Theoretical aspects of peptide imprinting: screening of MIP (virtual) binding sites for their interactions with amino acids, di- and tripeptides

Julie Settipani, Kal Karim, Alienor Chauvin, Si Mohamed Ibnou-Ali, Florian Paille-Barrere, Evgeny Mirkes, Alexander Gorban, Lee Larcombe, Michael J. Whitcombe, Todd Cowen & Sergey A. Piletsky

Published online: 10 May 2018.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Check for updates

Theoretical aspects of peptide imprinting: screening of MIP (virtual) binding sites for their interactions with amino acids, di- and tripeptides

Julie Settipani[a], Kal Karim[a], Alienor Chauvin[a], Si Mohamed Ibnou-Ali[a], Florian Paille-Barrere[a], Evgeny Mirkes[b], Alexander Gorban[b], Lee Larcombe[c], Michael J. Whitcombe[a], Todd Cowen[a] and Sergey A. Piletsky[a]

[a]Leicester Biotechnology Group, University of Leicester, Leicester, UK; [b]Department of Mathematics, University of Leicester, Leicester, UK; [c]Applied Exomics, Stevenage Bioscience Catalyst, Stevenage, UK

ABSTRACT

Molecular modelling and computational approaches were used to design (virtual) molecularly imprinted binding sited for 170 amino acids, dipeptides and tripeptides. Analysis of the binding energy of ligands to their corresponding virtual binding sites revealed a direct correlation between size of the ligand and its binding affinity. Only tripeptides were capable of forming binding sites in molecularly imprinted polymers (MIPs) that are capable, in theory, of binding the corresponding targets at micromolar concentrations. No appreciable specificity was demonstrated in binding of virtual binding sites and corresponding templates. It is possible to conclude that although tripeptide sequences are sufficiently long to form MIPs with relatively high affinity, the sequence of peptide epitopes should be substantially longer that three amino acid residues to ensure specificity of imprinted sites. This consideration will be useful for the design of highly efficient MIPs for proteins.

1. Introduction

The selective and specific recognition of polypeptide and protein targets is a fundamental challenge in bio-organic chemistry. Materials with selective protein binding properties, such as molecularly imprinted polymers (MIPs), have a host of applications, including diagnostic assays and sensors, process control, proteomics, downstream processing, cell labeling, bio-security and medicine. The specific binding properties of MIPs arise due to the formation of a cross-linked network in the presence of a molecular template, thereby creating a binding domain within the polymer that is complementary to the target, both in terms of a shape and chemical functionality. This chemical complementarity comes from the inclusion of functional monomers within the polymerization mixture that have favorable binding interactions with the template during the process of polymer formation. The binding sites (molecular imprints) must be cleared of template by extraction before the polymers can be used for selective recognition of their respective targets. The

---

imprinting of small molecules is aided by their ability to diffuse through a porous cross-linked monolith. Macromolecules, such as proteins would not be expected to diffuse out, or back into such a material. In the case of large templates therefore, locating the recognition sites on the surface of the MIP is necessary to allow rebinding to occur. A number of surface imprinting regimes have been designed to allow macromolecule-imprinted materials to be prepared, including micro-contact printing and bottom-up approaches. An alternative to imprinting whole protein molecules is the use of a characteristic peptide as template (the epitope approach).

The use of terminal sequences of target proteins as a surrogate template for the whole macromolecule was first proposed by Rachkov and Minoura in 2000. [1] They demonstrated that materials capable of binding the nonapeptide oxytocin, under aqueous conditions, could be prepared by imprinting a tetrapeptide with the same three amino acid residues in the N-terminus (Pro-Leu-Gly-NH$_2$). The epitope approach was soon adapted by others to prepare materials capable of recognizing larger proteins, such as cytochrome c and even viruses. The concept of an 'epitope' has been borrowed from immunochemistry. In that context, it is the region of an antigen that is recognized by the variable domain of an antibody and could span adjacent residues on the antigen surface that are non-adjacent in the primary sequence. In imprinting terms however we normally mean an epitope to be a peptide with a single sequence (taken from the primary sequence of the target) that, when used as a template, generates a selective imprint for the said target. While many examples are terminal sequences, for some targets the protein termini may be inaccessible and a loop or other surface displayed region may be selected. A further advantage of using the epitope approach is that it is not necessary to isolate the target protein to use as template, provided that some structural information is known.

