Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Bioorganic & Medicinal Chemistry Letters 14 (2004) 1389–1393
Bioorganic & Medicinal Chemistry Letters

Dual surface selection methodology for the identification of thrombin binding epitopes from hotspot biased phage-display libraries

Srivats Rajagopal, Roberto Meza-Romero and Indraneel Ghosh*

Department of Chemistry, University of Arizona, Tucson, AZ 85721-0041, USA

Received 7 July 2003; accepted 8 September 2003

Abstract—Protein libraries biased towards amino-acid residues found at so-called 'hotspots' were incorporated into the beta-sheet region of the thermostable variant (HTB1) of the B1 domain of the immunoglobulin (IgG) binding protein G and expressed as gene 3 fusions on M13 bacteriophage. The HTB1 library (2.2×10^9) variants with a minimal 12 amino acid basis set were selected for binding IgG, to ensure structural conservation, and subsequently to thrombin to evolve a thrombin-binding function. We believe that this dual surface selection strategy will have great utility in evolving new bi-functional proteins without compromising structure. Furthermore the discrete beta-sheet epitopes identified by our methodology will lend itself to small-molecule mimicry of beta-sheets.

1. Introduction

The design of small proteins, peptides and small-molecules to both recognize protein surfaces and disrupt protein–protein interactions is of great current interest for the development of inhibitors as well as for generating tools for delineating proteome function.[1–3] The most challenging aspect of this approach is the recognition of large (>600 A^2) and exceedingly flat surface areas as opposed to well-organized clefts at enzyme active sites. Most design approaches to date rely on residue-grafting, where structural data guides the selection of essential residues involved in binding that can be subsequently grafted on other proteins/peptides or constrained molecular scaffolds. Structure guided efforts have been used to disrupt protein dimers,[3] that include HIV-1 protease,[4] HSV ribonucleotide reductase[5] and HLH-E47[6] among others. Helical grafting has been another common theme and is best described in experiments from the DeGrado,[7] Serrano,[8] Schepartz,[9] and Hamilton[10] laboratories, where helical motifs were used to mimic natural helix–protein interactions in calmodulin/calcineurin, IL-4/IL-4 receptor, Bcl2/Bak and gp41 respectively. However, the same knowledge based design that allows protein-grafting and dimerization-inhibition excludes the discovery of new modalities of targeting protein surfaces that do not exist in a natural biological context.

In order to expand the scope of protein surface targeting non-structure guided selection strategies have also been successfully employed.[11,12] The underlying structural basis of this discovery based approach for identifying new modes of protein–protein interactions was detailed in a seminal paper by Wells et al., which clearly established that structurally divergent epitopes could recognize the same surface or 'hotspot'[13] on immunoglobulins.[14] The major hurdle for establishing a rational structural context for powerful selection techniques such as antibody and peptide-phage display is that the recognition motifs reside on loops rather than on well-defined helices or sheets. Selection approaches that can provide a structurally conserved helix or a sheet for protein-recognition will immediately allow for the grafting of residues to synthetic scaffolds as recently demonstrated for Bcl-X(L) by Hamilton et al.[15]

In this paper we describe our dual surface selection approach that allows us to extensively mutate a beta-sheet structural element in a protein while retaining its structural integrity. Our methodology has the potential to provide well-defined epitopes whose ligand binding properties can be rationalized in a structural context. We demonstrate our methodology using the B1 domain of the immunoglobulin (IgG) binding protein G. The B1 protein has been well characterized by both high resolution NMR spectroscopy[16] and X-ray crystallography.[17] This protein (Fig. 1) displays a compact globular fold and

* Corresponding author. E-mail: ghosh@u.arizona.edu

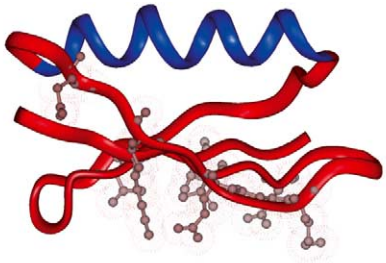
0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2003.09.098

recognizes IgG with high affinity ($K_a = 3\times10^7$ M^{-1}) using the helical surface. Furthermore, recent work in protein design has resulted in the identification of a hyperthermophilic variant[18] of B1, HTB1, which has a melting temperature greater than 100 °C. Thus, HTB1 should provide the necessary rigid framework for the proposed randomization of surface exposed beta-sheet residues (Fig. 1). The stability of the selected mutant proteins in the HTB1 scaffold should serve as protection against proteolysis, a major hindrance in peptide based drug design. Furthermore, literature precedence suggests that a similar protein, peptostreptococcal protein L (62 residues), can retain both IgG binding and native-like folding characteristics with variants that have 9 substitutions in the 14 residues that were mutagenized.[19,20]

