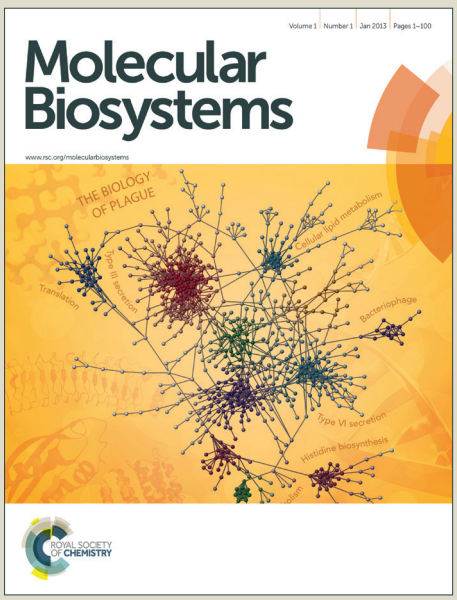
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

Molecular BioSystems

Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Design of gliadin peptide analogues with low affinity for the celiac disease associated HLA-DQ2 protein

Domenico Fraccalvieri,[a] Stefano Motta,[a] Guido Galliani,[b] Linda Cavaletti,[c] Laura Bonati[a]*

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Abstract

The HLA-DQ2.5 receptors bind gluten-derived peptides and present them to the T cells in the intestinal mucosa thus inducing the development of immune responses typical in the celiac disease. On the basis of the X-ray structure of the domain of HLA-DQ2.5 bound to the DQ2.5-glia-α1a epitope, fifteen peptides were designed with the aim of lowering the epitope binding affinity, thus reducing the autoimmune response. Hydroxylation of Pro residues was proposed as a suitable functionalization given that both enzymatic and chemical synthetic methods are available. Then, a computational study on the effects of Pro hydroxylation on HLA-DQ2.5 binding was performed by molecular docking. A docking protocol able to reproduce the binding geometry of the known crystallographic complex was set up and applied to the designed DQ2.5-glia-α1a analogues. Among them, the one including four di-hydroxylated Pro residues was predicted to lower the binding affinity to a greatest extent. Therefore, the same functionalization was computationally tested also for other celiac disease relevant epitopes, DQ2.5-glia-α1b and DQ2.5-glia-α2, and their ability in inhibiting the binding to HLA-DQ2.5 was confirmed. On this basis, these hydroxylated peptides are expected to significantly affect the gluten activity involved in celiac disease and, after experimental validation, a synthetic method will be developed for introducing this gluten modification directly in flour. The proposed approach is a promising tool to study the binding of other gliadin and glutenin derived T-cell epitopes as well as their variants.

Keywords: gliadin, gluten, HLA-DQ2, molecular docking, celiac disease

Introduction

Formation of complexes between human leukocyte antigen (HLA) DQ proteins and gluten-derived peptides is a key step in the development of (auto)immune responses typical in the celiac disease.

Wheat gluten consists of two protein subcomponents, gliadins and glutenins, both containing repetitive sequences rich in proline and glutamine, residues that are not preferred substrates for any human digestive enzyme. This renders gluten relatively resistant to gastrointestinal proteolysis and allows the persistence in the small intestine of potentially immunogenic gluten-derived peptides.

The HLA-DQ molecules expressed by celiac patients, mainly HLA-DQ2.5 and to a lesser extent HLA-DQ8, bind these peptides and present them to $CD4^+$ T cells within the intestinal mucosa, thus inducing the strong pathologic cascade culminating in destruction of the duodenal mucosal architecture and in production of auto- and anti-gliadin antibodies. [1,2] Deamidation of gluten peptides by tissue transglutaminase (TG2) action [3,4] enhances their affinity for HLA-DQ2, making them better T-cell antigens.[5,6] The large variety of gliadin and glutenin derived T-cell epitopes characterized to date were recently reviewed and a standard nomenclature was proposed.[7]

One of the gluten epitopes that is most frequently recognized by celiac patients and that has been best characterized is DQ2.5-glia-α1a (previously named DQ2-α-I or α9).[7,8] The availability of the X-ray crystal structure of the soluble domain of HLA-DQ2.5 bound to the deamidated form of this epitope[9] made it possible to understand the general structural characteristics of the HLA-DQ2.5 complexes and to elucidate the molecular determinants of protein-peptide binding. In this complex, the binding groove is formed by the N-terminal domains of the DQ2 heterodimer and the eleven residues bound peptide (LQPFPQPELPY) adopts the left-handed polyproline II (PPII) helical conformation. Nine residues of the epitope interact with specific binding sites in the binding cleft, known as P1 – P9 registers (Figure 1A).

On the basis of these structural information and the modelling efforts, several groups have characterized the peptide-binding properties of HLA-DQ2 and have designed peptide ligands with

higher affinity for HLA-DQ2 with the aim of providing these compounds to celiac patients as therapeutically effective DQ2 blocking agents.[10,11,12]Another strategy proposed to overcome gluten peptides-DQ2 (and -DQ8) binding relies on the treatment of gluten with enzymes able to abolish the immunostimulatory activity of gliadin. Transamidation of gliadin extracted from wheat flour directly pre-treated with microbial transglutaminase and lysine methyl ester, as well as transamidation of gluten by both bacterial or human recombinant transglutaminases and lysine as amine donor, showed to be effective in blocking the T-cell mediated gliadin activity.[13,14]

Similarly to these last approaches, our goal is to alter gluten sequences in order to lower the epitope binding affinity for HLA-DQ2, thus reducing the (auto)immune response by modifying gliadin directly in flours. Among different functionalizations of gliadin that could lower the epitope affinity, we suggest the hydroxylation of Pro residues as the most suitable synthetic strategy since both enzymatic and chemical systems are available for this purpose.

The most studied hydroxylation reaction of proline to give 4-trans-hydroxyproline is the biochemical process stabilizing protocollagen in animals.[15] It mainly acts on the sequence X-Pro-Gly, which is transformed into the sequence X-Hyp-Gly. Hydroxylation rates depend on the nature of amino acid X.[16] Similar enzymes hydroxylate proline in position 3.[17] Moreover, there are Hypoxia Inducible Factors (HIFs) hydroxylases[18] which show less strict requirements as to the position of proline.[19] Prolyl-4-hydroxylases have also been described from algae and plants. They show different substrate specificity.[20] 3- and 4-hydroxyproline are found in bacteria as well, where hydroxylases are active on free proline instead of the peptide-bound proline.[21]

Chemical hydroxylations are based on Fenton's[22] and Udenfriend's[23] systems. Reduced oxygen species hydroxylate both free and peptide-bound proline and lysine.[24] Chemical systems are currently used for modifying food grade gelatins.[25] Regio- and stereoselectivity are usually reduced.

Before embarking on an experimental screening, we decided to approach a preliminary proof of concept by computational analysis, which is the main topic of the present paper. Here we report the computational study on the effects of hydroxylating the four Pro residues of DQ2.5-glia-α1a on HLA-DQ2.5 binding affinity. For this purpose, we developed and tested a docking protocol able to reproduce the binding geometry of the HLA-DQ2.5:DQ2.5-glia-α1a crystallographic complex. Then we performed docking calculations on fifteen DQ2.5-glia-α1a analogues obtained by substituting, one at a time or simultaneously, the Pro residues with the 3- or 4-hydroxyproline (3- and 4-Hyp), or with the 3,4-dihydroxyproline (3,4-diHyp). By evaluating their binding free energies and comparing them to that of the native peptide, we selected a functionalization that resulted as the most effective in lowering the binding affinity for HLA-DQ2.5. The same analysis performed on the DQ2.5-glia-α2[8] epitope confirmed our prediction. The analysis and comparison of the effects of the proposed functionalization in three celiac disease relevant epitopes, DQ2.5-glia-α1a, DQ2.5-glia-α1b,[26] and DQ2.5-glia-α2 allowed to analyze the molecular determinants of effectiveness in lowering their binding affinity for HLA_DQ2.