While the epitope approach is a very powerful and eminently practical method for preparing imprinted materials capable of binding proteins (including membrane-bound proteins) and viruses, the rules concerning epitope selection are still somewhat unclear. In a preliminary study to investigate the effect of sequence and peptide length on the affinity and specificity of imprints, we have set out to model *in silico* a subset of peptides and to assess the cross-reactivity of the imprints for all members of the ensemble.

2. Results and discussion

2.1. Selection of amino acid, di- and tripeptides target templates

The total number of possible peptides made from the 20 natural amino acids is immense, and so therefore is the total number of MIPs which can be made for these targets. For practical reasons we have therefore decided to restrict the set of amino acids used in this research to only five amino acids (G, S, L, E and K). The rationale behind this selection originates from how frequently these amino acids appear in 35,576 known epitopes (Figure 1). [2] The assumption we have made to justify this choice is that the factors that make a particular amino acid sequence a good epitope for antibody binding should also apply to the production of high quality binding sites in imprinted polymers. [3] Considering all possible di- and tripeptides that can be made from these amino acids, there are 155 discrete structures to consider as templates (5 amino acids, 25 dipeptides and
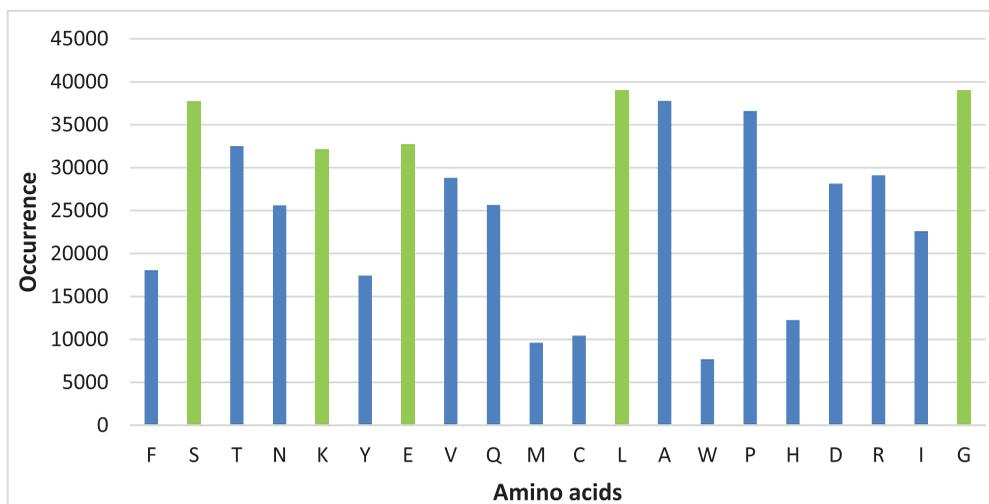
Figure 1. Instances of the appearance of each of the 20 natural amino acids in 35,576 known epitopes (The selected amino acids were S, K, E, L and G) [2].

125 tripeptides). At this stage we did not proceed with design of MIPs for larger peptides for the reason that it will require too much computational time and power to make it practical.

Binding site models for each of the 155 templates were designed by following the protocol outlined in the experimental section. This protocol is a variation of the method for the design of MIPs, developed in our group and validated by a number of research groups worldwide. [4–7] Once all the MIP receptors were modelled, they were probed using the program Surflex-Dock. This uses an empirical scoring function to dock ligands into receptor binding site. Docking is guided by the *protomol*, an idealized representation of the ligand cavity within the binding site. For example, Figure 2 shows the shape of the protomol for the Glu-Leu-Leu template.

