JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Matched Peptides: Tuning Matched Molecular Pair Analysis for Biopharmaceutical Applications

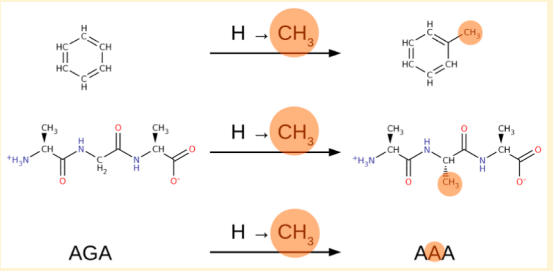
Julian E. Fuchs,*[,†] Bernd Wellenzohn,[‡] Nils Weskamp,[‡] and Klaus R. Liedl[†]

[†]Theoretical Chemistry, Faculty of Chemistry and Pharmacy, University of Innsbruck, Innrain 82, 6020 Innsbruck, Austria

[‡]Research Germany/Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach an der Riss, Germany

Supporting Information

ABSTRACT: Biopharmaceuticals hold great promise for the future of drug discovery. Nevertheless, rational drug design strategies are mainly focused on the discovery of small synthetic molecules. Herein we present matched peptides, an innovative analysis technique for biological data related to peptide and protein sequences. It represents an extension of matched molecular pair analysis toward macromolecular sequence data and allows quantitative predictions of the effect of single amino acid substitutions on the basis of statistical data on known transformations. We demonstrate the application of matched peptides to a data set of major histocompatibility complex class II peptide ligands and discuss the trends captured with respect to classical quantitative structure−activity relationship approaches as well as structural aspects of the investigated protein−peptide interface. We expect our novel readily interpretable tool at the interface of cheminformatics and bioinformatics to support the rational design of biopharmaceuticals and give directions for further development of the presented methodology.



INTRODUCTION

Biopharmaceuticals are defined as pharmaceutical products consisting of (glyco)proteins and/or nucleic acids.[1] Therefore, this class of drugs mainly comprises peptide hormones, recombinant proteins, monoclonal antibodies, and therapeutic antibodies. Biopharmaceuticals allow access to new target classes and are therefore considered more innovative than small-molecule drugs.[2] Accordingly, a record number of 11 new biopharmaceuticals were approved by the FDA in 2014.[3] Therefore, biopharmaceuticals hold promise to claim a larger share of the drug market in the future.[4] Additionally, biosimilars are increasingly entering the market after patent expiry of original biopharmaceutical products.[5]

Biopharmaceuticals generally pose new challenges for the drug discovery process, which has historically been focused on small molecules. This includes their analytical characterization,[6] delivery and formulation[7,8] after optimization of the biotechnological production process,[9,10] and their molecular properties.[11] Computational modeling techniques hold great promise to handle the complexity of the generated data and, for example, to guide affinity optimization of therapeutic proteins[12] or peptides.[13,14] Peptide drugs are often considered as the border between small-molecule drugs and biopharmaceuticals, as their synthesis is mainly chemistry-driven.[15]

Traditionally, quantitative structure−activity relationship (QSAR) and quantitative structure−property relationship (QSPR) modeling approaches neglect the three-dimensional (3D) structure of the peptides and proteins and are thus 2D-based. Nevertheless, approaches using 3D interaction fields[16] or comparative modeling techniques have been described.[17] These 3D techniques have to cover the bioactive conformation of the usually highly flexible peptide ligands, which poses additional challenges for modeling.[18] In a pioneering study, Sneath derived the first molecular descriptors for the 20 natural amino acids and applied them in QSAR modeling.[19] Later, these 2D descriptors were refined to capture chemically intuitive information via the *Z*-scale model[20] or the isotropic surface area/electronic charge index (ISA/ECI) model.[21] In contrast to substitution matrices frequently applied in bioinformatics (e.g., PAM,[22] BLOSUM[23]), these descriptors are designed to reflect chemical in contrast to evolutionary similarity. Amino acid descriptors have typically been used to derive QSAR equations by linear regression techniques.[24]

Over the past decade, the innovative cheminformatic concept of "matched molecular pair analysis"[25] has been gaining increasing attention. Herein, pairs of molecules with a single difference in chemical structure are analyzed with respect to changes in a physicochemical or biological property.[26] Data mining in large databases (e.g., bioactivities stored in ChEMBL[27] or in-house data sets[28]) allows trends from matched molecular pairs or matched molecular series to be applied subsequently for prediction of substitution effects in new molecules.[29] A key advantage of matched molecular pair analysis is the direct chemical interpretability of predictions ("white box") based on local SAR rules.[30] Recently, efforts have

Received: July 31, 2015
Published: October 26, 2015

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

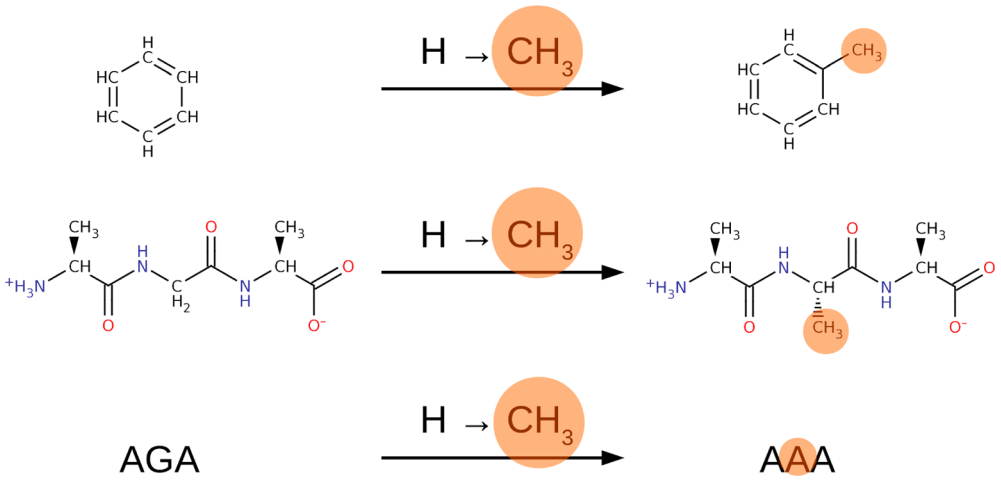


Figure 1. Matched molecular pairs and matched peptides. The correspondence of matched molecular pairs to matched peptides is exemplified by the chemical transformation of benzene to toluene by exchange of a hydrogen for a methyl group (orange). The same transformation is involved when an alanine-glycine-alanine tripeptide is exchanged with alanine-alanine-alanine. The latter transformation may be easily encoded when using standard one-letter amino acid codes. This representation allows for fast processing of large databases connecting sequences with respective molecular properties.

