doi:10.1016/j.jmb.2008.01.093 J. Mol. Biol. (2008) 377, 1518–1528

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies

Sara Birtalan[1], Yingnan Zhang[1], Frederic A. Fellouse[1], Lihua Shao[2], Gabriele Schaefer[2] and Sachdev S. Sidhu[1]*

[1]*Department of Protein Engineering, Genentech Inc, 1 DNA Way, South San Francisco, CA 94080, USA*

[2]*Department of Translational Oncology, Genentech Inc, 1 DNA Way, South San Francisco, CA 94080, USA*

Received 14 January 2008; received in revised form 28 January 2008; accepted 31 January 2008
Available online 12 February 2008

Synthetic antibody libraries with restricted chemical diversity were used to explore the intrinsic contributions of four amino acids (Tyr, Ser, Gly and Arg) to the affinity and specificity of antigen recognition. There was no correlation between nonspecific binding and the content of Tyr, Ser or Gly in the antigen-binding site, and in fact, the most specific antibodies were those with the highest Tyr content. In contrast, Arg content was clearly correlated with increased nonspecific binding. We combined Tyr, Ser and Gly to generate highly specific synthetic antibodies with affinities in the sub-nanomolar range, showing that the high abundance of Tyr, Ser and Gly in natural antibody germ line sequences reflects the intrinsic capacity of these residues to work together to mediate antigen recognition. Despite being a major functional contributor to co-evolved protein–protein interfaces, we find that Arg does not contribute generally to the affinity of naïve antigen-binding sites and is detrimental to specificity. Again, this is consistent with studies of natural antibodies, which have shown that nonspecific, self-reactive antibodies are rich in Arg and other positively charged residues. Our findings suggest that the principles governing naïve molecular recognition differ from those governing co-evolved interactions. Analogous studies can be designed to explore the roles of the other amino acids in molecular recognition. Results of such studies should illuminate the basic principles underlying natural protein–protein interactions and should aid the design of synthetic binding proteins with functions beyond the scope of natural proteins.

Keywords: phage display; antibody engineering; designed CDR; affinity; specificity

Edited by I. Wilson

Introduction

Most biological processes are mediated by protein–protein interactions,[1] and thus, there is great interest in the engineering of binding proteins that can be used to detect and manipulate endogenous proteins. Computational methods for the design of protein–protein interactions hold considerable promise, but inadequate knowledge of the physical basis for affinity and specificity has limited the practical application of such strategies.[2–4] Instead, most synthetic binding proteins have been derived with combinatorial approaches, which use *in vitro* evolution techniques to select specific binders from large, diverse populations.[5–10] With phage and ribosome display technologies, antibody frameworks and other scaffolds have been used to derive synthetic binding sites capable of recognizing diverse protein antigens with affinities and specificities that rival or exceed those of natural antibodies. As these display methods are amenable to automation, they may be particularly useful for applications in pro-

*Corresponding author. E-mail address: sidhu@gene.com.
Present address: F. A. Fellouse, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5.

Abbreviations used: CDR, complementarity-determining region; CDR-H*n*, heavy-chain CDR 1, 2 or 3 (where $n = 1$, 2, or 3); CDR-L3, light-chain CDR3; Fab, antigen-binding fragment; HER2, epidermal growth factor receptor 2; PBS, phosphate-buffered saline; VEGF, vascular endothelial growth factor; WT, wild type.

teomics research, which would benefit tremendously from high-throughput procedures for the generation of binding proteins.[11–13]

Combinatorial libraries of antibodies with man-made antigen-binding sites can be precisely designed in terms of chemical diversity, as advanced mutagenesis methods allow for the incorporation of controlled combinations and proportions of the 20 genetically encoded amino acids at any number of sites within a scaffold.[13] To develop highly functional naïve libraries that can be used as universal sources of ligands, it might be beneficial to bias chemical diversity in favor of those amino acids that are most likely to contribute favorably to affinity and/or specificity. Studies of natural proteins can provide guidelines, but it is not obvious what aspects of natural systems should be incorporated in optimal synthetic designs. This problem is particularly acute when different systems suggest different conclusions, as is the case for protein–protein interactions that have co-evolved over millions of years compared to naïve antibody repertoires that respond to antigen challenge over a short time frame.