Results and Discussion

Binding of DQ2.5-glia-α1a and different Hyp and diHyp containing analogues

The crystallographic structure of HLA-DQ2.5 in complex with the DQ2.5-glia-α1a epitope (PDB ID: 1S9V)[9] includes two heterodimers: HLA-DQ2-I, with the A and B chains, and HLA-DQ2-II, with the D and E chains. A DQ2.5-glia-α1a peptide with eleven residues (LQPFPQPELPY) is bound to each dimer and presents the typical PPII helix elongated conformation.[9] The HLA-DQ2-II structure was selected for this computational study because of the low number of unsolved atomic coordinates. The structure presents the typical fold of the Major Histocompatibility Complex (MHC) class II proteins: the binding groove is delimited by two parallel α helices, each belonging to one of the dimer chains, and by a β-sheet of eight strands (four from each chain) at the back.[9]

From the analysis of the electronic properties of the residues in the binding cleft (Figure 1A) it emerges that the first three sites (P1 – P3 registers) are lined by hydrophobic residues and contain some polar side-chains, the central region (registers P4 – P7) exhibits a positive Molecular Electrostatic Potential (MEP) mainly due to the presence of Arg-E70 and Lys-E71 residues, and the P8 - P9 registers are mainly characterized by polar residues. The PPII arrangement of the peptide favours its structural complementarity with the binding cleft (Figure 1A); only the first two residues (Leu and Gln) are outside the protein binding site, whereas the remaining nonamer gives stabilizing interactions with the P1 - P9 registers, through both the backbone and the side-chains. In particular, the most relevant interactions between the peptide residues and the corresponding registers are: at P4 interactions involving the Gln residue; at P6, a complex network of H-bonds and electrostatic interactions with the Glu residue derived from deamidation by the TG2 enzyme, that contributes strongly to the complex stabilization[9]; at P9, important interactions with Tyr.[9] In this latter register, despite bulky hydrophobic residues have been identified as optimal anchor residues for a series of DQ2 ligands,[27] in this crystal structure the C-terminal Tyr side-chain of the epitope is positioned outside the pocket, where the hydroxyl group favourably interacts with external residues like Arg-D76. Finally, only two of the four Pro residues provide favourable interactions with the hydrophobic P1 and P3 sites while the two Pro in proximity of the P5 and P8 registers point the side-chain ring toward the solvent.

A peptide-protein docking protocol, based on the use of the Glide XP program,[28] was set up by selecting the computational choices able to correctly predict the geometry and interactions observed in the X-ray structure (redocking). To provide a complete atomic description of the interactions, the unsolved side-chain structures (Leu 1 and Gln 2 in the peptide and Asp-E135 and Gln-E136 in the receptor, lying far from the binding site) were modelled with the Prime program[29] by using libraries of known rotamers. To obtain the ligand input conformations for docking, a preliminary conformational analysis was performed on the free peptide. A number of experimental and theoretical studies have been focused on the conformational dynamics of polyproline peptides in aqueous solution and their propensity to adopt a PPII arrangement was proposed to vary as a function of the peptide length and the amino acidic sequence.[30,31,32,33] On the other hand, peptides bound to class II MHC proteins always adopt the PPII

helix conformation to optimally fit the binding groove. [33,34] Following the hypothesis that the shift of the dynamic equilibrium of gliadin epitopes toward this specific conformation is induced by the peptide-protein binding, the X-ray backbone structure of the bound peptide was maintained, and only the side-chain flexibility was included in the conformational analysis of the free peptide. A recent study performed by circular dichroism experiments on several DQ2.5-glia-α1a derivatives gave a strong confirmation to our hypothesis by proving that no direct correlation exists between the PPII helix propensity of the free peptides in solution and their HLA-DQ2 binding ability.[35]

To extend the geometrical sampling ability of the Glide XP docking program, not only the global minimum obtained by conformational analysis of the free peptide, but multiple (five) conformations corresponding to local minima in the potential energy landscape were submitted to docking. Then a semi-rigid docking protocol was tested, in which the experimental protein structure was kept rigid and the peptide flexibility was partially included during docking. In this stage, only the side-chain flexibility of the peptide was considered in the docking sampling.

Redocking results are shown in Figure 1B, where the experimental binding geometry and the docking pose are superimposed. It can be observed that the correspondence between protein registers and epitope residues as well as the orientation of the side-chains involved in stabilizing interactions were fully reproduced in the docking pose. Deviations from the experimental geometry are limited to the side-chains of the first two residues (Leu and Gln), not interacting with the protein binding registers, and of Phe at P2 that points toward the solvent. The good agreement is confirmed by the RMSD value between the peptide backbone atoms in the X-ray structure and in the docking pose (0.61 Å).

On the basis of these encouraging results, the same computational protocol (see Methods for details) was extended to the study of fifteen DQ2.5-glia-α1a analogues obtained by hydroxylation of the Pro residues, to analyze the effects of these functionalizations on the HLA-DQ2 binding affinity.

Twelve analogues were designed by substituting each Pro residue (at the P1, P3, P5 and P8 registers), one at a time, with the (2S,3R)-3-hydroxyproline (3-Hyp), the (2S,4R)-4-hydroxyproline (4-Hyp), and the (2S,3S,4S)-3,4-dihydroxyproline (3,4-diHyp). Moreover, three analogues were built by substituting all the Pro residues simultaneously with the 3-Hyp, 4-Hyp and 3,4-diHyp residues. The computational protocol that was set up for DQ2.5-glia-α1a, including conformational analysis of the free peptide and semi-rigid docking of five selected peptide conformations to HLA-DQ2, was extended to all the hydroxylated peptides. To analyze the stabilities of the obtained complexes and to compare them with that of DQ2.5-glia-α1a, the docking poses were re-scored by calculating their binding free energies with the Prime MM-GBSA program[29] (see Methods for details). The geometric and energetic characteristics of the best docking poses obtained are reported in Table 1.

To analyze the potential geometrical distortion due to the different Pro hydroxylations, the backbone orientations of the peptides in the binding groove were compared to that of the native peptide. In all the poses, the conformations adopted by the fifteen hydroxylated epitopes are very close to the reference X-ray structure of DQ2.5-glia-α1a. In fact the RMSD values, considering the complete peptide backbone atoms, are less than 0.80 Å. The RMSD values are lower if only the portion of the backbone directly involved in the interactions with the P1 – P9 registers is considered. The second and most relevant analysis was performed by comparing the estimated binding free energies of the fifteen Hyp and diHyp containing analogues to that of the DQ2.5-glia-α1a peptide. Different effects of Pro hydroxylation on the binding free energy were observed. From our calculations, both single and double hydroxylations of Pro residues at P5 and P8 are predicted to give binding affinities very similar to the reference peptide, with a slight stabilization of the complex (from 1 to 5 kcal mol^{-1}). This can be attributed to the Pro side-chain orientation toward the solvent, where the hydroxyl groups of the Hyp and diHyp residues can give additional favorable interactions with the water molecules. Hydroxylation of Pro at P1 gave opposite results: the 4-Hyp slightly destabilizes the complex due to steric clash with the residues in the P1 site (in particular with Trp-D43), whereas the 3-Hyp gives stabilizing effects. Finally, hydroxylations of Pro at P3 are predicted to always decrease the complex stability and this effect is enhanced in the peptides containing Hyp and diHyp at P3 along with other hydroxylations. Independently, a very recent work confirmed this computational prediction by experiments with a HPLC-based peptide exchange assay. The Authors demonstrated that four peptides derived from the DQ2.5-glia-α1a and containing Hyp at P3 fail to form the complex with HLA-DQ2.[35] From data in Table 1 we predict that, among the DQ2.5-glia-α1a functionalizations here proposed, the most effective one in decreasing the binding affinity for HLA-DQ2.5 is the simultaneous hydroxylation of all the Pro residues. In particular, DQ2.5-glia-α1a-4-Hyp-1-3-5-8 is predicted to destabilize the complex of about 12 kcal mol^{-1} and DQ2.5-glia-α1a-3,4-diHyp-1-3-5-8 of 25 kcal mol^{-1}.