Residues in red (Fig. 1) of the parent HTB1 protein were chosen for randomization by phage-display. These residues (Fig. 1) exist at the solvent exposed interface and are not involved in either core-packing, necessary for stability or IgG recognition, necessary for selection. Phage display technology[21,22] provides the ideal methodology for the generation of randomized HTB1 (rTB1) variants.

Our work also builds upon the empirical observations of 'hot spots' in protein interfaces[23,24] that suggests that there is a definite preference for certain amino acid residues such as tryptophan (21%), arginine (13.3%) and tyrosine (12.3%) at protein interfaces. In this paper we describe the application of the 'hot spot' hypothesis, to construct a HTB1 mutant library (rTB1) deliberately biased towards coding for **Trp**, **Arg**, **Tyr**, and **His** using only 12 of the 20 naturally occurring amino acids. We also describe the selection of structured, well-folded hTB1 variants that are able to bind thrombin as a test case.

Thrombin is a widely studied serine protease implicated in the arterial and venous thrombosis. It initiates clotting of blood by cleaving fibrinogen to fibrin and activating platelets[25] and coagulation factors like V, VIII and XIII. The thrombin surface is highly charged; flanking the acidic region at the end of the active site are two basic patches, which constitute the fibrinogen, hirudin and thrombomodulin binding exosite-I and the heparin binding exosite-II.[26] Hirudin which is one of the strongest inhibitors of the thrombin–fibrinogen interaction binds the anion binding exosite I via a carboxy terminus tail with a K_d of 0.3 pM[27] whereas the more highly electropositive exosite II binds heparin with a K_d of 6-10μM.[28] Thus, the multiple modes of thrombin recognition make it an ideal test case to validate our methodology.



```
wt B1     MAQTYKLILNGKTLKGETTTEAVDAAT-
HT B1     MAQTFKLIINGKTLKGEITIEAVDAAE-
rTB1      MAQTFXLXIXGKTLXXXIXIXAVDAAE-
```

Figure 1. Ribbon diagram of B1 where residues in ball and stick reflect sites targeted for randomization, and the blue helix is required for IgG binding. Shown below are the amino acid sequences for the first two strands of wild type B1 (wtB1), the hyperthermophilic variant (HTB1) and the proposed randomized sequence (rTB1), changes from wt B1 are in bold and residues (X) reflect sites for randomization.

2. Materials and methods

2.1. Library construction

The phage display vector pHTB1 was obtained by cloning a synthetic construct encoding for HTB1 into the pCANTAB5E phage-display vector (Pharmacia) between the Sfi I and Not I restriction sites. The HTB1 construct incorporates a unique restriction site, Pst I, at residue 26 allowing for creating two alternative cassettes for library incorporation (Sfi I/Pst I or Pst I/Not I). The rTB1 library was created by inserting a Sfi I/Pst I containing cassette that incorporated randomized codons NXY at each position marked for randomization in Figure 1. N represents any of the four bases, X represents an equal mix of A/G, and Y represents an equal mix of T/G. This generates 16 possible codons encoding **Tyr**, **His**, Asn, Asp, Gln, Lys, Glu, Cys, **Arg**, Ser, Gly, and **Trp**.

2.2. Panning for anti-thrombin hTB1 variants

We have utilized the described design strategy in our approach to successfully create a library of 2.2×10^9 rTB1 protein variants in the pHTB1 phagemid vector and expressed them as gene III fusions on M13 phage using standard protocols.[9,21] Phagemid stocks were prepared from PEG precipitates of culture broths of XL-1 Blue cells that contained the rTB1 library and superinfected with M13KO7 helper phage.[29] We sequenced 40 clones from the starting library to clearly establish that our library was representative of the requisite diversity at both the codon level and at each randomized site. An initial panning step performed against immobilized human IgG resulted in the recovery of 42 + 6% of our library relative to HTB1 displaying phage. This subset of the original library retained their ability to bind immobilized human IgG and provided the starting library for subsequent selections.