As well as the initial set of ligands, comprising 5 amino acids, 25 dipeptides and 125 tripeptides, the remaining 15 amino acids were also added to the screening and docking database to give a final library of 170 ligands. For each ligand, the receptor and corresponding protomol were generated and used for docking of all 170 ligands. After the completion of a docking run, the binding score of each ligand was calculated. The final table of results contains 155 columns (one for each MIP receptor) and 170 rows (one for each ligand). This represents 26,350 empirical scores used in evaluation of molecular interactions between created models of binding sites and corresponding amino acids, di- and tripeptides. It is important to discuss here the meaning of empirical binding scores. Overall they reflect, albeit indirectly, pK_d values of corresponding ligands. The true pK_d values of course cannot be deduced from the modelling data, but empirical binding scores can serve as a good indication of possible affinity of the templates. Thus, the empirical scores smaller than 4.0 (equivalent of $K_d \leq 100 \ \mu M$) are generally considered to be an indicators of weak affinity. [8] Only ligand–receptor pairs with empirical binding score >6.0 (mean pK_d) can potentially be considered as having high affinity. [9] As expected,
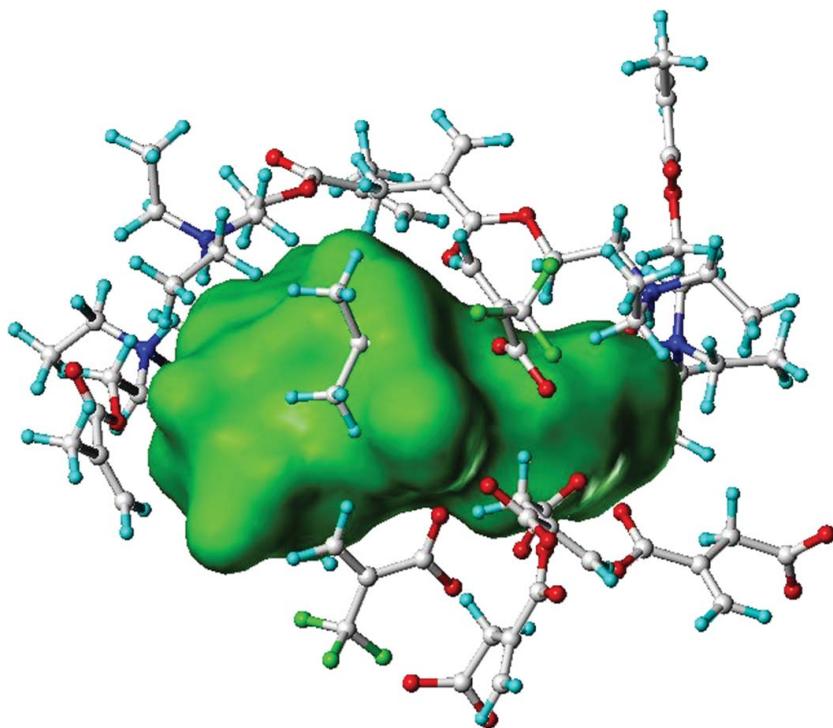
Figure 2. Protomol representation of Glu-Leu-Leu template (green) interacting with its virtual binding site.

all the ligands studied have *some* binding to each model of MIP binding site, as can be seen from examples of histograms presented in Figure 3.

As can be seen from these examples not every peptide template can produce high quality imprinted binding sites. Thus, tripeptide Lys-Leu-Gly produces receptor capable of binding to all template targets with very low binding energy represented by binding score below 3.0. Glu-Glu-Leu however can produce receptors with high affinity to a number of targets with binding score 5.0–6.0.

An interesting observation was noted regarding the effect of the templates' size on its binding strength (Table 1), in that it is possible to conclude that only tripeptides are capable of forming high quality binding sites with micromolar affinities for imprinted polymers. This is largely supported by experimental observations. [10–12]