Binding of three DQ2.5-α-gliadin epitopes and their Hyp and diHyp containing analogues

To verify if the most effective functionalization predicted for the HLA-DQ2:DQ2.5-glia-α1a complex gives the same effects also for other celiac disease relevant epitopes, the docking study was extended to DQ2.5-glia-α1b[26] and DQ2.5-glia-α2[8] (in the following the peptide names will be shortened to glia-α1a, glia-α1b, glia-α2). These α-gliadin epitopes belong to the group of dominant gluten-derived epitopes known to stimulate responses of T cells from DQ2.5 positive patients.[7,36] Both show high sequence similarity with glia-α1a. From the sequence alignment[8, 26] of their nine amino-acid core (Figure 2), it can be observed that the glia-α1b nonamer differs from that of glia-α1a only for the residue at P2 (Tyr instead of Phe), whereas glia-α2 differs for a sequence backward shift of two positions in the receptor registers that causes a different relative position of the third Pro residue (at P6 instead of P5).

Since the structures of the two epitopes in complex with HLA-DQ2.5 have not been determined to date, for our computational studies their backbone was modelled from the known X-ray structure of glia-α1a.[9] As in the previous study, a preliminary conformational analysis of each free peptide was performed by keeping the backbone structure rigid, and five conformational minima were retained and submitted to docking. To focus the attention on the interactions in the HLA-DQ2 binding groove,

only the core nonamers were docked. In this stage, the docking approach previously described was modified by including the entire peptide flexibility. In fact, while it is expected that glia-α1b adopts a structural arrangement in the binding groove similar to that of glia-α1a, the bound glia-α2 epitope could present a different fold due to the different relative position of the Pro residues. On the other hand, docking results obtained for the glia-α1a nonamer with this modified protocol confirmed its ability in finding a pose in good agreement with the crystal structure. The RMSD of the backbone atoms is lower than 1 Å (Table 2) and all the five conformations used as input structures produced a reliable pose (backbone RMSD in Table S1† for details).

The best pose obtained with this computational protocol for each epitope is shown in Figure 2, where only the backbone fold and the position of the Pro residues are evidenced. As before, the relative orientations of the peptides' backbone with respect to the nine registers were compared. All the three poses respected the correspondence between the peptide residues and the HLA-DQ2 registers. Moreover, similarly to what observed for the glia-α1a nonamer, for both the glia-α1b and the glia-α2 epitopes the binding poses obtained from all the five peptide conformations met this requirement (Table S1† for details).

As expected, due to the different relative position of the third Pro residue, the best pose of glia-α2 differs from the other two in the region of the P4-P7 registers (Figure 2). The RMSD values in the three regions corresponding to the P1-P3, P4-P7, and P8-P9 registers (Table 2) confirm that the glia-α2 backbone deviates from that of the other peptides only in the P4-P7 region (RMSD greater than 2 Å). This different conformation does not alter the number of stabilizing H-bonds with the receptor estimated by Glide XP in the docking poses (Table 2). However the residues involved in the H-bond network are different: for glia-α2 five H-bonds are in the P4-P7 region, due to the specific backbone conformation that allows better contacts with Lys-E71 and surrounding side-chains (Table S2† for details).

The best pose of glia-α1b is very similar to that of glia-α1a. An interesting difference regards the residue at P9 where, differently to both the X-ray structure[9] and the docking pose of glia-α1a, the Tyr side-chain is positioned inside the P9 pocket. This difference explains the relatively high RMSD value (2,58 Å) between the backbone of the two peptides in the P8-P9 region (Table 2). However this does not alter the estimated binding free energy ($\Delta\Delta G_{bind}$ between the best poses of glia-α1a and glia-α1b of about 2 kcal/mol) and both the external and internal orientations of the Tyr side-chain are observed for glia-α1a and glia-α1b among the five docking poses, with little differences in $\Delta\Delta G_{bind}$ (Table S1† for details). It is conceivable that both these orientations are assumed by residues at P9 in gliadin epitopes and this could explain the apparent discrepancy between the anchor role observed for bulky hydrophobic residues at this position[27] and the external positioning of the Tyr side-chain in the glia-α1a crystal structure.[9]

The prediction made on the basis of the analysis of 15 analogues of glia-α1a, that the functionalization most effective in decreasing the HLA-DQ2 binding affinity is the simultaneous di-hydroxylation of all the Pro residues, can be reasonably extended to glia-α1b, because of the same relative position of Pro residues in the sequence. On the contrary, to further verify the role of the single Pro functionalization in glia-α2, the same systematic approach adopted for glia-α1a was extended to this epitope. It included: design of the 15 glia-α2 analogues (Table S2† for details); conformational search to select five conformers; docking calculations including all the degrees of freedom of the peptides (see Methods for details).

All of the 15 glia-α2 analogues were correctly placed in the HLA-DQ2 binding groove with RMSD of the backbone lower than 2 Å with respect to the glia-α2 best pose (Table S2† for details). Moreover, comparison of the estimated binding free energy of each analogue to that of the native peptide highlighted that Pro hydroxylations in the P6 and P8 registers slightly alter ΔG_{bind} (from -3 to 5 kcal mol^{-1}). This can be attributed either to interactions with the solvent (P8) or to interactions with a polar register (P6). On the contrary, all of the Pro functionalizations in both the P1 and P3 registers were predicted to destabilize the complex (from 3 to 25 kcal mol^{-1}). Finally, in agreement with the results obtained for glia-α1a, both the single and the double simultaneous hydroxylations of all the Pro residues were predicted to destabilize the complex, and the most effective analogue resulted glia-α2-3,4-diHyp-1-3-6-8 (about 34 kcal mol^{-1}).

On the basis of the results obtained for both the glia-α1a and glia-α2 Hyp and diHyp containing analogues, docking calculations were extended to the glia-α1b nonamer with all the four Pro residues substituted by 3,4-diHyp residues. Like for the other epitopes, all the degrees of freedom of the peptide were considered in the docking process.

The effects of simultaneous di-hydroxylation of all the Pro residues on binding to HLA-DQ2 were carefully analyzed and compared among the three epitopes.

In all the three functionalized peptides, destabilizing effects of diHyp residues were firstly revealed by some difficulties of Glide XP in correctly assigning peptide residues to receptor registers. Differently from the native nonamers, only 1 to 3 poses correctly fitting the nine registers were obtained, starting from the five input conformations (Table S1† for details). From an energetic point of view, results confirmed that the double hydroxylation of all the Pro residues is effective in destabilizing the three complexes with HLA-DQ2 (Table 2). For glia-α1a, the diHyp containing analogue and the native nonamer have a ΔG_{bind} difference of about 25 kcal mol^{-1}, thus confirming the preliminary docking results obtained with the rigid backbone constrains (compare Tables 1 and 2). For the other two epitopes, this modification of the Pro residues reduces the affinity for HLA-DQ2 of about 30 (glia-α1b) and 34 kcal mol^{-1} (glia-α2).

