nrd4625 (2015).

3 Howlader, N. e. a., SEER Cancer Statistics Review, 1975–2010. National Cancer Institute Surveillance, Epidemiology and End Results Program [online] http://seer.cancer.gov/csr/1975_2010/. 2013

4 Albano, F. *et al.* The expression of inhibitor of bruton's tyrosine kinase gene is progressively up regulated in the clinical course of chronic lymphocytic leukaemia conferring resistance to apoptosis. *Cell Death Dis* **9**, 13, doi:10.1038/s41419-017-0026-3 (2018).

5 Campo, E. *et al.* The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* **117**, 5019-5032, doi:10.1182/blood-2011-01-293050 (2011).

6 Burger, J. A. & Wiestner, A. Targeting B cell receptor signalling in cancer: preclinical and clinical advances. *Nat Rev Cancer* **18**, 148-167, doi:10.1038/nrc.2017.121 (2018).

7 Chiorazzi, N. & Ferrarini, M. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annu Rev Immunol* **21**, 841-894, doi:10.1146/annurev.immunol.21.120601.141018 (2003).

8 Matutes, E. *et al.* The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia* **8**, 1640-1645 (1994).

9 Maharaj, K., Sahakian, E. & Pinilla-Ibarz, J. Emerging role of BCR signaling inhibitors in immunomodulation of chronic lymphocytic leukemia. *Blood Adv* **1**, 1867-1875, doi:10.1182/bloodadvances.2017006809 (2017).

10 Jiménez de Oya, N. *et al.* Pathogen-specific B-cell receptors drive chronic lymphocytic leukemia by light-chain-dependent cross-reaction with autoantigens. *EMBO Mol Med* **9**, 1482-1490, doi:10.15252/emmm.201707732 (2017).

11 Vangapandu, H. V. *et al.* B-cell Receptor Signaling Regulates Metabolism in Chronic Lymphocytic Leukemia. *Mol Cancer Res* **15**, 1692-1703, doi:10.1158/1541-7786.MCR-17-0026 (2017).

12 Gaidano, G., Foà, R. & Dalla-Favera, R. Molecular pathogenesis of chronic lymphocytic leukemia. *J Clin Invest* **122**, 3432-3438, doi:10.1172/JCI64101 (2012).

13 Schroeder, H. W. & Cavacini, L. Structure and function of immunoglobulins. *J Allergy Clin Immunol* **125**, S41-52, doi:10.1016/j.jaci.2009.09.046 (2010).

14 Matsuuchi, L. & Gold, M. R. New views of BCR structure and organization. *Curr Opin Immunol* **13**, 270-277 (2001), doi: 10.1016/S0952-7915(00)00215-6

15 Avalos, A. M., Meyer-Wentrup, F. & Ploegh, H. L. B-cell receptor signaling in lymphoid malignancies and autoimmunity. *Adv Immunol* **123**, 1-49, doi:10.1016/B978-0-12-800266-7.00004-2 (2014).

16 Avalos, A. M. & Ploegh, H. L. Early BCR Events and Antigen Capture, Processing, and Loading on MHC Class II on B Cells. *Front Immunol* **5**, 92, doi:10.3389/fimmu.2014.00092 (2014).

17 Chan, T. D. & Brink, R. Affinity-based selection and the germinal center response. *Immunol Rev* **247**, 11-23, doi:10.1111/j.1600-065X.2012.01118.x (2012).

18 Wu, X. *et al.* Immunoglobulin somatic hypermutation: double-strand DNA breaks, AID and error-prone DNA repair. *J Clin Immunol* **23**, 235-246 (2003), doi:10.1023/A:102457171.

19 Iaccino, E. *et al.* Monitoring multiple myeloma by idiotype-specific peptide binders of tumor-derived exosomes. *Mol Cancer* **16**, 159, doi:10.1186/s12943-017-0730-8 (2017).

20 Stamatopoulos, K. *et al.* Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood* **109**, 259-270, doi:10.1182/blood-2006-03-012948 (2007).

21 Stamatopoulos, K. *et al.* Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood* **106**, 3575-3583, doi:10.1182/blood-2005-04-1511 (2005).

22 Ghiotto, F. *et al.* Similarities and differences between the light and heavy chain Ig variable region gene repertoires in chronic lymphocytic leukemia. *Mol Med* **12**, 300-308, doi:10.2119/2006–00080.Ghiotto (2006).

23 Stevenson, F. K. & Caligaris-Cappio, F. Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood* **103**, 4389-4395, doi:10.1182/blood-2003-12-4312 (2004).

24 Tobin, G. The immunoglobulin genes: structure and specificity in chronic lymphocytic leukemia. *Leuk Lymphoma* **48**, 1081-1086, doi:10.1080/10428190701342034 (2007).

25 Messmer, B. T. *et al.* Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J Exp Med* **200**, 519-525, doi:10.1084/jem.20040544 (2004).

26 Fais, F. *et al.* Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest* **102**, 1515-1525, doi:10.1172/JCI3009 (1998).

27 Murray, F. *et al.* Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood* **111**, 1524-1533, doi:10.1182/blood-2007-07-099564 (2008).

28 Bomben, R. *et al.* Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. *Br J Haematol* **144**, 492-506, doi:10.1111/j.1365-2141.2008.07469.x (2009).

29 Ten Hacken, E., Gounari, M., Ghia, P. & Burger, J. A. The importance of B cell receptor isotypes and stereotypes in chronic lymphocytic leukemia. *Leukemia,* doi:10.1038/s41375-018-0303-x (2018).

