J|A|C|S
COMMUNICATIONS

Published on Web 04/03/2003

Expanding the DNA-Recognition Repertoire for Zinc Finger Proteins beyond 20 Amino Acids

Derek Jantz and Jeremy M. Berg*

Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205

Received December 10, 2002; E-mail: jberg@jhmi.edu

Designed sequence-specific nucleic acid-binding proteins hold great promise for the directed regulation of gene expression. To date, tandem arrays of Cys_2His_2 zinc finger domains have proven to be the most effective framework for the development of novel DNA-binding proteins.[1] Despite considerable effort, however, only a limited number of DNA sequences have been targeted with a high degree of specificity. In particular, attempts to design zinc finger domains specifically recognizing any base other than G in the 5′-most position of the domain-binding site have met with limited success.[2] These restrictions on zinc finger design appear to be due to the restricted number of combinations of functional groups and side-chain lengths among the 20 amino acids that normally occur in proteins.

Structural studies of Cys_2His_2 zinc finger proteins in complex with DNA have revealed a conserved $\beta\beta\alpha$ architecture stabilized by the coordination of a zinc atom by two cysteines and two histidines.[3] These proteins exhibit a highly modular mode of DNA recognition in which binding specificity is conferred primarily by three amino acids within each zinc finger domain. Typically, the amino acids in positions −1, 3, and 6 relative to the start of the α-helix make specific contacts with three consecutive bases of DNA. Appropriate alterations made to these residues can produce changes in DNA-binding specificity.[4]

The zinc finger protein QNK-QDK-RHR has previously been described and crystallographically characterized.[5] This protein consists of three tandem Cys_2His_2 zinc fingers with the contact residues $Q_{-1}N_3K_6$, $Q_{-1}D_3K_6$, and $R_{-1}H_3R_6$ and binds with high affinity to the double-stranded DNA site 5′-$G^A/_G G^G/_T C^A/_G$GAA. The Arg residue in position 6 of the carboxyl-terminal domain recognizes the 5′-most G in the binding site with high specificity through the formation of two hydrogen bonds from the Arg side-chain guanidinium group to N7 and O6 of the guanine base (Figure 1A). Glutamine residues in each of the N-terminal two fingers make similar bidentate contacts with their cognate DNA bases. Because the side-chain carboxamide of glutamine has both a hydrogen-bond donor and a hydrogen-bond acceptor, this residue tends to specify A through contacts with the N7 and N6 of the adenine base.

As a means of evaluating the importance of side-chain length in zinc finger design, a mutant of QNK-QDK-RHR was made in which the Arg residue in position 6 of the C-terminal finger was mutated to a glutamine (RH**R** → RH**Q**). The DNA-binding activities of both the parent and the mutant proteins were then characterized using a fluorescence anisotropy-based assay in which association constants were determined for an assortment of DNA binding sites representing all possible point variants in the 5′-most DNA triplet (the bases contacted by the C-terminal finger, Figure 2A and B).[6] As anticipated, the parent protein binds to its optimal site with an association constant of $4 \pm 1 \times 10^8$ M^{-1} and shows excellent specificity for G at the first position of the triplet. The Gln variant, on the other hand, binds its optimal sequence with 13-fold lower affinity and shows essentially no discrimination at position 1. These observations are consistent with the notion that the Gln side chain is too short to reach the DNA when Arg_{-1} and His_3 are the other binding residues.

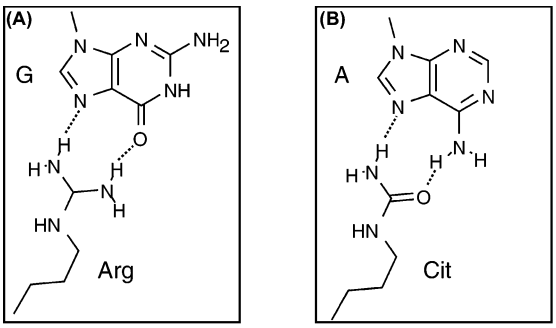

Figure 1. (A) Interaction for Arg-G recognition in QNK-QDK-RHR observed crystallographically and (B) the proposed interaction for citrulline-A recognition in QNK-QDK-RHCit.