Tripeptide templates create binding sites with relatively low specificity (Table 2). It is not obvious why any particular template does not always create unique binding site specific for the same molecule. Thus Glu-Glu-Gly as template creates binding site that binds stronger to Leu-Leu-Leu and Lys-Ser-Glu than to Glu-Glu-Gly. Looking closely however at the structure of the binding site composed of monomers such as the charged form of diethylaminoethyl methacrylate (DEAEM) and itaconic acid (Figure 4) it is clear that these monomers will have stronger non-specific bonds with Leu and Lys. From analysis of these data we can conclude that the overall contribution of non-specific binding in the case of MIPs formed by tripeptides is too high to assure specificity of binding sites.
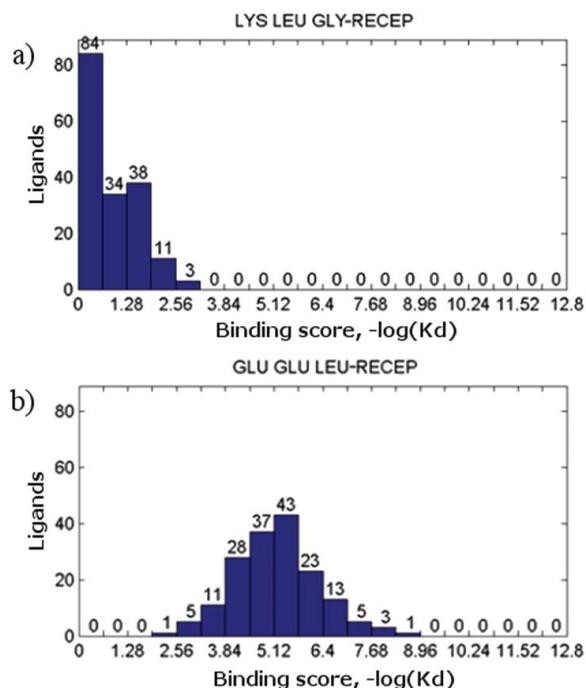
Figure 3. Number of ligands binding to receptor model for tripeptide Lys-Leu-Gly (a) and Glu-Glu-Leu (b) by total scores obtained with Surflex-Dock.

Looking at the Venn diagram (Figure 5) it is possible to see for example that at least 94 targets can bind to binding sites created by three templates: Glu-Glu-Gly, Glu-Gly-Ser and Glu-Leu-Glu. Only two targets bind specifically to Glu-Glu-Gly, one to Glu-Gly-Ser and one to Glu-Leu-Glu.

Low specificity of binding sites is an important issue which may hinder development of MIPs specific for di- and tripeptides. This is not too dissimilar to antibodies, which are difficult to raise against short peptides. [13] With very few exceptions however the recognition of short peptides by MIPs has very little practical value. More interesting and important is the recognition of peptide epitopes that constitute part of whole proteins by MIPs. Materials capable of doing this can find applications in bioanalytical chemistry, proteomics research, biotechnology and medicine. [14–16] For biotechnological applications it might be useful to identify suitable peptide tags that can be included in engineered proteins for their recognition by MIP adsorbents. These tags can compete with His and FLAG tags that have limitations such as relatively low affinity and high polarity. [17,18] We believe that our work will provide insight and guidance for selecting appropriate peptide sequences for creating efficient MIPs for peptides and proteins.

Table 1. Average (mean) pK_d for template–receptor pairs.

Template/receptor	Binding sites created by amino acids	Binding sites created by dipeptides	Binding sites created by tripeptides
Amino acids	2.90	2.64	3.79
Dipeptide	2.94	3.35	5.25
Tripeptide	2.13	2.83	5.56

Table 2. Preferential binding of amino acids, dipeptides and tripeptides by selected MIPs binding sites formed by tripeptides. Only tripeptide templates are included capable of forming MIPs with mean pKd > 6.0.