From a methodological point of view, the choice of docking multiple peptide conformations to extend the sampling capability of Glide XP turned out very appropriate in particular for the hydroxylated analogues. In fact, while five reliable and similar binding poses were obtained from the different input conformations of the native epitopes, destabilization of the diHyp containing analogues was strongly evidenced by the few poses obtained with the peptide correctly placed in the receptor registers, in addition to the higher ΔG_{bind} values (Table S1† for details). It has to be noticed that similar effects were observed by using both the docking approaches, one including only the side-chain flexibility and the other considering the entire peptide

flexibility (data not shown).

To explain the molecular determinants of the predicted destabilization, it is interesting to compare the docking poses of each diHyp containing analogue with that of the corresponding native epitope.

The conformations adopted by the glia-α1a-3,4-diHyp and glia-α1b-3,4-diHyp epitopes are very similar. As shown in Figure 3 for glia-α1a, the Pro hydroxylation mainly affects the binding pose in the P1 - P3 region. As previously highlighted both from docking predictions and experimental evidences[35], it results that Pro hydroxylations at P3 greatly reduce the binding affinity of this epitope. As shown in Table 2, the poses of glia-α1a and glia-α1b are characterized by nine H-bonds, while the corresponding functionalized analogues present six or seven H-bonds. The destabilization is more evident comparing the backbone H-bond network of the peptides. For example, seven of the nine intermolecular H-bonds of glia-α1a involve the peptide backbone (Table 2), with respect to the two of glia-α1a-diHyp. In particular, despite diHyp generates a favourable H-bond network with the backbone of Tyr-D9 in P3, the presence of this residue negatively affects the H-bond network in the P1 – P3 region. The backbone H-bonds of glia-α1a involve the peptide residues in P2 (Phe), P4 (Gln) and P6 (Glu) and the receptor residues Asn-E82, Tyr-D9 and Asn-D62. These interactions are completely missing in the functionalized peptide pose (Table S3† for details). The same differences in this H-bond network are evident also for glia-α1b (Table S3† for details).

In its docking pose (Figure 4), the di-hydroxylated glia-α2 backbone shows an overall displacement from that of the native peptide of 3.04 Å (Table 2), thus causing a reduction in the number of favourable interactions between the peptide and the receptor. The number of H-bonds decreases from nine to six and, in particular, the backbone H-bonds decrease from seven to three (Table 2). Also in this case, the Pro residues at P1 and P3 negatively affect the interaction, causing a displacement of the backbone of 2.52 Å in this region. In fact, both the hydroxyl groups of the diHyp in P3 are involved in H-bonds (with Tyr-D9 and Tyr-D22), leading to the loss of the backbone H-bond stabilizing network in this region. In addition, for this peptide the destabilization is associated also with the loss of the H-bond network mediated by Lys-E71 in the P4 - P7 region (Figures 4B and 4C and Table S3†). In the native peptide this network includes the sidechain of Glu (P4) and the backbone of Leu (P5), while in the functionalized peptide no H-bonds are observed in the P4 – P7 region (Table S3† for details).

Conclusions

One of the most promising strategies to block the T-cell mediated gluten activity involved in celiac disease is the treatment of gluten aimed at inhibiting binding of the gluten-derived peptides to HLA-DQ2 (and -DQ8) proteins.[13, 14] Given the high content of Pro residues in gluten proteins and the availability of several enzymatic[16-21] and chemical[22-26] systems for proline hydroxylation, we propose the modification of Pro to Hyp or diHyp as a suitable synthetic strategy to alter gluten sequences directly in flour in order to lower epitope binding affinity for HLA-DQ2.

In this paper, a series of 15 Hyp and diHyp containing analogues of DQ2-5-glia-α1a, one of the epitopes most frequently recognized by celiac patients,[7, 8] was designed and submitted to a preliminary computational screening. A peptide-protein docking protocol was set up to study their binding to HLA-DQ2.5 and redocking results on the X-ray structure of the HLA-DQ2:DQ2.5-glia-α1a complex[9] confirmed its ability to correctly predict the binding geometry for these systems.

Among the designed glia-α1a analogues that contain a unique Hyp or diHyp residue, the ones hydroxylated at the position corresponding to the P3 receptor register were predicted to be the most suitable to inhibit binding. Recently, this prediction was confirmed by experiments indicating that glia-α1a analogues containing 4-Hyp at P3 fail to bind HLA-DQ2.[35] From our computational studies, it results that the destabilization of the peptide-protein complex is strongly increased by simultaneous hydroxylation or di-hydroxylation of all the Pro residue (at the P1, P3, P5 and P8 receptor positions). At a molecular level, binding affinity reduction for these epitopes is mainly due to the loss of stabilizing interactions in the P1 – P3 region.

These findings were confirmed by an analogous analysis of the docking poses of fifteen glia-α2 derived epitopes with single and multiple Hyp or diHyp residues.

The comparison of the results obtained for three of the most important celiac disease relevant epitopes,[7,36] DQ2.5-glia-α1a, DQ2.5-glia-α1b and DQ2.5-glia-α2, indicated that the di-hydroxylation of all the Pro residues is effective in lowering the binding affinity to HLA-DQ2.5 of all these peptides. Given the highly similar nine amino-acid core of glia-α1b and glia-α1a and the conservation of Pro residues at the same receptor registers, binding modes and intermolecular interactions obtained for glia-α1b and glia-α1b-3,4-diHyp were very similar to those predicted for glia-α1a and glia-α1a-3,4-diHyp. On the contrary, different binding modes were predicted both for the native glia-α2 and its diHyp containing analogue as a consequence of the different nonamer sequence and Pro residue positions. For this epitope, the loss of interactions with the receptor induced by diHyp residues is not only localized at the nonamer N-term (P1 - P3 registers) but it also involves the P4 - P7 region.

On the whole, the results confirmed our hypothesis on the destabilizing effects of Pro hydroxylation in gluten-derived peptide HLA-DQ binding and indicated the complete hydroxylation of the four Pro residues of glia-α1a, -α1b and -α2 as the most effective functionalization. Therefore, these hydroxylated analogues are expected to significantly affect gluten activity involved in celiac disease. This proposal need to be submitted to experimental validation before proceeding toward the development of the most suitable synthetic method for modifying gluten in flour.

Since the computational approach here proposed was able to highlight differences in the epitope binding in terms of both the binding geometry and the gain or loss of specific intermolecular interactions, it results as a promising tool to gain insight into the differences in binding of the large variety of gliadin and glutenin derived T-cell epitopes characterized to date[7] and of their variants.

Materials and Methods

Peptide conformational search

Conformational search of the peptides in the free state was performed using the MacroModel v9.9[37] program included in Maestro v9.3.[38] Backbone atoms were considered frozen, whereas sidechains were considered flexible. Conformational analysis was performed by using the Monte Carlo Multiple Minimum (MCMM) random search algorithm[39] with a maximum of 40000 MC steps, and 100 MC steps for each rotable bond. The Truncated Newton Conjugate Gradient (TNCG)[40] minimization algorithm was used, with a maximum number of iterations of 1500 and convergence on a gradient threshold of 0.05 kJ mol^{-1} $Å^{-1}$. The OPLS-2005 force field[41] was selected and water solvation was included by using the Generalized Born/Solvent Accessible (GB/SA)[42] implicit model. The conformations obtained were geometrically clustered using a RMSD cut-off of 2 Å and five representative conformations were selected for each peptide.