Human thrombin (T1063, Sigma) (5 μM) was immobilized in a microtiter plate at 4 °C in PBS overnight prior to phage binding. The protein containing wells were washed 4 times with PBS/Tween-20 and used directly in the presence of detergent (Tween-20) to prevent non-specific binding (BSA blocking gave unreliable results). PEG precipitates of overnight cultures of HTB1 variants (rTB1) expressed on M13KO7 phage were incubated with thrombin for 3 h at 4 °C (rounds 1–4) followed by 8 washes with buffer A (50 mM NaCl, 0.05% Tween-20, 150 mM PBS) followed by elution with high salt (1 M NaCl)

at pH 2. The final 9th round selection entailed a 1 h incubation time (to populate rTB1 variants with faster k_{on} and slower k_{off}) in 600 mM high salt buffer and elution with 2M NaCl at pH 2. The input and output titers of phage were measured after each round and at least 10–20 output clones were sent for sequencing after the completion of each round of panning.

2.3. Protein sub-cloning and purification

Two of the anti-thrombin rTB1 variants, 9sr3 and 9sr10, were sub-cloned into the BamH1/Xma I site of the pQE30 his-tag vector (Qiagen) using appropriate primers and confirmed by DNA sequencing. The proteins were expressed in XL-I blue cells by induction with IPTG at an OD of 1. Cells were harvested after 6–12 h and lysed. All proteins were purified from the soluble fractions on Ni-NTA (Qiagen) resin following the manufacturers protocol followed by buffer exchange into buffer A (20 mM Tris at pH 8.0) and subsequent passage over a size exclusion column (Sephadex G-50, Pharmacia) in buffer A. Protein purity was judged to be >95% by SDS-PAGE.

2.4. Circular dichroism (CD) spectroscopy

CD spectra were recorded on an Aviv 62A-DS spectropolarimeter. All measurements reported were carried out at 25 °C in 20 mM Tris buffer and 100 mM NaCl, using a 1 mm pathlength cuvette. Protein concentrations were determined by intrinsic absorbance at 280 nm under denaturing conditions.

2.5. Thrombin inhibition assay

The activity of two thrombin binding proteins, 9sr3 and 9sr10, were assayed for thrombin inhibition using a fluorogenic substrate (Benzoyl-Phe-Val-Arg-AMC; Calbiochem). The assays were conducted in 20 mM Tris buffer at pH 8 (0.01% Tween-20). Thrombin was incubated with different concentrations of the phage-derived proteins at room temperature for 30 min followed by addition of substrate. The cleavage of substrate was followed by excitation at 370 nm and the emission at 450 nm over two min. Hirudin (Sigma) was used as a positive control.

3. Results and discussion

3.1. Initial selection results

In order to judge whether the 12 amino acid 'hotspot' biased library was suitable for selecting thrombin binding proteins we wanted to see if there was any increase in the overall composition of Trp, Tyr, Arg and His during the course of the in vitro evolution experiment. The results (Fig. 2) clearly show that there is a significant increase in Trp, Tyr, Arg, His, and surprisingly, Ser. There is also a concomitant decrease in Gln, Asp, Glu and Cys whereas other residues are relatively constant. This gross increase in the types of amino acids supports the hotspot hypothesis to a first approximation and calls

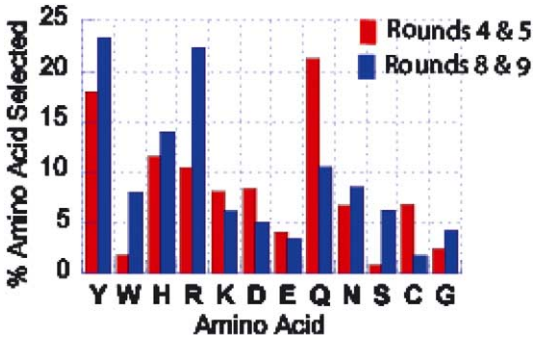


Figure 2. Histogram depicting the change in residue composition within library members (20 clones or 160 amino acids each) going from rounds 4 & 5 to rounds 8 & 9.

for further evaluation of the utility of using subsets of amino acids for targeting protein–protein interactions.