Set of MIPs	Set of targets
Glu-Glu-Gly	Leu-Leu-Leu, Lys-Ser-Glu
Glu-Gly-Ser	Gly
Glu-Leu-Glu	Ser, Gly-Gly-Glu
Glu-Glu-Gly	Leu-Gly-Ser, Leu-Leu-Gly
Glu-Gly-Ser	
Glu-Glu-Gly	Glu-Leu, Gly-Glu, Gly-Gly, Gly-Leu, Gly-Lys, Gly-Ser, Leu-Leu, Lys-Lys, Ser-Glu, Ser-Leu, Ser-Lys,
Glu-Leu-Glu	Glu-Ser-Ser, Gly-Glu-Lys, Gly-Glu-Ser, Gly-Gly-Leu, Gly-Gly-Lys, Gly-Gly-Ser, Gly-Leu-Glu, Gly-Leu-Leu, Gly-Leu-Lys, Gly-Lys-Ser, Gly-Ser-Glu, Gly-Ser-Leu, Gly-Ser-Lys, Gly-Ser-Ser, Leu-Glu-Lys, Leu-Gly-Glu, Leu-Gly-Leu, Leu-Gly-Lys, Leu-Leu-Lys, Leu-Leu-Ser, Leu-Lys-Glu, Leu-Lys-Leu, Leu-Lys-Ser, Leu-Ser-Glu, Leu-Ser-Leu, Leu-Ser-Ser, Lys-Glu-Ser, Lys-Gly-Lys, Lys-Gly-Ser, Lys-Leu-Glu, Lys-Leu-Leu, Lys-Lys-Ser, Lys-Ser-Ser, Ser-Glu-Leu, Ser-Gly-Lys, Ser-Gly-Ser, Ser-Leu-Leu, Ser-Leu-Lys, Ser-Leu-Ser, Ser-Ser-Lys
Glu-Gly-Ser	Glu, Leu, Gly-Leu-Gly
Glu-Leu-Glu	
Glu-Glu-Gly	Lys, Glu-Glu, Glu-Gly, Glu-Lys, Glu-Ser, Leu-Glu, Leu-Gly, Leu-Lys, Leu-Ser, Lys-Glu, Lys-Gly, Lys-Leu,
Glu-Gly-Ser	Lys-Ser, Ser-Gly, Ser-Ser, Glu-Glu-Glu, Glu-Glu-Gly, Glu-Glu-Leu, Glu-Glu-Lys, Glu-Glu-Ser, Glu-Gly-