Molecular docking

Molecular docking was carried out using the XP version of the Glide v5.8[28,43,44] program included in Maestro v9.3.[38] Glide XP uses multiple hierarchical filters to search for possible locations of the ligand in the active-site of the receptor. The structure and the physico-chemical properties of the receptor are represented on a grid. Ligand conformational flexibility is handled in Glide XP by an extensive conformational search, augmented by a heuristic screen that rapidly eliminates unsuitable conformations. The final scoring of the poses is carried out by the Schrödinger's proprietary GlideScore XP scoring function.[45]

To model the interactions between HLA-DQ2 and the α-gliadin peptides, a grid side of 50 Å (corresponding to the maximum dimension allowed) was set up. The grid was centered in the averaged X, Y, Z coordinates of the receptor residues within 5 Å from DQ2.5-glia-α1a in the crystal structure of the complex.

For each peptide, the five conformers previously selected were employed as input conformations. Two different protocols were developed to include the peptide flexibility during docking: a) for the 15 Hyp and diHyp containing analogues of DQ2.5-glia-α1a the backbone conformation was kept fixed and only the flexibility of the peptide sidechains was considered; b) for the 15 Hyp and diHyp containing analogues of DQ2.5-glia-α2 as well as for the comparison of DQ2.5-glia-α1a, DQ2.5-glia-α1b, DQ2.5-glia-α2 and their diHyp containing analogues, the degrees of freedom of the whole peptide were included. The other Glide XP parameters were set up to widen the docking funnel. In particular, the number of poses per ligand to pass to the grid refinement calculation were increased from 5000 to 50000, and the maximum number of poses per ligand to pass to the grid refinement calculation was increased from 400 to 1000; the initial rough-score cut-off was increased from 100 to 500 kcal mol^{-1}; the maximum number of steps taken by the conjugate gradient minimization algorithm was increased from 100 to 1500. At the end of the Glide process, only the best pose was saved for each input conformation of each peptide.

Both to select the best pose for each peptide and to evaluate the binding free energy (ΔG_{bind}) of the obtained protein-ligand complexes, molecular mechanics generalized Born/surface area (MM-GBSA) calculations were performed, by using an energy minimization protocol for the free ligand, the free protein, and their complex as a basis for calculating ΔG_{bind}. This calculation was performed using Prime MM-GBSA v2.0,[29] excluding entropic terms, with a flexible receptor shell within 7 Å from the ligand.

Acknowledgments

Financial support by University of Milano-Bicocca and Italian Regione Lombardia (European Social Fund ESF) is gratefully acknowledged.

Abbreviations

TG2, tissue transglutaminase; PPII, polyproline II; HIFs, Hypoxia Inducible Factors; 3-Hyp, (2S,3R)-3-hydroxyproline; 4-Hyp, (2S,4R)-4-hydroxyproline; 3,4-diHyp, (2S,3S,4S)-3,4-dihydroxyproline; MHC, Major Histocompatibility Complex; MEP, Molecular Electrostatic Potential; ΔG_{bind}, binding free energy; MCMM, Monte Carlo Multiple Minimum; TNCG, Truncated Newton Conjugate Gradient; GB/SA, Generalized Born/Solvent Accessible; MM-GBSA, molecular mechanics generalized Born/surface area.

Notes and references

[a] *Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza 1, 20126 Milano, Italy, E-mail: laura.bonati@unimib.it*
[b] *Pharma Marine AS, Terøy, 6280 Søvik, Norway*
[c] *FIIRV, Fondazione Istituto Insubrico di Ricerca per la Vita, Via R. Lepetit 34, 21040 Gerenzano (VA), Italy*

† Electronic Supplementary Information (ESI) available. Table S1: Docking results, expressed as Prime MM-GBSA ΔG_{bind}, for the selected five conformers of α-gliadin epitopes and their diHyp containing analogues,Table S2: Docking results, expressed as Prime MM-GBSA ΔG_{bind}, for DQ2.5-glia-α2 and 15 Hyp and diHyp containing analogues, Table S3: H-bond networks in the best docking pose of each nonamer.See DOI: 10.1039/b000000x/

1. P. H. Green and C. Cellier, *N. Engl. J. Med.*, 2007, **357**, 1731-1743.
2. D. Schuppan, Y. Junker and D. Barisani, *Gastroenterology*, 2009, **137**, 1912-1933.
3. O. Molberg, S. N. McAdam, R. Korner, H. Quarsten, C. Kristiansen, L. Madsen, L. Fugger, H. Scott, O. Noren, P. Roepstorff, K. E. A. Lundin, H. Sjostrom and L. M. Sollid, *Nat. Med.*, 1998, **4**, 713-717.
4. L. Shan, O. Molberg, I. Parrot, F. Hausch, F. Filiz, G. M. Gray, L. M. Sollid and C. Khosla, *Science*, 2002, **297**, 2275-2279.
5. H. Sjöström, K. E. Lundin, O. Molberg, R. Körner, S. N. McAdam, D. Anthonsen, H. Quarsten, O. Norén, P. Roepstorff, E. Thorsby and L. M. Sollid, *Scand. J. Immunol.*, 1998, **48**, 111-115.
6. Y. van de Wal, Y. Kooy, P. van Veelen, S. Peña, L. Mearin, G. Papadopoulos and F. Koning, *J. Immunol.*, 1998, **161**, 1585-1588.
7. L. M. Sollid, S. W. Qiao, R. P. Anderson, C. Gianfrani and F. Koning, *Immunogenetics*, 2012, **64**, 455-460.
8. H. Arentz-Hansen, R. Körner, O. Molberg, H. Quarsten, W. Vader, Y. M. Kooy, K. E. Lundin, F. Koning, P. Roepstorff, L. M. Sollid and S. N. McAdam, *J. Exp. Med.*, 2000, **191**, 603-612.