been made to put purely ligand-based matched molecular pairs into structural context and thereby identify the structural background of observed bioactivity trends.[31,32]

Herein we expand the scope of matched molecular pairs to the analysis of macromolecular data from proteins and peptides and introduce matched peptides, a concept we expect to hold great promise for the development of biopharmaceuticals. As an example application, we investigate peptide binding to the major histocompatibility complex class II (MHC II), a surface receptor crucial for T-cell activation in immune response.[33,34] A crystal structure of the receptor shows that the peptide is bound to a hydrophobic surface groove that is flanked by two α-helices.[35] Through the availability of structural information, most modeling approaches aiming at the prediction of peptide binding to MHC molecules employ machine learning techniques,[36] e.g., MULTIPRED.[37] Large peptide data sets have been compiled and used for the optimization of consensus approaches based on machine learning methods.[38] Application of these techniques allows for the optimization of peptides with desired immunological properties.[39] Quantitative modeling techniques are rarely applied toward MHC binding but include classical amino acid-descriptor-based QSAR methods[40] as well as molecular dynamics simulation approaches.[41] Here we apply the novel matched peptides strategy to the prediction of MHC II binding affinities and demonstrate the direct interpretability of the predictions in a structural context.

METHODS

Matched Molecular Pairs and Matched Peptides. Identification of matched molecular pairs involves an exhaustive pairwise matching of molecular graphs. To simplify this task, molecules are usually fragmented to aid the search for corresponding substructures.[42] Older implementations additionally required a definition of the allowed transformations within matched pairs,[43] thus prohibiting the identification of unknown chemical modifications associated with a change in the molecular property under investigation.

In the context of peptide and protein data, a molecular transformation corresponds to a point mutation. Therefore, sequences differing by a single character correspond to matched peptides. For identification of these single substitutions, a pairwise sequence alignment that can be performed by standard bioinformatics methodologies is required. Sequence alignment for matched peptides is trivial since they consistently differ by a single amino acid. This sequence alignment step simplifies the graph matching problem for small molecules described above, since linear peptides have the advantage of having defined C- and N-termini as well as identical chemical backbones (see Figure 1). Furthermore, insertions and deletions between sequence pairs can be considered as trivial additions to the exchange of single amino acids and thus may also be involved as additional transformations in matched peptides. Matched peptides therefore represent a special case of matched molecular pairs that are easy to detect on the sequence level.

Code implementation was performed using standard Python tools in combination with a custom node in KNIME[44] aiming to identify all sequence pairs differing in a single sequence position and thus forming matched peptide pairs. Since all of the sequences analyzed in the current study were point mutations relative to a consensus sequence and had a constant length, no gaps occurred, and the presented analysis is therefore independent of gap penalties in the alignment step.

Affinity differences observed for matched peptides were aggregated over all peptide positions, assuming that the effects of amino acid exchanges are independent of their position and thus reflect an average of all binding-site environments. We will demonstrate in the Results and Discussion section that this assumption is in general valid for MHC II binders and also discuss ways to cover position-specific aspects in matched peptide analysis.

Analyzed Data Set. We analyzed experimental binding affinities for a panel of 198 peptides toward MHC II molecules from Marshall et al.[45] Fluorescence-based assays were applied to obtain IC_{50} values in the nano- and subnanomolar range for all of the peptides using a 12-point inhibition curve for binding with three replicates each. Peptide sequence were grouped around the template peptide sequence AAYAAAAAAAAAA, where the central 11 amino acids of the peptide with length 13 were varied. For positions 2 and 4 to 12, all 20 natural amino acids were tested, while only seven apolar amino acids (F, I, L,

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

M, V, W, Y) were tested for position 3. All of the sequences represent single-point mutations around the template sequence. Additionally, the length of the peptide (13 amino acids) was kept constant, and thus, no insertions or deletions were present among the matched peptide sequences. All of the presented analyses are based on negative decadic logarithms of reported IC_{50} values and their ratios based on molar units.

Statistical Framework. On the basis of statistical analysis of bioactivity data from ChEMBL, the expected standard deviation of the IC_{50} data from a homogeneous source is 0.2 log units.[46] Therefore, an average effect size of at least 0.20 log units establishes statistical significance versus the null hypothesis (no change in activity) at the $p = 0.05$ level with at least 10 matched molecular pairs or matched peptides.[47] Therefore, the term "significant transformations" used later on explicitly refers to those amino acid substitutions associated with an effect on the binding affinity that is statistically significantly different to zero.

Correlation to Amino Acid Descriptors. We correlated trends in bioactivities (affinity shifts) observed from analysis of matched peptides to differences in amino acid properties. Therefore, we employed three descriptors from the *Z*-scale approach describing hydrophilicity (z_1), steric bulk (z_2), and electronic properties (z_3).[20] Furthermore, we analyzed correlations to differences in the isotropic surface area (ISA) and the electronic charge index (ECI).[21] Correlations between activity differences and property differences were assessed via calculation of Pearson's linear correlation coefficient r and Spearman's rank correlation coefficient ρ to capture both quantitative and qualitative dependences. Statistical analyses were performed using R.[48]

Structural Interpretation and Correlation to Peptide Specificity. To interpret the bioactivity data in structural context, we compared the observed trends to a cocrystal structure of an MHC II in complex with an antigenic peptide (PDB entry 1AQD[49]). We visualized the structure in Pymol[50] and extracted polar contacts as well as electrostatic properties of the binding-site region using default settings.

The specificities of respective MHC II binding-site regions were assessed on the basis of binding affinity distributions for single amino acids. Therefore, we converted the affinity ratios to decadic log units and analyzed the distribution of binding affinities for each single site. In the case of a highly specific region, major differences in binding affinity are expected, corresponding to a narrow peak in the distribution. On the contrary, a completely nonspecific position shows an equal distribution of binding affinities. Such experimental distributions can be converted to single values depicting local specificity via an information-entropy-based approach, as demonstrated earlier for amino acid distributions in protease substrates.[51] Thereby, an entropy of 0 corresponds to the highest specificity, whereas a value of 1 corresponds to maximum binding promiscuity with constant binding affinities. All of the peptide residues except for position 3, where only seven of the 20 amino acids were tested, were examined individually.

RESULTS AND DISCUSSION

Trends in Binding Affinity from Matched Peptide Analysis. Applying matched peptides, we extracted information on quantitative changes in MHC II binding affinity induced by single-point mutations. On the basis of 198 peptide sequences and their respective experimental binding affinities, we extracted trends on how amino acid substitutions increase and decrease molecular interactions via matched peptide analysis. In total we extracted 2117 matched peptides that formed the basis of the statistical evaluation. The order of identified matched pairs was normalized to consistently reflect gains in affinity.