Glu-Leu-Glu	Glu, Glu-Gly-Gly, Glu-Gly-Leu, Glu-Gly-Lys, Glu-Gly-Ser, Glu-Leu-Glu, Glu-Leu-Gly, Glu-Leu-Leu, Glu-Leu-Lys, Glu-Leu-Ser, Glu-Lys-Glu, Glu-Lys-Gly, Glu-Lys-Leu, Glu-Lys-Lys, Glu-Lys-Ser, Glu-Ser-Glu, Glu-Ser-Gly, Glu-Ser-Leu, Glu-Ser-Lys, Gly-Glu-Glu, Gly-Glu-Gly, Gly-Glu-Leu, Gly-Gly-Gly, Gly-Leu-Ser, Gly-Lys-Glu, Gly-Lys-Gly, Gly-Lys-Leu, Gly-Lys-Lys, Gly-Ser-Gly, Leu-Glu-Glu, Leu-Glu-Gly, Leu-Glu-Leu, Leu-Glu-Ser, Leu-Gly-Gly, Leu-Leu-Glu, Leu-Lys-Gly, Leu-Lys-Lys, Leu-Ser-Gly, Leu-Ser-Lys, Lys-Glu-Glu, Lys-Gly-Gly, Lys-Glu-Leu, Lys-Gly-Lys, Lys-Gly-Glu, Lys-Gly-Gly, Lys-Gly-Leu, Lys-Leu-Gly, Lys-Leu-Lys, Lys-Leu-Ser, Lys-Lys-Glu, Lys-Lys-Gly, Lys-Lys-Leu, Lys-Lys-Lys, Lys-Ser-Gly, Lys-Ser-Leu, Lys-Ser-Lys, Ser-Glu-Glu, Ser-Glu-Gly, Ser-Glu-Lys, Ser-Glu-Ser, Ser-Gly-Glu, Ser-Gly-Gly, Ser-Gly-Leu, Ser-Leu-Glu, Ser-Leu-Gly, Ser-Lys-Glu, Ser-Lys-Gly, Ser-Lys-Leu, Ser-Lys-Lys, Ser-Lys-Ser, Ser-Ser-Glu, Ser-Ser-Gly, Ser-Ser-Leu, Ser-Ser-Ser
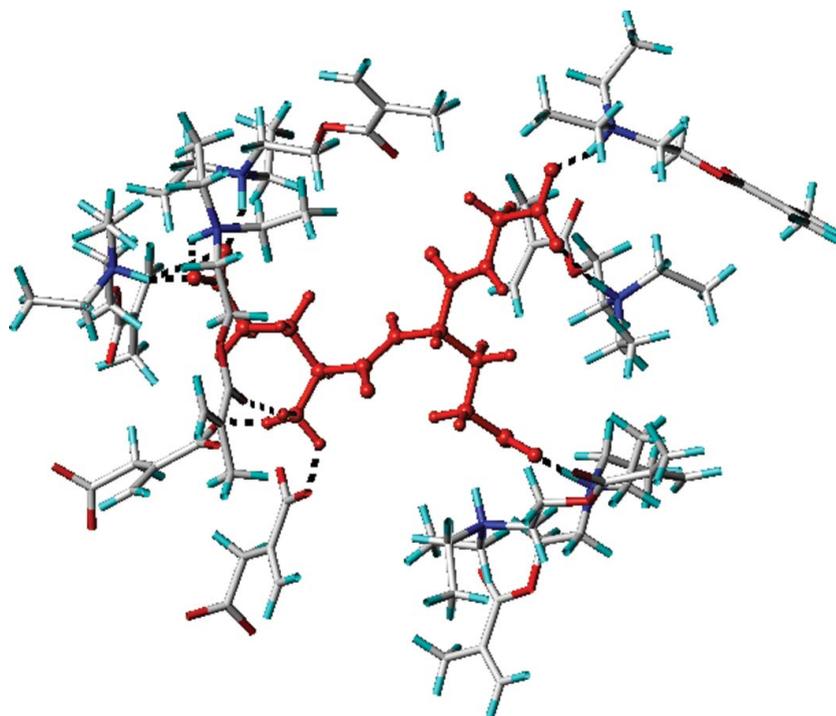
Figure 4. Structure of binding site for Glu-Glu-Gly (red ball and stick) surrounded by seven DEAEM and two itaconic acid monomer (stick – white (C), red (O), blue (N) and cyan (H)).
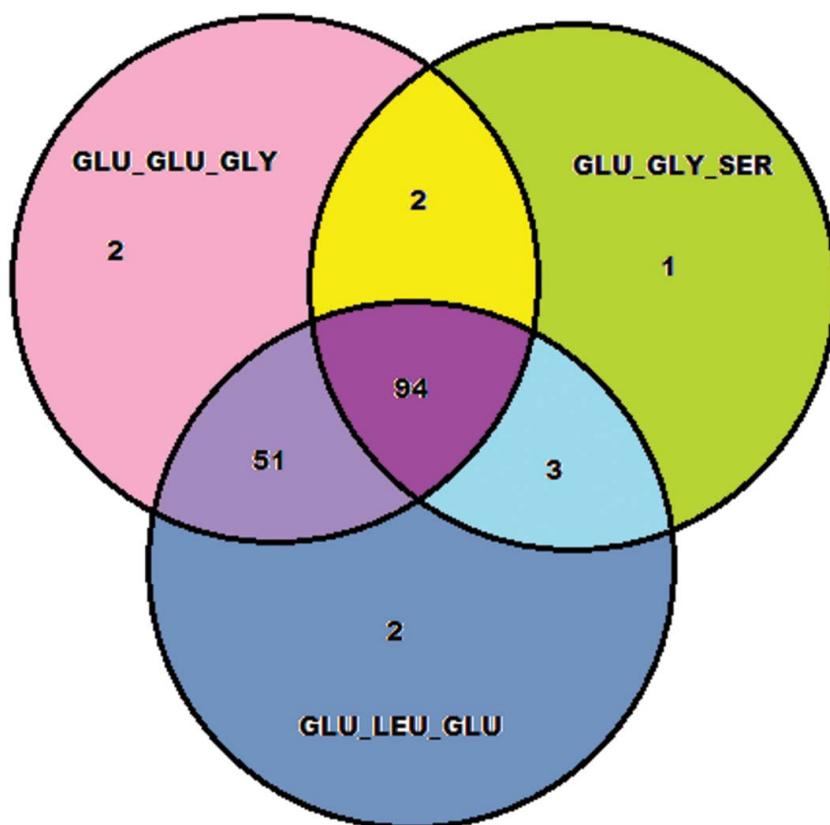
Figure 5. Venn diagram for tripeptides which bind amino acids, di- and tripeptides. Composition of each area is presented in Table 2.