3.2. Structural landscape

Some of the sequences identified after 9 rounds of panning are shown in Figure 3. The rigid surface of the beta-sheet along with our dual selection approach allows us to align our phage-display selected protein-binding library based on a beta-hairpin structural motif as opposed to the more traditional straight-chain alignment, which lacks any structural information. We only show 3 distinct families of sequences (9sr4 and 9sr 8 were seen multiple times and are currently under investigation). What is immediately apparent by a simple scan of Figure 3 is that the minimal 12 amino acid 'hotspot' biased beta-sheet libraries do indeed display distinct epitopes or consensus-clusters for protein recognition and provide a rich diversity of sequences for both functional and structural studies directed towards understanding the nature of protein–protein interactions.

3.3. Characterization of two thrombin binders

Two of the thrombin binding proteins, 9sr3 and 9sr10, were selected for further study. 9sr3 was chosen because the incorporation of 4 Asn and 1 Gln residue could possibly compromise structure, a prerequisite in our dual surface selection methodology. 9sr10 was chosen as a member of the same family with a similar charge on the N-termini as 9sr3 but a negative charge on the C-termini.

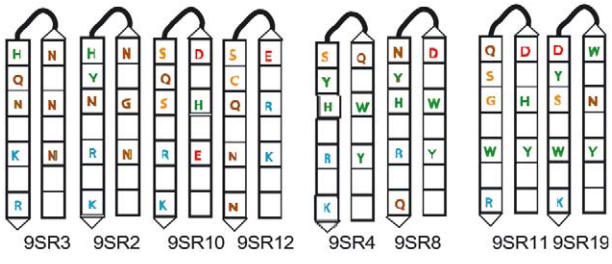


Figure 3. The representation of three families of clones identified against thrombin from round 9. The sequences are represented on an antiparallel beta-sheet motif as all the clones still maintain their ability to bind IgG, which implies structural integrity.

The purified proteins were investigated for structural integrity by CD spectroscopy and for thrombin binding by using an inhibition assay with a fluorogenic substrate.

The CD spectra (Fig. 4) clearly shows that both 9sr3 and 9sr10 are structured in solution and have spectra similar to that of the authentic hTB1 protein.[18] 9sr3 is estimated to be 10% less structured than 9sr10 based upon the minima at 220 nm. Thus, unlike the smaller avian pancreatic polypeptide fold[9] our proteins maintain structure at room temperature while incorporating multiple substitutions.

The two proteins, 9sr3 and 9sr10, inhibited thrombin with IC_{50} values of 7.8 μM and 2.8 μM respectively (Fig. 5). Since our selection protocol does not directly select for thrombin inhibitors but rather for thrombin binders, it is interesting that both proteins inhibit thrombin activity. It is possible that 9sr10 with a Asp and Glu residue in strand 1 of the beta-sheet mimics the carboxy-tail of hirudin and binds the anion binding exosite I of thrombin, thus increasing its binding capacity relative to 9sr3 that lacks negative charges and is less structured than 9sr10. Future experiments will clearly establish the mode of thrombin binding for the various protein epitopes that we have selected using our dual surface selection strategy.

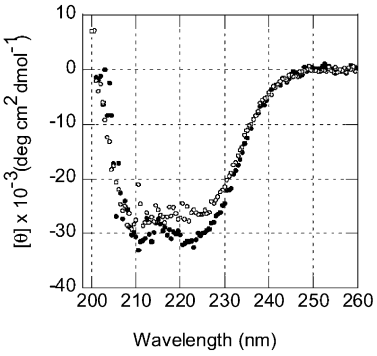


Figure 4. Circular dichroism spectra of 9sr3 (o) and 9rs10 (•) at 7 mM in a pH 8.0 (20 mM Tris and 100 mM NaCl) buffer at room temperature.

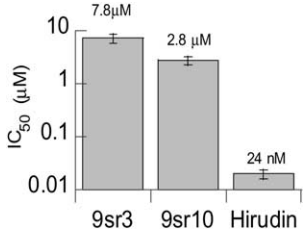


Figure 5. Inhibition of thrombin cleavage by 9s3, 9sr10 and Hirudin under our assay conditions. Thrombin concentration was fixed at 1 nM and substrate at 25 μM in these assays. The data is the average of three independent measurements. The parent HTB1 has no activity under these conditions.