The amino acid substitutions with the strongest effects on the observed binding affinities are summarized in Table 1.

Table 1. Transformations with Major Effects on the MHC II Binding Affinity[a]

transformation	pairs	mean affinity difference [log units]	SEM [log units]
P → C	10	0.715	0.291
P → Y	10	0.672	0.329
P → L	10	0.624	0.354
P → M	10	0.620	0.320
D → C	10	0.606	0.147
P → S	10	0.606	0.332
P → N	10	0.596	0.366
P → V	10	0.594	0.339
P → A	19	0.587	0.215
K → C	11	0.585	0.250

[a]Matched peptides were used to extract the 10 substitutions leading to the largest changes in affinity. These transformations were normalized to reflect affinity increases and sorted according to decreasing effect size; the standard error of the mean (SEM) is indicated as measure of statistical uncertainty. Removal of proline residues increases the binding strength to MHC II molecules, as does the removal of charged residues. On the contrary, inclusion of cysteine residues increases the binding affinity to the receptor.

Overall, for 88 of 190 transformations (44%) we observed a significant change in binding affinity. The strongest effect was achieved by a replacement of proline by cysteine, leading to a gain of 0.715 log units, which corresponds to 5 times stronger binding. The standard error of the mean (SEM) observed over 10 examples for this transformation was 0.291 log units, which is much smaller than the average effect size. This indicates that replacement of proline by cysteine indeed leads to an increase in binding affinity largely independent of the peptide position where the transition occurs.

Several additional substitutions of proline among the transformations with the strongest effects on binding affinity indicates that this residue is in general detrimental to MHC II binding. Replacement by smaller residues is favored, and especially the inclusion of cysteine residues leads to major gains in binding affinity. Additionally, replacement of charged residues is associated with gains in binding affinity. Within the top 10 transformations we found the substitution of aspartate and lysine by cysteine, both of which led to an affinity gain of approximately 0.6 log units, corresponding to a factor of 4 on a linear scale. The aspartate to cysteine transformation is associated with a particularly small SEM of 0.147, indicating particularly conserved effects over the whole binding-site region.

In addition to 88 transformations with significant effects on the binding strength, we characterized 102 transformations associated with only minor changes in MHC II binding. Here we observed the absence of amino acids described to be associated with particularly weak or strong binding. Therefore, the frequency of cysteine residues and charged residues within these transformations was reduced or those residues were completely missing. We found several transformations involving

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

small amino acids that appear to be readily interchangeable within MHC II binding peptides (see Table 2). This behavior is illustrated by the substitution of valine by isoleucine, which is associated with a mean affinity difference smaller than 0.001 log units as well as a small SEM of 0.059 log units over 11 peptide pairs. This indicates that the subtle transformation involving an addition of a methylene group in the side chain does not alter the binding constant independent of the position of the exchanged amino acid. On the contrary, some substitutions involving major chemical changes do not affect the MHC II binding affinities significantly. This includes, for example, the substitution of a small glycine residue by an aromatic histidine residue, which points to the minor importance of residue size in MHC II binding. On average, this transformation is associated with an affinity gain of less than 0.01 log units. The larger SEM of 0.154 log units over 10 pairs indicates some dependence on the position of the transformation in this case.

Table 2. Transformations with Little Effect on Binding Affinity[a]

transformation	pairs	mean affinity difference [log units]	SEM [log units]
V → I	11	<0.001	0.059
T → H	10	0.001	0.129
V → N	10	0.002	0.107
I → W	11	0.003	0.114
L → F	11	0.003	0.173
V → W	11	0.003	0.128
G → T	10	0.008	0.124
A → V	19	0.008	0.073
G → H	10	0.009	0.154
N → S	10	0.010	0.133

[a]Matched peptides were used to search for amino acid substitutions with the smallest changes in experimentally measured binding affinity to MHC II. The top 10 transformations were sorted according to increasing effect on the binding affinity. Statistical uncertainty is shown by the SEM. The complete absence of proline residues and charged amino acids indicates their major impact on the observed binding affinities to MHC II. The smallest effect is observed for the replacement of a valine by an isoleucine, corresponding to the addition of a methylene group.

On the basis of the same matched peptide analysis, we aimed to identify the most favored and unfavored residues in MHC II binding peptides. Since the peptide transformations have been arranged to reflect gains in binding affinity, we counted the occurrence of all 20 amino acids on the left side of the transformation (N_{left}, smaller activity) and on the right side (N_{right}, higher activity) among all 88 pairs showing a significant change in binding affinity. The differences in these occurrences ($N_{right} - N_{left}$), which are given in Figure 2A, enable qualitative identification of favorable and unfavorable amino acids. We found six amino acids to be mainly disfavored in MHC II binding proteins: proline, all four charged amino acids (aspartate, glutamate, lysine, and arginine), and the polar amino acid glutamine. Glutamine shows a smaller negative effect (a total of five pairs with decreased affinity) than the other five amino acids, all of which exhibit very similar disruptive effects (a total of 14 or 15 pairs with decreased affinity). On the other end of the amino acid ranking, tyrosine was found to enhance the MHC II binding affinity in a total of nine significant peptide pairs, followed by cysteine, which was identified as a favorable replacement in a total of eight cases. Mostly small amino acids follow in the ranking, including alanine, methionine, asparagine, and serine. The difference in size might also explain the marked difference observed in comparisons of peptides containing asparagine versus glutamine. The smaller asparagine appears to be favorable for MHC II binding (+6 pairs), whereas glutamine is unfavorable (−5 pairs).