Many studies of structural databases have revealed that co-evolved interfaces are enriched for large Tyr, Trp and Arg residues, which are capable of mediating a wide array of intermolecular interactions.[4,14–18] Furthermore, extensive alanine-scanning analyses have shown that these residues are often "hot spots," which contribute favorably to the binding energy.[19,20]

However, a somewhat different view of molecular recognition emerges from studies of antibody diversity. The combining sites of antibodies are formed from six complementarity-determining region (CDR) loops, and of these, the third CDR of the heavy chain (CDR-H3) is the most diverse and most important for antigen recognition.[21–23] An analysis of the germ line contribution to CDR-H3 diversity indicates that naïve loops are dominated by Tyr and small residues (Gly, Ser, Ala and Thr).[23] In contrast, Trp residues are not particularly abundant, and Arg residues are rare, but become more abundant in functional antibodies that have undergone affinity maturation (Fig. 1).[23,24] Thus, unlike co-evolved protein–protein interactions, molecular recognition by naïve antibodies relies heavily on Tyr residues but not on Arg and Trp, and consequently, Tyr side chains mediate ~25% of the antigen contacts in functional antibodies.[25] The prevalence of small residues in CDR-H3 loops implies that conformational flexibility is also important for effective antigen recognition.[21,23,25] While these observations provide intriguing possibilities for synthetic library design, there is a clear need for an empirical method that can compare alternative design strategies.

Recently, we have shown that phage-displayed synthetic antibody libraries can be used not only to derive novel binding proteins, but also to study directly the impact of chemical diversity on naïve molecular recognition.[13,26,27] We have reduced the complexity of the antigen-binding site by targeting diversity to those positions that are most often involved in antigen recognition and by biasing chemical diversity in favor of those amino acids that are most likely to establish favorable contacts. In the most extreme example of simplification, we have shown that functional antibodies can be obtained from libraries built on a single antigen-binding fragment (Fab) framework with diversity restricted to four CDR loops and only two amino acids (Tyr and Ser).[28]

These minimalist systems provide a unique opportunity for assessing directly the contributions of different chemical diversity to naïve antigen recognition. We reasoned that the limited addition of chemical diversity to the Tyr/Ser background may be used to precisely gauge the intrinsic capacity of different amino acids for mediating binding affinity and specificity. Herein, we use this approach to explore contributions to the affinity and specificity of antigen-binding sites by Tyr, Ser and two other amino acids that play distinct roles in molecular recognition. Gly was investigated because of its unique flexibility and abundance in naïve antibodies,[23] and Arg was investigated because of its important role

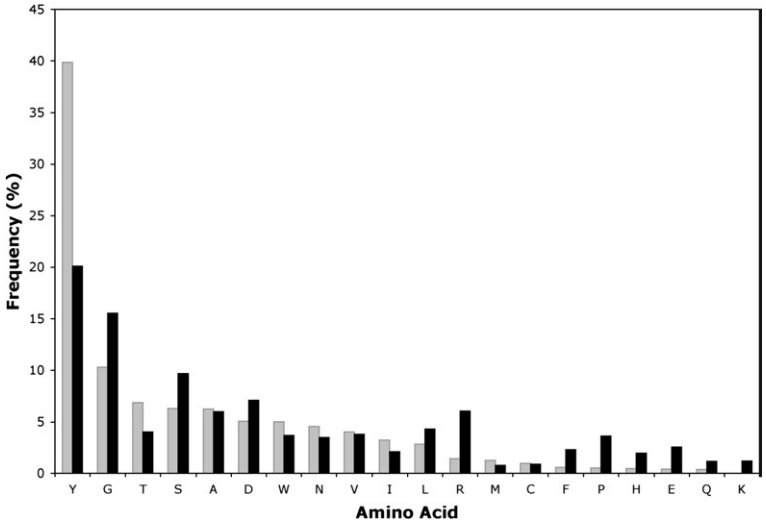

Fig. 1. Relative amino acid frequencies in germ line DH and JH gene segments that can contribute to naïve CDR-H3 loops (grey bars) and in functional CDR-H3 loops (black bars). The data are from the analysis reported by Zemlin *et al.*[23] and show the averages for human and mouse sequences. The analysis of functional CDR-H3 loops included the residues between position 94 and the position preceding position 101, exclusive. The amino acids are arranged from left to right in descending order of frequency in the germ line.