9. C. Y. Kim, H. Quarsten, E. Bergseng, C. Khosla and L. M. Sollid, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 4175-4179.
10. U. Jüse, Y. van de Wal, F. Koning, L. M. Sollid and B. Fleckenstein, *Hum. Immunol.*, 2010, **71**, 475-481.
11. V. V. Kapoerchan, M. Wiesner, U. Hillaert, J. W. Drijfhout, M. Overhand, P. Alard, G. A. van der Marel, H. S. Overkleeft and K. F., *Mol. Immunol.*, 2010, **47**, 1091-1097.
12. J. Xia, M. Siegel, E. Bergseng, L. M. Sollid and C. Khosla, *J. Am. Chem. Soc.*, 2006, **128**, 1859-1867.
13. C. Gianfrani, R. A. Siciliano, A. M. Facchiano, A. Camarca, M. F. Mazzeo, S. Costantini, V. M. Salvati, F. Maurano, G. Mazzarella, G. Iaquinto, P. Bergamo and M. Rossi, *Gastroenterology*, 2007, **133**, 780-789.
14. L. Elli, L. Roncoroni, M. Hils, R. Pasternack, D. Barisani, C. Terrani, V. Vaira, S. Ferrero and M. T. Bardella, *Hum. Immunol.*, 2012, **73**, 992-997.
15. K. L. Gorres and R. T. Raines, *Crit. Rev. Biochem. Mol. Biol.*, 2010, **45**, 106-124.
16. K. I. Kivirikko, Y. Kishida, S. Sakakibara and D. J. Prockop, *Biochim. Biophys. Acta*, 1972, **271**, 347-356.
17. R. M. Gryder, M. Lamon and E. Adams, *J. Biol. Chem.*, 1975, **250**, 2470-2474.
18. G. L. Wang, B. H. Jiang, E. A. Rue and G. L. Semenza, *Proc. Natl. Acad. Sci. U.S.A.*, 1995, **92**, 5510-5514.
19. D. Li, M. Hirsilä, P. Koivunen, M. C. Brenner, L. Xu, C. Yang, K. I. Kivirikko and J. Myllyharju, *J. Biol. Chem.*, 2004, **279**, 55051-55059.
20. M. K. Koski, R. Hieta, M. Hirsilä, A. Rönkä, J. Myllyharju and R. K. Wierenga, *J. Biol. Chem.*, 2009, **284**, 25290-25301.
21. J. E. Baldwin, R. A. Field, C. C. Lawrence, V. Lee, J. K. Robinson and C. J. Schofield, *Tetrahedron Lett.*, 1994, **35**, 4649-4652.
22. E. R. Stadtman, *Annu. Rev. Biochem.*, 1993, **62**, 797-821.
23. S. Udenfriend, C. T. Clark, J. Axelrod and B. B. Brodie, *J. Biol. Chem.*, 1954, **208**, 731-739.
24. R. L. Trelstad, K. R. Lawley and L. B. Holmes, *Nature*, 1981, **289**, 310-312.
25. M. S. Hoque, S. Benjakul and T. Prodpran, *Food Chem.*, 2011, **128**, 878-888.
26. H. Arentz-Hansen, S. N. McAdam, Ø. Molberg, B. Fleckenstein, K. E. Lundin, T. J. Jørgensen, G. Jung, P. Roepstorff and L. M. Sollid, *Gastroenterology*, 2002, **123**, 803-809.
27. H. Quarsten, G. Paulsen, B. H. Johansen, C. J. Thorpe, A. Holm, S. Buus and L. M. Sollid, *Int. Immunol.*, 1998, **10**, 1229-1236.
28. *Schrödinger LLC, Glide, version 5.8*, New York, NY, 2012.
29. *Schrödinger LLC, Prime, version 3.1*, New York, NY, 2012.
30. I. Parrot, P. C. Huang and C. Khosla, *J. Biol. Chem.*, 2002, **277**, 45572-45578.
31. S. Doose, H. Neuweiler, H. Barsch and M. Sauer, *Proc. Natl. Acad. Sci. U.S.A*, 2007, **104**, 17400-17405.
32. M. Moradi, V. Babin, C. Roland, T. A. Darden and C. Sagui, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 20746-20751.
33. A. A. Adzhubei, M. J. Sternberg and A. A. Makarov, *J. Mol. Biol.*, 2013, **425**, 2100-2132.
34. T. S. Jardetzky, J. H. Brown, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger and D. C. Wiley, *Proc. Natl. Acad. Sci. U.S.A.*, 1996, **93**, 734-738.
35. S. D. Unudurthi, K. Hotta and C. Y. Kim, *ACS Chem. Biol.*, 2013, **8**, 2383-2387.
36. D. Stepniak, M. Wiesner, A. H. de Ru, A. K. Moustakas, J. W. Drijfhout, G. K. Papadopoulos, P. A. van Veelen and F. Koning, *J. Immunol.*, 2008, **180**, 3268-3278.
37. *Schrödinger LLC, MacroModel, version 9.9*, New York, NY, 2012.
38. *Schrödinger LLC, Maestro, version 9.3*, New York, NY, 2012.
39. G. Chang, W. C. Guida and W. C. Still, *J. Am. Chem. Soc.*, 1989, **111**, 4379-4386.
40. A. R. Leach, *Molecular modelling : principles and applications*, Essex, 2001.
41. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474-6487.
42. W. C. Still, A. Tempczyk, R. C. Hawley and T. A. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127-6129.
43. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739-1749.
44. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J. Med .Chem.*, 2004, **47**, 1750-1759.
45. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2006, **49**, 6177-6196.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Table 1 – Docking results, expressed as Prime MM-GBSA ΔG_{bind}, for DQ2.5-glia-α1a and 15 Hyp and diHyp containing analogues

Name	ΔG_{bind} (kcal mol^{-1})	$\Delta\Delta G_{bind}$ [a] (kcal mol^{-1})	backbone RMSD (Å)[b] All peptide	P1-P9
3-Hyp-8	-212.15	-4.96	0.60	0.21
3-Hyp-1	-209.57	-2.39	0.60	0.21
3,4-diHyp-8	-209.51	-2.33	0.62	0.28
3,4-diHyp-5	-209.04	-1.85	0.61	0.24
3-Hyp-5	-208.97	-1.79	0.64	0.31
4-Hyp-5	-208.91	-1.73	0.62	0.28
4-Hyp-8	-208.49	-1.31	0.62	0.25
DQ2.5-glia-α1a	-207.18	--	0.61	0.24
4-Hyp-1	-206.41	0.77	0.61	0.25
3,4-diHyp-1	-204.66	2.52	0.64	0.33
4-Hyp-3	-203.30	3.89	0.65	0.33
3,4-diHyp-3	-202.28	4.90	0.75	0.45
3-Hyp-1-3-5-8	-202.24	4.94	0.64	0.31
3-Hyp-3	-201.40	5.78	0.67	0.38
4-Hyp-1-3-5-8	-195.67	11.51	0.68	0.37
3,4-diHyp-1-3-5-8	-182.54	24.65	0.73	0.43

[a] with respect to the DQ2.5-glia-α1a docking pose

[b] with respect to the DQ2.5-glia-α1a crystal structure

Table 2 Docking results, expressed as Prime MM-GBSA ΔG_{bind}, for the nonamers of native DQ2.5-glia-α1a, DQ2.5-glia-α1b and DQ2.5-glia-α2 and their corresponding 3,4-diHyp analogues

Name	ΔG_{bind} (kcal mol^{-1})	$\Delta\Delta G_{bind}$ [a] (kcal mol^{-1})	backbone RMSD (Å) P1-P9	P1-P3	P4-P7	P8-P9	#H-bonds [d] Total	backbone [e]
DQ2.5-glia-α1a	-193.45	--	0.93 [b]	--	--	--	9	7
DQ2.5-glia-α1b	-195.48	--	1.24 [b]	0.27 [c]	0.29 [c]	2.58 [c]	9	6
DQ2.5-glia-α2	-182.78	--	2.60 [b]	1.17 [c]	2.37 [c]	1.07 [c]	9	7
DQ2.5-glia-α1a-3,4-diHyp	-168.60	24.85	1.80 [a]	2.62 [a]	1.73 [a]	1.33 [a]	6	2
DQ2.5-glia-α1b-3,4-diHyp	-165.62	29.87	0.94 [a]	1.70 [a]	1.38 [a]	2.25 [a]	7	4
DQ2.5-glia-α2-3,4-diHyp	-148.43	34.34	3.04 [a]	2.52 [a]	1.81 [a]	1.02 [a]	6	3

[a] with respect to the corresponding native nonamer

[b] with respect to the DQ2.5-glia-α1a crystal structure

[c] with respect to the DQ2.5-glia-α1a docking pose

[d] Computed by Glide XP[28] (donor-acceptor distance = 2.5 Å, donor angle = 120°, acceptor angle = 90°)