2.2. Conclusions

Virtual molecularly imprinted binding sited were formed for 170 amino acids, dipeptides and tripeptides. The binding analysis has proven that only tripeptides were capable of forming high affinity binding sites in MIPs (with $pK_d > 6.0$). No appreciable specificity was demonstrated in binding of virtual binding sites and corresponding templates. Only tripeptide templates were capable of forming binding sites with unique specificity for corresponding targets. It is possible to conclude that although tripeptide sequences are sufficiently long to form MIPs with relatively high affinity, the sequence of peptide epitope should be substantially longer to ensure specificity of imprinted sites. This consideration will be useful for the design of highly efficient MIPs for proteins.

3. Materials and methods

All modelling was performed using the Sybyl 7.3 software package (Certara Inc.) on an HP EliteDesk G1 Tower PC running CentOS Linux 7. The protocol for the rational design of MIPs has been previously described to select monomers for a given template, creating a database for screening the template using the LEAPFROG™ algorithm.

3.1. Minimization of ligands

Using the Biopolymer Build Biopolymer function a set of 155 ligands were drawn in their charged state and minimized as previously described. [4–7] The 155 ligand set comprised of the initial set of five amino acid ligands G, S, L, E and K and all the combinations of di- and tripeptides using these five amino acids resulting in 25 dipeptides and 125 tripeptides (155 in total) in their charged forms.

3.2. Creation of MIP receptor

The 155 ligands were used as templates to create 155 binding site models (one for each template) following the computational protocol that is a variation of the method for the design of MIPs, developed in our group by taking a library of 25 commonly used functional monomers and screening them against the template. In this case a library of 12 monomers were used for screening against the 155 templates a shown in Table 3.

The protocol was to a build a receptor for each template using repeated LEAPFROG runs was carried out stepwise. Each of the monomers in the library was then probed for its possible interaction with the template using the LEAPFROG^{TM} algorithm that was used to screen the library of functional monomers for their possible interactions with the template. The program was applied for 20,000 iterations. The results from this were examined and the empirical binding score evaluated. The monomer (monomer one) giving the highest binding score represented the best binding monomer that forms the strongest complex with the template from the first screen.

The monomer:template complex from the first screen was then used as a template and screening using LEAPFROG was repeated. Prior to the second screen, monomer one and template were merged together in the same orientation and conformation found in the complex (using freeze view all function) and monomer one was given zero charge so that in the second screen the monomer library will only complex with the template and not the first monomer to find monomer two. This was repeated until the template was saturated with monomers, and stopped when the monomer screen resulted in interaction with monomers and not the template. At this point the MIP receptor has been created for the template and this was carried out on all 155 ligand templates resulting in 155 MIP receptors ensuring the charges for each receptor were added at the end and saved.

3.3. Creation of protomol for docking

The 155 MIP receptors generated were prepared for docking experiments against a database of ligands by creating a protomol for each MIP receptor to be screened or docked

Table 3. Database of 12 monomers used to create binding site for 155 templates.

Monomers	Neutral	Charged
Acrylamide	✓	
Bisacrylamide	✓	
2-(Diethylamino)ethyl methacrylate	✓	✓
2-Hydroxyethyl methacrylate	✓	
Itaconic acid	✓	✓
Methacrylic acid	✓	✓
2-(Trifluoromethyl)acrylic acid	✓	✓
1-Vinylimidazole	✓	✓

against a database. Within SYBYL software, Surflex-Dock (BioPharmics LLC) was used as the docking program. Surflex-Dock uses an empirical scoring function and a patented search engine to dock ligands into receptor's binding site. Docking is guided by the proto-mol, an idealized representation of a ligand that makes every potential interaction with the binding site. The protomol was prepared for each MIP receptor using the Surflex-Dock manual in the Surflex-Dock (SFXC) Docking Mode. The receptor preparation is carried out by extraction of the ligand (template) and generating the protomol using the automatic protomol generation mode.