Quantitative Effects of Amino Acid Exchanges. A similar analysis can be performed on the basis of the average effect size rather than the occurrence of amino acids on each side of the transformation. Here, all of the transformations, including those with insignificant effects on the binding affinity, were analyzed to yield a quantitative ranking of amino acid contributions to MHC II binding affinities (see Figure 2B). Consistent with the other presented analyses, proline is associated with a major decrease in MHC II binding affinity representing the strongest effect observed within the data set. On average, 0.474 log units can be gained by replacement of a proline with any other natural amino acid. A replacement of either charged amino acid leads to a gain of between 0.31 and 0.36 log units, thus halving the binding affinity. On the other end of the spectrum, the introduction of cysteine and tyrosine residues is favored and leads to a gain in affinity by 0.26 to 0.28 log unit. Several mainly small and hydrophobic residues are

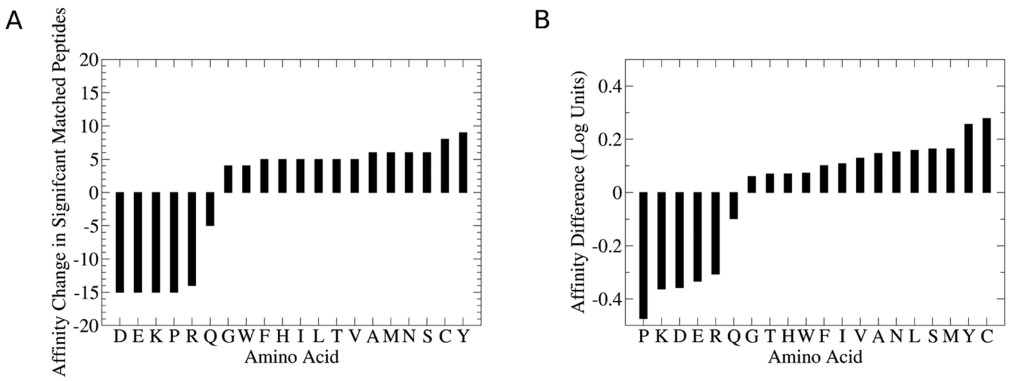


Figure 2. Amino acids favored and disfavored in MHC II binding. On the basis of matched peptides, we identified amino acids frequently associated with a loss in binding affinity. (A) Differences in the number of significant matched peptides leading to a gain versus a loss of affinity in MHC II binding. (B) Absolute average differences in bioactivity when exchanging an amino acid with any other natural amino acid. Proline residues as well as charged amino acids are strongly disfavored in MHC II binding. On the other end of the spectrum, cysteine and tyrosine residues enhance peptide–MHC II interactions.

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

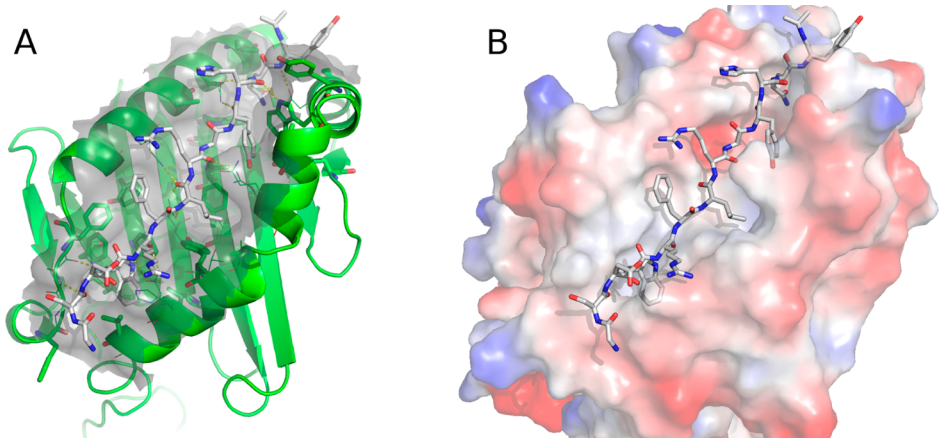


Figure 3. Structural interpretation of MHC II−peptide interactions. On the basis of the cocrystal structure of MHC II and a high-affinity peptide ligand,[49] we analyzed molecular interactions. (A) The peptide ligand (sticks in elemental colors with carbon in gray) is bound to a broad surface groove of the MHC II (green cartoon and lines with semitransparent surface). Hydrogen bonding (yellow dots) is mainly observed via the peptide backbone. (B) An electrostatic map of the MHC II binding site is shown (blue, positively charged; white, neutral; red, negatively charged). The peptide is bound to a predominantly neutral region of the binding site. Additionally, most of the amino acid side chains are bound to solvent-exposed regions on the surface. These binding-site properties explain the promiscuity of MHC II, which is crucial for its immunological function.

slightly favored and show affinity increases of around 0.1 log units on average. Hydrophobic residues have been described to drive association of protein−protein interactions in general.[52] As shown by statistical analysis of crystal structure data, aliphatic amino acids and tyrosine residues predominantly form cores of protein−protein interaction areas.[53]

To allow a comparison with a conventional peptide analysis method, we divided our data set of 198 peptides into two halves according to higher and lower binding affinity to MHC II and analyzed enriched amino acids among both sets. We report residues with an enrichment factor larger than 1.5 in decreasing order of their enrichment. We found that hydrophobic cysteine, methionine, isoleucine, valine, and phenylalanine are enriched among high-activity binders. Low-affinity binders on the other hand show a disproportionately high content of aspartate, glutamate, proline, arginine, and tryptophan. The observed trends for favored and disfavored residues are similar to those from the pairwise comparisons conducted for matched peptide analysis. Nevertheless, the results from matched peptide analysis provide deeper insights since experimental uncertainty can be handled directly and no classification into affinity classes via an arbitrary cutoff is required.

Correlation to Amino Acid Descriptors. As we observed clear correlations between affinity differences and chemical properties of amino acids, we tested the performance of classical QSAR amino acid descriptors in reproducing those. Therefore, we calculated property differences for all of the transformations in five different dimensions based on z-scales (three descriptors) as well as the ISA/ECI scheme (two descriptors). Then we correlated the property differences to the experimentally measured affinity differences identified via matched peptide analysis and analyzed them using Pearson's linear correlation coefficient and Spearman's rank correlation coefficient.

We observed weak correlations between individual descriptor differences of substituted amino acids and the associated bioactivity changes of 190 matched peptide pairs (see Figure S1 for correlation plots). The most pronounced correlation was identified for the z_2 axis representing residue size. Here we found an inverse correlation ($r = -0.33$, $\rho = -0.32$), indicating that a reduction in residue size is associated with an increase in binding affinity. This index is closely followed by the ECI, which designates both positively and negatively charged residues with large values and therefore indicates polarity ($r = -0.31$, $\rho = -0.31$). The observed inverse correlation indicates that a reduction in polarity is favored in MHC II binding. The third axis contributing to binding affinity is the z_1 descriptor that reflects hydrophobicity. We again observed an inverse correlation ($r = -0.29$, $\rho = -0.27$) and conclude in agreement with the ECI results that a reduction in polarity favors MHC II binding. The other two descriptors (z_3 and ISA) show correlation coefficients smaller than or equal to $r = 0.1$. Thus, electronic properties and residue surface area were found to be less important in MHC II recognition. The lack of correlation to the surface area appears surprising since a reduction in residue size was identified as favorable. We attribute this seeming contradiction to the inclusion of solvation factors in the ISA calculation, which leads to the lowest ISA values for asparagine and aspartate even though these amino acids have a larger molecular weight than, for example, glycine.