4. Summary

In this paper we outline our methodology for selecting new functional epitopes while maintaining structure using the dual surface selection approach. We have demonstrated the feasibility of our methodology using the HTB1 protein scaffold and selected well-organized thrombin binding epitopes by phage-display. We have also demonstrated the utility of using a minimal 12 amino acid containing basis set enriched in amino acids preferentially found at protein interfaces. We believe our approach can facilitate the transfer of protein binding activity onto smaller synthetic scaffolds. Finally, it is worth speculating that Nature has utilized a similar strategy for evolving new protein functions using pre-existing scaffolds through an intermediate dual binding evolutionary stage.

Acknowledgements

This work was supported by the Research Corporation (Research Innovation Award), ACS (PRF TypeG) and the Leukemia and Lymphoma Society of America. We thank Min Zhou and Scott Meyer for helpful advice.

References and notes

1. Peczuh, M. W.; Hamilton, A. D. *Chemical Reviews* **2000**, *100*, 2479.
2. Cochran, A. G. *Current Opinion in Chemical Biology* **2001**, *5*, 654.
3. Zutshi, R.; Brickner, M.; Chmielewski, J. *Curr. Opin. Chem. Biol.* **1998**, *2*, 62.
4. Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J. *J. Am. Chem. Soc.* **1997**, *119*, 4841.
5. Dutia, B. M.; Frame, M. C.; Subaksharpe, J. H.; Clark, W. N.; Marsden, H. S. *Nature* **1986**, *321*, 439.
6. Ghosh, I.; Chmielewski, J. *Chemistry & Biology* **1998**, *5*, 439.
7. Ghirlanda, G.; Lear, J. D.; Lombardi, A.; DeGrado, W. F. *J. Mol. Biol.* **1998**, *281*, 379.
8. Domingues, H.; Cregut, D.; Sebald, W.; Oschkinat, H.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 652.
9. Chin, J. W.; Schepartz, A. *Angew. Chem., Int. Ed.* **2001**, *40*, 3806.
10. Ernst, J. T.; Kutzki, O.; Debnath, A. K.; Jiang, S.; Lu, H.; Hamilton, A. D. *Angew. Chem., Int. Ed.* **2001**, *41*, 278.
11. Park, S. H.; Raines, R. T. *Nat. Biotechnol.* **2000**, *18*, 847.
12. Atwell, S.; Ultsch, M.; DeVos, A. M.; Wells, J. A. *Science* **1997**, *278*, 1125.
13. DeLano, W. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14.
14. DeLano, W. L.; Ultsch, M. H.; de Vos, A. M.; Wells, J. A. *Science* **2000**, *287*, 1279.
15. Kutzki, O.; Park, H. S.; Ernst, J. T.; Orner, B. P.; Yin, H.; Hamilton, A. D. *J. Am. Chem. Soc.* **2002**, *124*, 11838.
16. Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657.
17. Achari, A.; Hale, S. P.; Howard, A. J.; Clore, G. M.; Gronenborn, A. M.; Hardman, K. D.; Whitlow, M. *Biochemistry* **1992**, *31*, 10449.
18. Malakauskas, S. M.; Mayo, S. L. *Nat. Struct. Biol.* **1998**, *5*, 470.

19. Gu, H. D.; Yi, Q. A.; Bray, S. T.; Riddle, D. S.; Shiau, A. K.; Baker, D. *Protein Sci.* **1995**, *4*, 1108.
20. Oneil, K. T.; Hoess, R. H.; Raleigh, D. P.; Degrado, W. F. *Proteins* **1995**, *21*, 11.
21. Lowman, H. B. *Methods Mol. Biol.* **1998**, *87*, 249.
22. Lowman, H. B. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 401.
23. Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1.
24. Janin, J.; Chothia, C. *J. Biol. Chem.* **1990**, *265*, 16027.
25. Fenton, J. W. *Ann. N.Y. Acad. Sci.* **1986**, *485*, 5.
26. Stubbs, M. T.; Bode, W. *Trends Biochem. Sci.* **1995**, *20*, 23.
27. Dodt, J.; Kohler, S.; Schmitz, T.; Wilhelm, B. *J. Biol. Chem.* **1990**, *265*, 713.
28. Olson, S. T.; Halvorson, H. R.; Bjork, I. *J. Biol. Chem.* **1991**, *266*, 6342.
29. Starovasnik, M. A.; Braisted, A. C.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10080.