30 Chiorazzi, N., Rai, K. R. & Ferrarini, M. Chronic lymphocytic leukemia. *N Engl J Med* **352**, 804-815, doi:10.1056/NEJMra041720 (2005).

31 Goodnow, C. C. *et al.* Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334**, 676-682, doi:10.1038/334676a0 (1988).

32 Seifert, M. *et al.* Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med* **209**, 2183-2198, doi:10.1084/jem.20120833 (2012).

33 Dühren-von Minden, M. *et al.* Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. *Nature* **489**, 309-312, doi:10.1038/nature11309 (2012).

34 Catera, R. *et al.* Chronic lymphocytic leukemia cells recognize conserved epitopes associated with apoptosis and oxidation. *Mol Med* **14**, 665-674, doi:10.2119/2008-00102.Catera (2008).

35 Mangini, M. *et al.* Peptide-guided targeting of GPR55 for anti-cancer therapy. *Oncotarget* **8**, 5179-5195, doi:10.18632/oncotarget.14121 (2017).

36 Schiavone, M. *et al.* Design and characterization of a peptide mimotope of the HIV-1 gp120 bridging sheet. *Int J Mol Sci* **13**, 5674-5699, doi:10.3390/ijms13055674 (2012).

37 Buhl, L., Szecsi, P. B., Gisselø, G. G. & Schafer-Nielsen, C. Surface immunoglobulin on B lymphocytes as a potential target for specific peptide ligands in chronic lymphocytic leukaemia. *Br J Haematol* **116**, 549-554 (2002), doi:10.1046/j.0007-1048.2001.03311.x.

38 Liu, Y. *et al.* Peptides that bind specifically to an antibody from a chronic lymphocytic leukemia clone expressing unmutated immunoglobulin variable region genes. *Mol Med* **19**, 245-252, doi:10.2119/molmed.2013.00082 (2013).

39 Seiler, T. *et al.* Characterization of structurally defined epitopes recognized by monoclonal antibodies produced by chronic lymphocytic leukemia B cells. *Blood* **114**, 3615-3624, doi:10.1182/blood-2009-01-197822 (2009).

40 Nilsson, F., Tarli, L., Viti, F. & Neri, D. The use of phage display for the development of tumour targeting agents. *Adv Drug Deliv Rev* **43**, 165-196 (2000), doi:10.1016/S0169-409X(00)00068-5.

41 Mimmi, S., Maisano, D., Quinto, I. & Iaccino, E. Phage Display: An Overview in Context to Drug Discovery. *Trends Pharmacol Sci*, doi:10.1016/j.tips.2018.12.005 (2018).

42 Palmieri, C. *et al.* In vivo targeting and growth inhibition of the A20 murine B-cell lymphoma by an idiotype-specific peptide binder. *Blood* **116**, 226-238, doi:10.1182/blood-2009-11-253617 (2010).

43 Mimmi, S. *et al.* Evidence of shared epitopic reactivity among independent B-cell clones in chronic lymphocytic leukemia patients. *Leukemia* **30**, 2419-2422, doi:10.1038/leu.2016.245 (2016).

44 Binder, M. *et al.* B-cell receptor epitope recognition correlates with the clinical course of chronic lymphocytic leukemia. *Cancer* **117**, 1891-1900, doi:10.1002/cncr.25755 (2011).

45 Baroni, M., Cruciani, G., Sciabola, S., Perruccio, F. & Mason, J. S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J Chem Inf Model* **47**, 279-294, doi:10.1021/ci600253e (2007).

46 Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **28**, 849-857 (1985), doi:10.1021/jm00145a002.

47 Cross, S., Baroni, M., Goracci, L. & Cruciani, G. GRID-based three-dimensional pharmacophores I: FLAPpharm, a novel approach for pharmacophore elucidation. *J Chem Inf Model* **52**, 2587-2598, doi:10.1021/ci300153d (2012).

48 Perruccio, F., Mason, J. S., Sciabola, S. & Baroni, M. (R. Mannhold, H. Kubinyi, G. Folkers and G. Cruciani, In Molecular Interaction Fields, 2006).

49 Cross, S. *et al.* GRID-based three-dimensional pharmacophores II: PharmBench, a benchmark data set for evaluating pharmacophore elucidation methods. *J Chem Inf Model* **52**, 2599-2608, doi:10.1021/ci300154n (2012).

50 Bowers, K. J. *et al.* Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters.

51 Acharya, C., Coop, A., Polli, J. E. & Mackerell, A. D. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr Comput Aided Drug Des* **7**, 10-22 (2011).

52 Team, R. RStudio: integrated development environment for R. RStudio. *Inc., Boston, MA* **14** (2015).

53 Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **118**, 11225-11236 (1996), doi:10.1021/ja9621760.

54 Wermuth, C. G. Pharmacophores and Pharmacophore Searches (eds T. Langer and R.D. Hoffmann). 3-13 (2006).

55 Chu, C. C. *et al.* Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. *Blood* **115**, 3907-3915, doi:10.1182/blood-2009-09-244251 (2010).

56 Lanemo Myhrinder, A. *et al.* A new perspective: molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies. *Blood* **111**, 3838-3848, doi:10.1182/blood-2007-11-125450 (2008).

57 Hoogeboom, R. *et al.* A mutated B cell chronic lymphocytic leukemia subset that recognizes and responds to fungi. *J Exp Med* **210**, 59-70, doi:10.1084/jem.20121801 (2013).

58 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935 (1983), doi:10.1063/1.445869.

59 Daura, X. *et al.* Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition* **38**, 236-240 (1999), doi: 10.1002/(sici)1521-3773(19990115)38:1/2<236::aid-anie236>3.0.co;2-m.