As these analyses included proline residues in the data set, we wondered whether removal of this residue with a different backbone, and thus a refined depiction of the side-chain properties, would increase the observed correlations. Therefore, we repeated the correlation analysis covering only the 171 pairs not affecting proline. We found that the magnitudes of all of the correlation coefficients consistently increased, pointing toward the special status of this amino acid. The correlation between the affinity change and the difference in ECI was strengthened from $r = -0.31$ to $r = -0.49$ upon removal of the uncharged but still disfavored proline residue. Similarly, the correlation of z_1 differences to the binding affinity changes jumped from $r = -0.29$ to $r = 0.40$ when proline was included. This points to the special importance of peptide backbone hydrogen bonding in MHC II interactions.

Structural Analysis of the MHC II Receptor−Peptide Complex. To interpret these SAR trends in a structural context, we investigated a cocrystal structure of the MHC II receptor with a cognate peptide of length 15 amino acids.[49] The peptide is bound in an extended conformation and is tightly bound to the receptor via 14 hydrogen bonds of the backbone carbonyls and amides (see Figure 3A). This large number of

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

interactions indicates the importance of backbone hydrogen bonding in receptor binding and therefore explains the observed trend that removal of prolines, which lack the backbone NH capable of hydrogen-bond formation, leads to stronger binding. In contrast, amino acid side chains are involved in fewer interactions with the receptor then the peptide backbone. Most of the side chains of the bound peptide are bound to shallow pockets rather than to pronounced cavities and show limited contact area. This explains the lack of particular amino acids being favored as a result of the absence of polar interactions with the receptor. This is further underlined by the absence of charged regions in the binding site (see Figure 3B). The binding-site region for side chains is mostly a flat solvent-exposed surface patch that therefore can bind several amino acids. The hydrophobicity of a large part of the binding site additionally leads to a preference for apolar amino acids due to energy gains by desolvation.[54] Since such hydrophobic and solvent-exposed interfaces are expected to be multispecific,[55] the MHC II may perform its key function in immunology to recognize and present diverse peptides from pathogens.[56]

Promiscuity in MHC II Receptor−Peptide Interactions. This promiscuous binding of peptides is also reflected in our data set, where the range of measured binding affinities is approximately 3 log units from the most tightly bound peptide to the weakest binder (see Figure S2 for affinity distributions). Many amino acid exchanges show negligible effects on the observed binding constant, in agreement with promiscuous binding. With the assumption of a conserved position of the peptide termini, which might be provided via the central binding register of approximately nine residues in MHC II,[57] the specificity for each of the respective peptide positions can be quantified as information entropy from the distribution of binding affinities for single-point mutants. An entropy of 0 depicts the highest specificity and thus affinity for only a single amino acid, while an entropy of 1 corresponds to completely unspecific binding with constant affinities for all binding partners.[51]

We found all of the peptide positions to be predominantly unspecific, with information entropies ranging from 0.83 for position 11 to 0.97 for position 10. At the most specific peptide position 11, seven amino acids (D, E, F, I, K, R, and Y) show binding affinities differing from the most favorable amino acid, cysteine, by at least 1 log unit. On the contrary, for the almost completely unspecific position 10 all of the binding affinities lie within 0.8 log units, with tyrosine as the most favored residue. This latter situation reflects the typical situation in MHC II peptide recognition, where seven of the 10 investigated pockets show an information entropy larger than 0.9. On the other hand, more specific positions coincide with classical MHC II specificity sites 4, 6, and 9,[58] which show information entropies of 0.90, 0.88, and 0.91, respectively. The overall promiscuity is also reflected by a comparison of activity rankings in respective binding pockets, an approach similar to a matched series. We found only weak correlations between subpocket profiles, with Spearman rank correlation coefficients between −0.32 and +0.57. In fact, 32% of the affinity profiles appear anticorrelated the inherent experimental error is neglected, which we expect to limit the applicability of this approach given the narrow affinity ranges within the data set.

The overall promiscuity shows the applicability of the matched pair approach, which treats all amino acid exchanges equally and does not include the specific position. This inherent limitation of the approach could be overcome by the use of context-specific matched peptides in analogy to context-specific matched pairs, where the chemical environment of the transformation is included.[59] In the standard implementation of matched pairs and peptides, transformations showing context-specific effects are attributed with higher standard errors. This can be seen in our data set for the proline to arginine transformation, which shows a weak average gain in binding affinity of 0.158 log units in 10 pairs. The standard error of the mean for this transformation is 0.400 log units, showing that the same transformation is favorable for binding in four cases and unfavorable in six. Here the position of the transformation and therefore the chemical context are crucial for the observed effect, although MHC II is overall highly unspecific. The strongest effect for the proline to arginine transformation is observed at position 4, where the binding affinity is increased by more than 3 orders of magnitude. This represents the most dramatic effect of a single substitution and occurs at position 4, where proline appears to be especially disfavored compared with all other amino acids (see Figure 4).

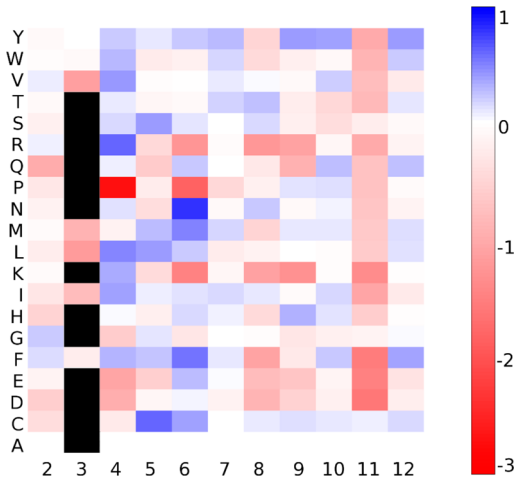


Figure 4. Context specificity of matched peptides: Negative decadic logarithms of MHC-II binding affinity ratios vs the template peptide AAYAAAAAAAAAA are shown as a heat map, where blue boxes represent a gain in affinity, white boxes indicate no change, and red boxes indicate affinity losses. Black boxes represent missing data at position 3, where only seven amino acids were tested experimentally. Substitutions at positions 4 and 6 lead to major differences in binding affinity, and proline is particularly disfavored at these positions. Context-specific analysis of affinity data rather than averaging over all peptide positions is expected to capture these effects in more detail.

Position 6 is the second peptide position where the presence of proline is strongly disfavored. Here replacement by asparagine yields a gain in affinity of 2.6 log units. In terms of context-specific matched peptides, positions 4 and 6 show the strongest affinity differences in the analyzed data set. We expect that context-specific matched pairs are especially required in order to reliably predict changes in binding affinities to specific receptors. Physico-chemical properties like solubility, on the contrary, are expected to be more independent of the chemical surroundings and thus easier to capture using standard matched peptides.