[e] H-bonds involving the backbone of the peptide

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

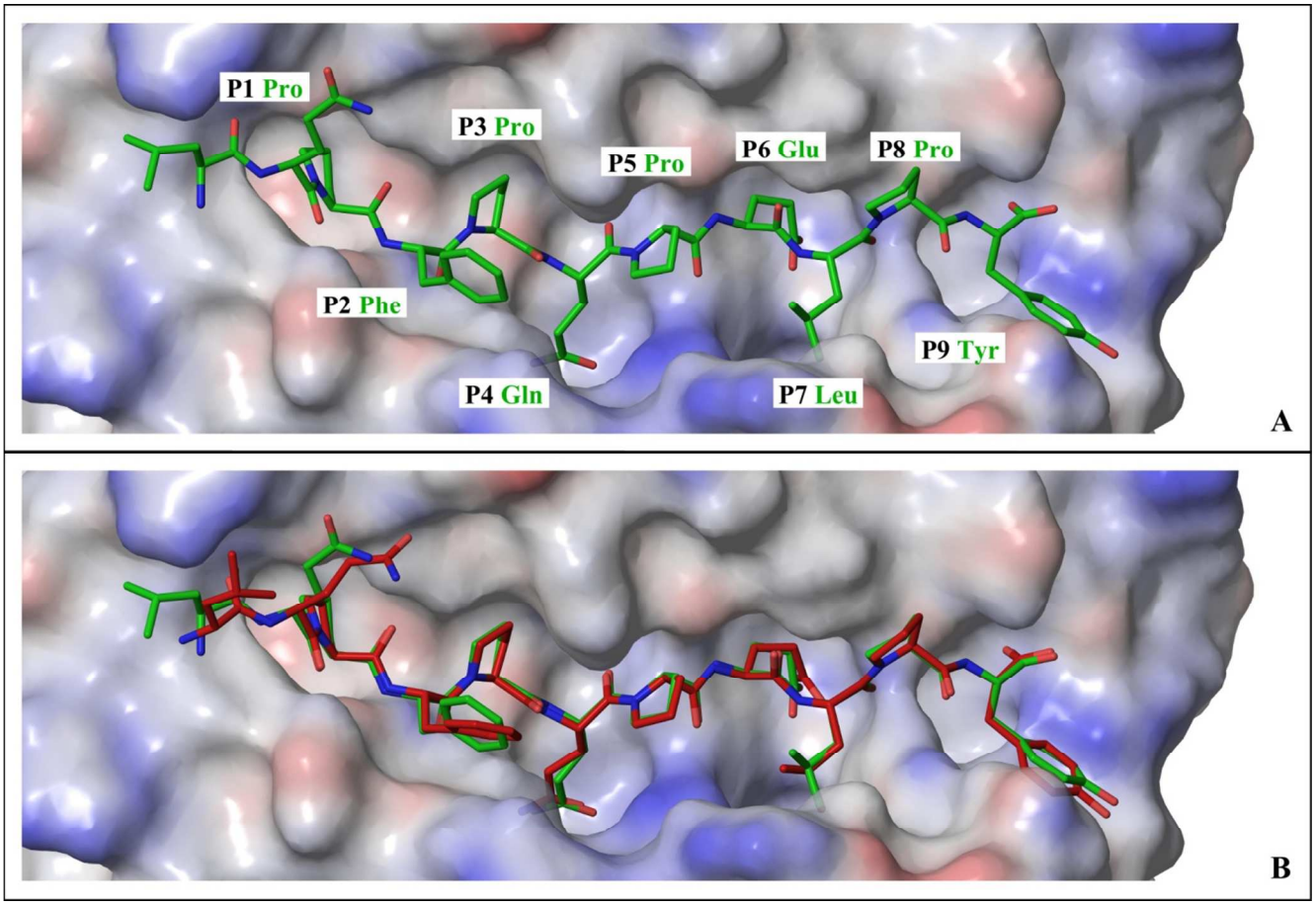


Fig.1 A) Details of the binding groove of HLA-DQ2.5 in the crystal structure of the complex with DQ2.5-glia-α1a (PDB ID: 1S9V)[9]. Registers are labeled P1 - P9 and the corresponding peptide residues are indicated. The receptor is represented by a surface colored according to the Molecular Electrostatic Potential values (blue = positive, red = negative). The peptide is represented by green sticks. B) The best docking pose of DQ2.5-glia-α1a (red) is superimposed to its crystal structure (green) and represented by sticks.

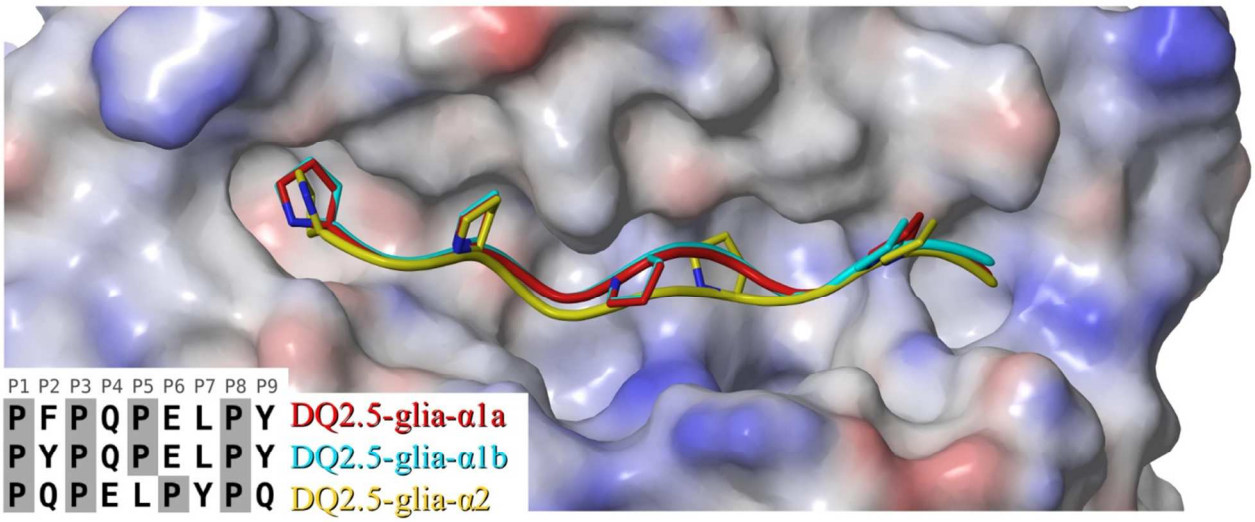


Fig. 2 Details of the best docking poses in the HLA-DQ2 binding groove. The receptor is represented as in Figure 1. The best poses (DQ2.5-glia-α1a - red, DQ2.5-glia-α1b – cyan, and DQ2.5-glia-α2 – yellow) are represented by cartoon trace of the backbone and sticks of the Pro residues. On the left corner, the nonamer sequences are reported and aligned to the corresponding HLA-DQ2 registers. Pro residues are highlighted by a grey background