3.3.1. Docking using Surflex-Dock

The database of ligands for docking was the 155 ligands plus the remaining 15 amino acids to give a final library of 170 ligands in an mdb database folder containing the files _tamol2dbloc and _tamol2dbset. This database folder is added as the ligand source and the docking is run that takes approximately 25 minutes. After the completion of a docking run, the binding score of each ligand was calculated. The final table of results contains 155 columns (one for each MIP receptor) and 170 rows (one for each ligand) with the docking score obtained using Surflex-Dock.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] A. Rachkov, N. Minoura. Recognition of oxytocin and oxytocin-related peptides in aqueous media using a molecularly imprinted polymer synthesized by the epitope approach. *J. Chromatogr. A*, **2000**, *889*, 111–118.
[2] R. Vita, J. A. Overton, J. A. Greenbaum, J. Ponomarenko, J. D. Clark, J. R. Cantrell, D. K. Wheeler, J. L. Gabbard, D. Hix, A. Sette, B. Peters. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*., **2015**, *43*, D405-D412.
[3] M. J. Whitcombe, I. Chianella, L. Larcombe, S. A. Piletsky, J. Noble, R. Porter, A. Horgan. The rational development of molecularly imprinted polymer-based sensors for protein detection. *Chem. Soc. Rev.*, **2011**, *40*, 1547–1571.
[4] S. Piletsky, E. Piletska, K. Karim, G. Foster, C. Legge, A. Turner. Custom synthesis of molecu-lar imprinted polymers for biotechnological application preparation of a polymer selective for tylosin. *Anal. Chim. Acta*, **2004**, *504*, 123–130.
[5] S. A. Piletsky, K. Karim, E. V. Piletska, C. J. Day, K. W. Freebairn, C. Legge, A. P. F. Turner. Recognition of ephedrine enantiomers by molecularly imprinted polymers designed using a computational approach. *Analyst*, **2001**, *126*, 1826–1830.
[6] E. Piletska, J. Kumire, T. Sergeyeva, S. Piletsky. Rational design and development of affinity adsorbents for analytical and biopharmaceutical applications. *J. Chin. Adv. Mater. Soc.*, **2013**, *1*, 229–244.
[7] I. Chianella, M. Lotierzo, S. A. Piletsky, I. E. Tothill, B. Chen, K. Karim, A. P. F. Turner. Ratio-nal design of a polymer specific for microcystin-LR using a computational approach. *Anal. Chem.*, **2002**, *74*, 1288–1293.
[8] T. A. Pham, A. N. Jain. Parameter estimation for scoring protein-ligand interactions using negative training data. *J. Med. Chem.*, **2006**, *49*, 5856–5868.
[9] J. J. Langham, A. E. Cleves, R. Spitzer, D. Kirshner, A. N. Jain. Physical binding pocket induc-tion for affinity prediction. *J. Med. Chem.*, **2009**, *52*, 6107–6125.

[10] O. Ramström, I. A. Nicholls, K. Mosbach. Synthetic peptide receptor mimics: highly stereoselective recognition in non-covalent molecularly imprinted polymers. *Tetrahedron: Asymmetry*, **1994**, *5*, 649–656.

[11] M. Kempe. Antibody-mimicking polymers as chiral stationary phases in HPLC. *Anal. Chem.*, **1996**, *68*, 1948–1953.

[12] E. Papaioannou, C. Koutsas, M. Liakopoulou-Kyriakides. Molecularly imprinted polymers for RGD selective recognition and separation. *Amino Acids*, **2009**, *36*, 563–569.

[13] H. M. Geyson, S. J. Barteling, R. H. Meloen. Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein. *Proc. Natl. Acad. Sci. USA*, **1985**, *82*, 178–182.

[14] H. Nishino, C.-S. Huang, K. J. Shea. Selective protein capture by epitope imprinting. *Angew. Chem. Int. Ed.*, **2006**, *45*, 2392–2396.

[15] A. Rachkov, N. Minoura. Towards molecularly imprinted polymers selective to peptides and proteins. The epitope approach. *Biochim. BIophys. Acta, Protein Struct. Mol. Enzymol.*, **2001**, *1544*, 255–266.

[16] C.-H. Lu, Y. Zhang, S.-F. Tang, Z.-B. Fang, H.-H. Yang, X. Chen, G.-N. Chen. Sensing HIV related protein using epitope imprinted hydrophilic polymer coated quartz crystal microbalance. *Biosens. Bioelectron.*, **2012**, *31*, 439–444.

[17] S. J. Costa, A. Almeida, A. Castro, L. Domingues, H. Besir. The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Appl. Microbiol. Biotechnol.*, **2013**, *97*, 6779–6791.

[18] J. J. Lichty, J. L. Malecki, H. D. Agnew, D. J. Michelson-Horowitz, S. Tan. Comparison of affinity tags for protein purification. *PREP*, **2005**, *41*, 98–105.