Future Potential of Matched Peptide Analysis. In summary, we have introduced a new quantitative modeling tool at the interface of cheminformatics and bioinformatics: matched peptides. By means of this extension of matched

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

molecular pair analysis, protein, peptide, and nucleic acid sequence-related properties such as bioactivity, solubility, aggregation, pI, bioavailability, stability, and expression yield may be predicted. The inclusion of nonstandard residues and modified amino acids or nucleotides and cross-links is straightforward given sufficient training data. The coverage of cyclic sequences, which are of high interest because of the additional stability of cyclic peptides and proteins,[60] requires the use of special software tailored for alignment of cyclic sequences because of undefined terminal residues.[61] When matched peptide analysis is extended beyond the analysis of single-point mutations, cooperative effects in structure−activity relationships may be identified using nonadditivity cycles, as recently demonstrated for small molecules via matched molecular pair analysis.[62]

In comparison with the wealth of bioactivity data readily available for small-molecule ligands (e.g., via ChEMBL[27]), the data basis for peptide binding data is much sparser. To date there exist only a few databases listing peptide affinities to general targets (e.g., JenPep[63]) that could be used for matched peptide analysis using public domain data. Most of the open data are centered around immunology and MHC binding (e.g., IEDB[64]), hindering broad application of the technique in the academic environment. As a result of recent biotechnological advances in the development of protein, peptide, and nucleic acid microarrays, a plethora of qualitative and quantitative binding data are available.[65] Similar data can be obtained, e.g., from proteomics techniques,[66] protein-fragment complementation assays,[67] or phage display.[68] Additionally, synthetic access to peptides can be automated using solid-phase synthesis.[69] The decision of which peptide to synthesize next may be supported by using the presented matched peptide approach, which allows existing bioactivity data to be captured in an intuitive way and provides a qualitative and quantitative ranking of favored residues. Therefore, the matched peptide strategy shows synergy and complementarity with classical QSAR techniques.

CONCLUSION

We have presented matched peptides as an extension of standard matched molecular pair analysis that pushes the technology to the interface of cheminformatics and bioinformatics. Matched peptides correspond to single-point mutants, which are easily identified via pairwise sequence alignments, simplifying the data analysis. Differences in molecular properties may be identified via the matched peptide strategy and can subsequently be applied to identify SAR trends and predict properties of new sequences. Herein we have presented a statistical analysis of MHC II peptide binding data and discussed observed trends with respect to classical QSAR as well as structural data of the complex. We expect our presented methodology, which is readily applicable to peptides, proteins, and nucleic acids, to be of high relevance for the rational design of novel biopharmaceuticals.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00476.

Correlation plots of amino acid descriptors and activity differences extracted via matched peptide analysis as well as affinity distributions for respective peptide positions (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: julian.fuchs@uibk.ac.at.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the Austrian Science Fund (FWF) via Grants P23051 “Targeting Influenza Neuraminidase” and P26997 “Influence of Protein Folding on Allergenicity and Immunogenicity”. J.E.F. acknowledges financial support from the University of Innsbruck via the program “Nachwuchsförderung” as well as stimulating discussions with Christian Kramer and Anna Sophia Kamenik.

ABBREVIATIONS

MHC, major histocompatibility complex; QSAR, quantitative structure−activity relationship; QSPR, quantitative structure−property relationship; SAR, structure−activity relationship

REFERENCES

(1) Schellekens, H. Bioequivalence and the Immunogenicity of Biopharmaceuticals. *Nat. Rev. Drug Discovery* **2002**, *1*, 457−462.
(2) Miller, K. L.; Lanthier, M. Innovation in Biologic new Molecular Entities: 1986−2014. *Nat. Rev. Drug Discovery* **2015**, *14*, 83.
(3) Mullard, A. 2014 FDA Drug Approvals. *Nat. Rev. Drug Discovery* **2015**, *14*, 77−81.
(4) Crommelin, D. J.; Storm, G.; Verrijk, R.; de Leede, L.; Jiskoot, W.; Hennink, W. E. Shifting Paradigms: Biopharmaceuticals versus Low Molecular Weight Drugs. *Int. J. Pharm.* **2003**, *266*, 3−16.
(5) Brinckerhoff, C. C.; Schorr, K. Have the Biosimilar Floodgates been Opened in the United States? *Nat. Rev. Drug Discovery* **2015**, *14*, 303−304.
(6) Berkowitz, S. A.; Engen, J. R.; Mazzeo, J. R.; Jones, G. B. Analytical Tools for Characterizing Biopharmaceuticals and the Implications for Biosimilars. *Nat. Rev. Drug Discovery* **2012**, *11*, 527−540.
(7) Muller, R. H.; Keck, C. M. Challenges and Solutions for the Delivery of Biotech Drugs − a Review of Drug Nanocrystal Technology and Lipid Nanoparticles. *J. Biotechnol.* **2004**, *113*, 151−170.
(8) Mitragotri, S.; Burke, P. A.; Langer, R. Overcoming the Challenges in Administering Biopharmaceuticals: Formulation and Delivery Strategies. *Nat. Rev. Drug Discovery* **2014**, *13*, 655−672.
(9) Baldi, L.; Hacker, D. L.; Adam, M.; Wurm, F. M. Recombinant Protein Production by Large-Scale Transient Gene Expression in Mammalian Cells: State of the Art and Future Perspectives. *Biotechnol. Lett.* **2007**, *29*, 677−684.
(10) Shire, S. J. Formulation and Manufacturability of Biologics. *Curr. Opin. Biotechnol.* **2009**, *20*, 708−714.
(11) Mao, H.; Graziano, J. J.; Chase, T. M. A.; Bentley, C. A.; Bazirgan, O. A.; Reddy, N. P.; Song, B. D.; Smider, V. V. Spatially Addressed Combinatorial Protein Libraries for Recombinant Antibody Discovery and Optimization. *Nat. Biotechnol.* **2010**, *28*, 1195−1204.
(12) Lippow, S. M.; Wittrup, K. D.; Tidor, B. Computational Design of Antibody-Affinity Improvement beyond *in vivo* Maturation. *Nat. Biotechnol.* **2007**, *25*, 1171−1176.
(13) Andersson, I. E.; Andersson, C. D.; Batsalova, T.; Dzhambazov, B.; Holmdahl, R.; Kihlberg, J.; Linusson, A. Design of Glycopeptides used to Investigate Class II MHC Binding and T-cell Responses associated with Autoimmune Arthritis. *PLoS One* **2011**, *6*, e17881.
(14) Maccari, G.; Di Luca, M.; Nifosi, R.; Cardarelli, F.; Signore, G.; Boccardi, C.; Bifone, A. Antimicrobial Peptides Design by Evolu-