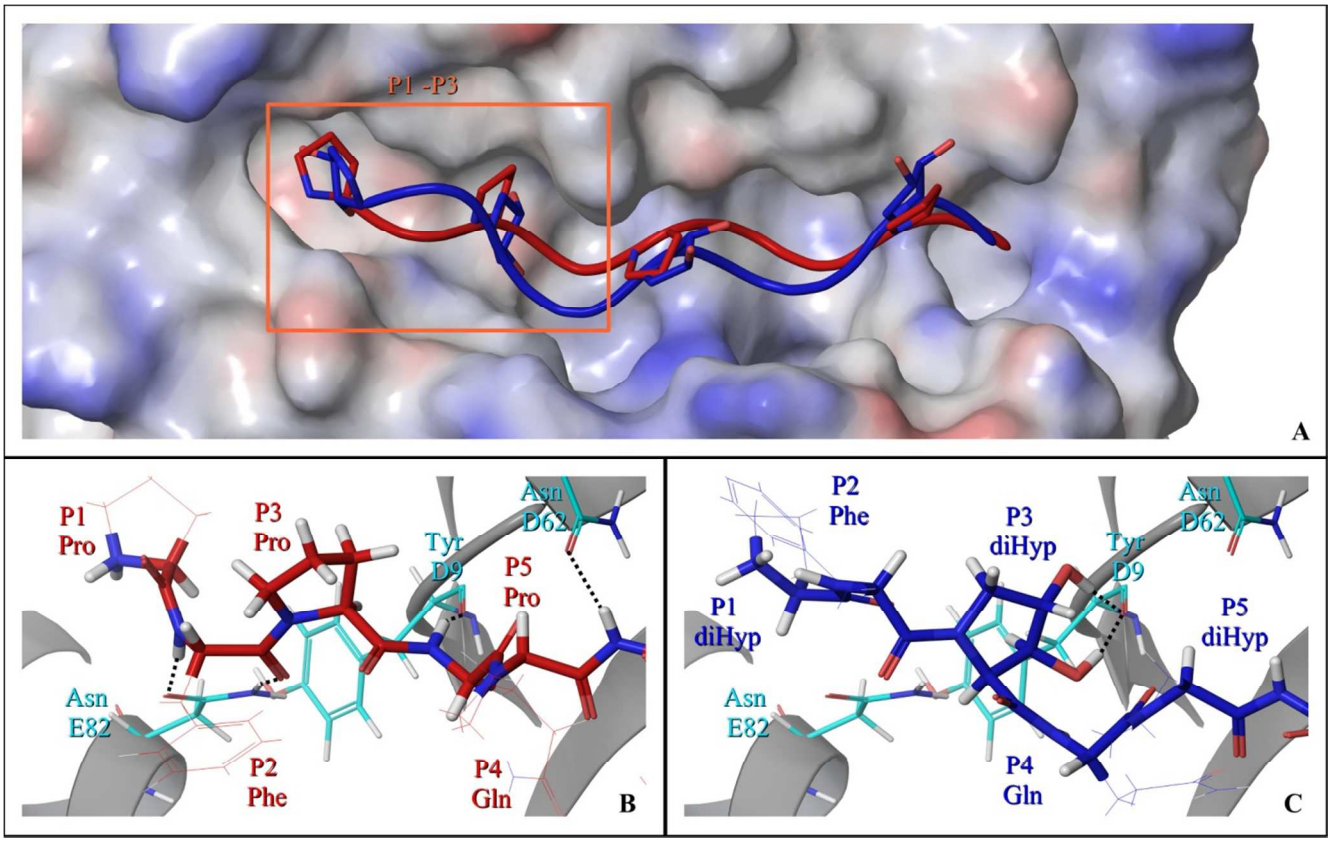


Fig. 3 A) Details of the best docking poses in the HLA-DQ2 binding groove. The receptor is represented as in Figure 1. The docking poses of DQ2.5-glia-α1a (red) and DQ2.5-glia-α1a-3,4-diHyp (blue) are structurally superimposed and the nonamers are represented by cartoon trace of the backbone and sticks of the Pro or 3,4-diHyp residues. The region of the binding groove involved in the putative H-bond network (P1 - P3) is highlighted by a box. B-C) Details of the interactions in P1 - P3. The nonamer backbones and the sidechains involved in the putative H-bonds are represented by sticks, the other sidechains are represented in wire and all the residues are labeled (DQ2.5-glia-α1a: red, DQ2.5-glia-α1a-3,4-diHyp: blue). For the receptor, the sidechains involved in the putative H-bonds are represented by cyan sticks and labels. The putative H-bonds are represented by black dashes

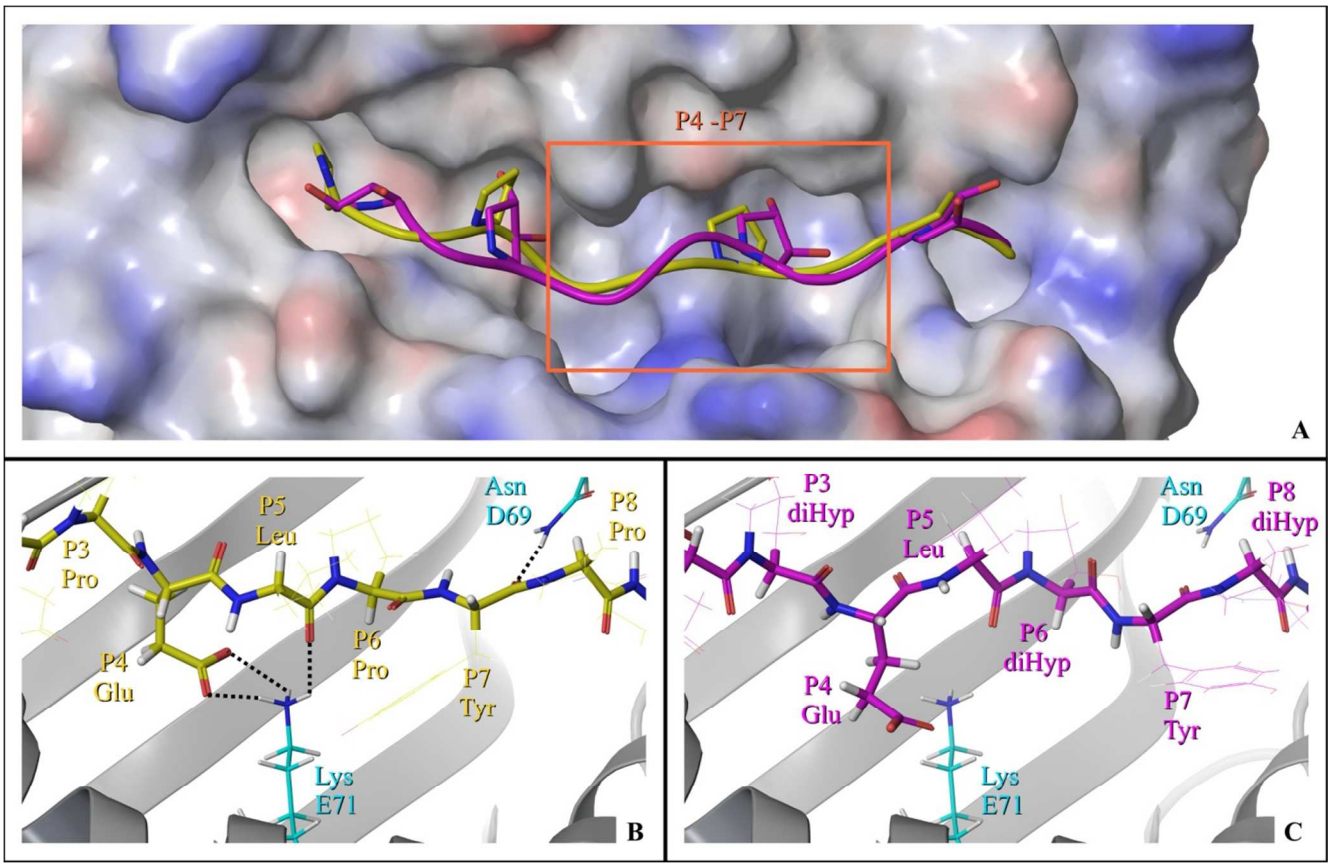


Fig. 4 A) Details of the best docking poses in the HLA-DQ2 binding groove. The receptor is represented as in Figure 1. The poses of DQ2.5-glia-α2 (yellow) and DQ2.5-glia-α2-3,4-diHyp (magenta) are structurally superimposed and the nonamers are represented by cartoon trace of the backbone and sticks of the Pro or 3,4-diHyp residues. The region of the binding groove involved in the putative H-bond network (P4 - P7) is highlighted by a box. B-C) Details of the interactions in P4 - P7. The nonamer backbones and the sidechains involved in the putative H-bonds are represented by sticks, the other sidechains are represented in wire and all the residues are labeled (DQ2.5-glia-α2: yellow, DQ2.5-glia-α2-3,4-diHyp: magenta). For the receptor, the sidechains involved in the putative H-bonds are represented by cyan sticks and labels. The putative H-bonds are represented by black dashes.