tionary Multiobjective Optimization. *PLoS Comput. Biol.* **2013**, *9*, e1003212.
(15) Kovalainen, M.; Mönkäre, J.; Riikonen, J.; Pesonen, U.; Vlasova, M.; Salonen, J.; Lehto, V.-P.; Järvinen, K.; Herzig, K.-H. Novel Delivery Systems for Improving the Clinical Use of Peptides. *Pharmacol. Rev.* **2015**, *67*, 541−561.
(16) Wu, S.; Qi, W.; Su, R.; Li, T.; Lu, D.; He, Z. CoMFA and CoMSIA Analysis of ACE-Inhibitory, Antimicrobial and Bitter-Tasting Peptides. *Eur. J. Med. Chem.* **2014**, *84*, 100−106.
(17) Borkar, M. R.; Pissurlenkar, R. R. S.; Coutinho, E. C. HomoSAR: Bridging Comparative Protein Modeling with Quantitative Structural Activity Relationship to Design new Peptides. *J. Comput. Chem.* **2013**, *34*, 2635−2646.
(18) Zaliani, A.; Gancia, E. MS-WHIM Scores for Amino Acids: A new 3D-Description for Peptide QSAR and QSPR Studies. *J. Chem. Inf. Model.* **1999**, *39*, 525−533.
(19) Sneath, P. H. Relations between Chemical Structure and Biological Activity in Peptides. *J. Theor. Biol.* **1966**, *12*, 157.
(20) Hellberg, S.; Sjöström, M.; Skagerberg, B.; Wold, S. Peptide Quantitative Structure-Activity Relationships, a Multivariate Approach. *J. Med. Chem.* **1987**, *30*, 1126−1135.
(21) Collantes, E. R.; Dunn, W. J., III. Amino Acid Side Chain Descriptors for Quantitative Structure-Activity Relationship Studies of Peptide Analogues. *J. Med. Chem.* **1995**, *38*, 2705−2713.
(22) Dayhoff, M. O.; Schwartz, R.; Orcutt, B. C. A Model of Evolutionary Change in Proteins. In *Atlas of Protein Sequence and Structure*; Natural Biomedical Research Foundation: Washington, DC, 1978; pp 345−358.
(23) Henikoff, S.; Henikoff, J. G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 10915−10919.
(24) Pissurlenkar, R. R. S.; Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. Encoding Type and Position in Peptide QSAR: Application to Peptides Binding to Class I MHC Molecule HLA-A*0201. *QSAR Comb. Sci.* **2007**, *26*, 189−203.
(25) Kenny, P. W.; Sadowski, J. Structure Modification in Chemical Databases. In *Cheminformatics in Drug Discovery*; Wiley-VCH: Weinheim, Germany, 2005; pp 271−285.
(26) Griffen, E.; Leach, A. G.; Robb, G. R.; Warner, D. J. Matched Molecular Pairs as a Medicinal Chemistry Tool. *J. Med. Chem.* **2011**, *54*, 7739−7750.
(27) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083−D1090.
(28) Brown, D. G.; Gagnon, M. M.; Boström, J. Understanding our Love Affair with p-Chlorophenyl: Present Day Implications from Historical Biases of Reagent Selection. *J. Med. Chem.* **2015**, *58*, 2390−2405.
(29) O'Boyle, N. M.; Boström, J.; Sayle, R. A.; Gill, A. Using Matched Molecular Series as a Predictive Tool to Optimize Biological Activity. *J. Med. Chem.* **2014**, *57*, 2704−2713.
(30) Cumming, J. G.; Davis, A. M.; Muresan, S.; Haeberlein, M.; Chen, H. Chemical Predictive Modelling to Improve Compound Quality. *Nat. Rev. Drug Discovery* **2013**, *12*, 948−962.
(31) Weber, J.; Achenbach, J.; Moser, D.; Proschak, E. VAMMPIRE: A Matched Molecular Pairs Database for Structure-Based Drug Design and Optimization. *J. Med. Chem.* **2013**, *56*, 5203−5207.
(32) Posy, S. L.; Claus, B. L.; Pokross, M. E.; Johnson, S. R. 3D Matched Pairs: Integrating Ligand- and Structure-Based Knowledge for Ligand Design and Receptor Annotation. *J. Chem. Inf. Model.* **2013**, *53*, 1576−1588.
(33) Rudolph, M. G.; Stanfield, R. L.; Wilson, I. A. How TCRs Bind MHCs, Peptides, and Coreceptors. *Annu. Rev. Immunol.* **2006**, *24*, 419−466.
(34) Birnbaum, M. E.; Mendoza, J. L.; Sethi, D. K.; Dong, S.; Glanville, J.; Dobbins, J.; Özkan, E.; Davis, M. M.; Wucherpfennig, K. W.; Garcia, K. C. Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell* **2014**, *157*, 1073−1087.
(35) Stern, L. J.; Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Urban, R. G.; Strominger, J. L.; Wiley, D. C. Crystal Structure of the Human Class II MHC Protein HLA-DR1 Complexed with an Influenza Virus Peptide. *Nature* **1994**, *368*, 215−221.
(36) Lafuente, E. M.; Reche, P. A. Prediction of MHC-Peptide Binding: A Systematic and Comprehensive Overview. *Curr. Pharm. Des.* **2009**, *15*, 3209−3220.
(37) Zhang, G. L.; Khan, A. M.; Srinivasan, K. N.; August, J. T.; Brusic, V. MULTIPRED: A Computational System for Prediction of Promiscuous HLA Binding Peptides. *Nucleic Acids Res.* **2005**, *33*, W172−W179.
(38) Wang, P.; Sidney, J.; Dow, C.; Mothe, B.; Sette, A.; Peters, B. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Comput. Biol.* **2008**, *4*, e1000048.
(39) Koch, C. P.; Perna, A. M.; Pillong, M.; Todoroff, N. K.; Wrede, P.; Folkers, G.; Hiss, J. A.; Schneider, G. Scrutinizing MHC-I Binding Peptides and their Limits of Variation. *PLoS Comput. Biol.* **2013**, *9*, e1003088.
(40) Wang, Y.; Zhou, P.; Lin, Y.; Shu, M.; Hu, Y.; Xia, Q.; Lin, Z. Quantitative Prediction of Class I MHC/Epitope Binding Affinity using QSAR Modeling Derived from Amino Acid Structural Information. *Comb. Chem. High Throughput Screening* **2015**, *18*, 75−82.
(41) Knapp, B.; Dunbar, J.; Deane, C. M. Large Scale Characterization of the LC13 TCR and HLA-B8 Structural Landscape in Reaction to 172 Altered Peptide Ligands: A Molecular Dynamics Simulation Study. *PLoS Comput. Biol.* **2014**, *10*, e1003748.
(42) Hussain, J.; Rea, C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J. Chem. Inf. Model.* **2010**, *50*, 339−348.
(43) Leach, A. G.; Jones, H. D.; Cosgrove, D. A.; Kenny, P. W.; Ruston, L.; MacFaul, P.; Wood, M.; Colclough, N.; Law, L. Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties: A Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure. *J. Med. Chem.* **2006**, *49*, 6672−6682.
(44) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization*; Springer: Berlin, 2007.
(45) Marshall, K. W.; Wilson, K. J.; Liang, J.; Woods, A.; Zaller, D.; Rothbard, J. B. Prediction of Peptide Affinity to HLA DRB1*0401. *J. Immunol.* **1995**, *154*, 5927−5933.
(46) Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of Mixed IC_{50} Data − A Statistical Analysis. *PLoS One* **2013**, *8*, e61007.
(47) Kramer, C.; Fuchs, J. E.; Whitebread, S.; Gedeck, P.; Liedl, K. R. Matched Molecular Pair Analysis: Significance and the Impact of Experimental Uncertainty. *J. Med. Chem.* **2014**, *57*, 3786−3802.
(48) R Core Team. *R: A Language and Environment for Statistical Computing*, version 3.2.0; The R Foundation: Vienna, Austria, 2015; http://www.R-project.org/.
(49) Murthy, V. L.; Stern, L. J. The Class II MHC Protein HLA-DR1 in Complex with an Endogenous Peptide: Implications for the Structural Basis of Specificity of Peptide Binding. *Structure* **1997**, *5*, 1385−1396.
(50) *The PyMOL Molecular Graphics System*, version 1.6.0.0; Schrödinger, LLC: New York, 2013; http://www.schrodinger.com/pymol/.
(51) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Margreiter, M. A.; Spitzer, G. M.; Wallnoefer, H. G.; Liedl, K. R. Cleavage Entropy as Quantitative Measure of Protease Specificity. *PLoS Comput. Biol.* **2013**, *9*, e1003007.
(52) Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Studies of Protein-Protein Interfaces: A Statistical Analysis of the Hydrophobic Effect. *Protein Sci.* **1997**, *6*, 53−64.

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323

(53) Bickerton, G. R.; Higueruelo, A. P.; Blundell, T. L. Comprehensive, Atomic-Level Characterization of Structurally Characterized Protein-Protein Interactions: The PICCOLO Database. *BMC Bioinf.* **2011**, *12*, 313.
(54) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(55) Chang, C. A.; McLaughlin, W. A.; Baron, R.; Wang, W.; McCammon, J. A. Entropic Contributions and the Influence of the Hydrophobic Environment in Promiscuous Protein-Protein Association. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7456−7461.
(56) Felix, N. J.; Allen, P. M. Specificity of T-Cell Alloreacitivity. *Nat. Rev. Immunol.* **2007**, *7*, 942−953.
(57) Mohan, J. F.; Unanue, E. R. Unconventional Recognition of Peptides by T Cells and the Implications for Autoimmunity. *Nat. Rev. Immunol.* **2012**, *12*, 721−728.
(58) Painter, C. A.; Stern, L. J. Structural Insights Into HLA-DM Mediated MHC II Peptide Exchange. *Curr. Top. Biochem. Res.* **2011**, *13* (2), 39−55.
(59) Papadatos, G.; Alkarouri, M.; Gillet, V. J.; Willett, P.; Kadirkamanathan, V.; Luscombe, C. N.; Bravi, G.; Richmond, N. J.; Pickett, S. D.; Hussain, J.; Pritchard, J. M.; Cooper, A. W. J.; Macdonald, S. J. F. Lead Optimization using Matched Molecular Pairs: Inclusion of Contextual Information for Enhanced Prediction of hERG Inhibition, Solubility, and Lipophilicity. *J. Chem. Inf. Model.* **2010**, *50*, 1872−1886.
(60) Wang, C. K. L.; Kaas, Q.; Chiche, L.; Craik, D. J. CyBase: A Database of Cyclic Protein Sequences and Structures, with Applications in Protein Discovery and Engineering. *Nucleic Acids Res.* **2008**, *36* (Suppl. 1), D206−D210.
(61) Fernandes, F.; Pereira, L.; Freitas, A. T. CSA: An Efficient Algorithm to Improve Circular DNA Multiple Alignment. *BMC Bioinf.* **2009**, *10*, 230.
(62) Kramer, C.; Fuchs, J. E.; Liedl, K. R. Strong Nonadditivity as a Key Structure-Activity Relationship Feature: Distinguishing Structural Changes from Assay Artifacts. *J. Chem. Inf. Model.* **2015**, *55*, 483−494.
(63) McSparron, H.; Blythe, M. J.; Zygouri, C.; Doytchinova, I. A.; Flower, D. R. JenPep: A Novel Computational Information Resource for Immunobiology and Vaccinology. *J. Chem. Inf. Model.* **2003**, *43*, 1276−1287.
(64) Peters, B.; Sidney, J.; Bourne, P.; Bui, H. H.; Buus, S.; Doh, G.; Fleri, W.; Kronenberg, M.; Kubo, R.; Lund, O.; Nemazee, D.; Ponomarenko, J. V.; Sathiamurthy, M.; Schoenberger, S.; Stewart, S.; Surko, P.; Way, S.; Wilson, S.; Sette, A. The Immune Epitope Database and Analysis Resource: From Vision to Blueprint. *PLoS Biol.* **2005**, *3*, e91.
(65) Houseman, B. T.; Huh, J. T.; Kron, S. J.; Mrksich, M. Peptide Chips for the Quantitative Evaluation of Protein Kinase Activity. *Nat. Biotechnol.* **2002**, *20*, 270−274.
(66) Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of Proteins and Proteomes by Mass Spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437−473.
(67) Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E. Universal Strategies in Research and Drug Discovery based on Protein-Fragment Complementation Assays. *Nat. Rev. Drug Discovery* **2007**, *6*, 569−582.
(68) Scott, J. K.; Smith, G. P. Searching for Peptide Ligands with an Epitope Library. *Science* **1990**, *249*, 386−390.
(69) Merrifield, R. B. Solid Phase Peptide Synthesis. 1. Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149−2154.

DOI: 10.1021/acs.jcim.5b00476
J. Chem. Inf. Model. 2015, 55, 2315−2323