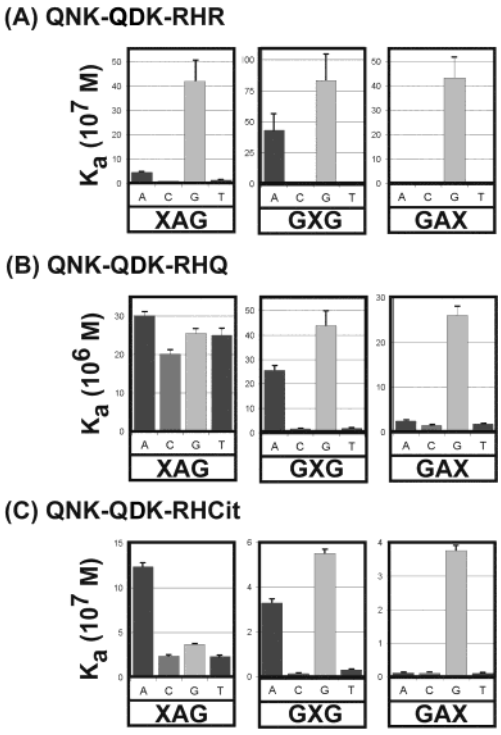

Figure 2. Association constants for the three-zinc finger proteins and 12 point variants of the 5′-**GAG**GCAGAA binding site determined by titrations monitoring fluorescence anisotropy changes of appropriately labeled oligonucleotides upon complexation with protein. (A) Data for QNK-QDK-RH**R**, (B) data for QNK-QDK-RH**Q**, (C) data for QNK-QDK-RH**Cit** containing the "unnatural" amino acid citrulline in position 6.

Overcoming this limitation required the incorporation of an amino

10.1021/ja029671i CCC: $25.00

acid in position 6 with a longer side chain and a functional group suitable for the recognition of a base other than G. Lys has a relatively long side chain, but generally prefers G (and to a lesser degree T) because, like Arg, it possesses only hydrogen-bond donors.[7] Generating novel specificity would, therefore, require an "unnatural" amino acid combining the length necessary to reach the DNA with functional groups capable of interacting favorably with bases other than G. We selected citrulline for this purpose because it has the same overall length as Arg and its ureido moiety possesses both a hydrogen donor and a hydrogen acceptor, suitable for interacting with A in a manner similar to that of Gln (Figure 1B).

To incorporate the atypical amino acid into the three-zinc finger protein, the third domain was synthesized chemically incorporating the citrulline into position 6 of the helix (RH**R** → RH**Cit**).[8] The synthetic finger 3 peptide was ligated to the first two fingers by expressed protein ligation to yield the three-finger protein QNK-QDK-RH**Cit**.[9] The DNA-binding characteristics of this semisynthetic protein were then examined as before (Figure 2C). QNK-QDK-RH**Cit** showed a clear preference for A in position 1, as anticipated, and bound its optimal sequence with an affinity of $1.2 \pm 0.1 \times 10^8$ M^{-1}. The observed level of discrimination for A is 3.4-fold over G, the base providing the second highest affinity. This is somewhat lower than the 9-fold discrimination demonstrated by QNK-QDK-RH**R**, but is much higher than that observed for QNK-QDK-RH**Q**.

The three proteins assayed show similar binding preferences at positions 2 and 3 of the first DNA triplet. Specificity at these positions is conferred by His in position 3 and Arg in position −1, respectively. Because the specificity at these positions is independent of the nature of the residue in position 6, the differences in DNA-binding affinity between the three proteins can be attributed to the interaction between the position 6 amino acid and the base in position 1. The higher overall affinity of QNK-QDK-RH**R** is likely due to the fact that Arg is positively charged and can interact favorably with DNA through electrostatic interactions in addition to the hydrogen-bonding interactions that are primarily responsible for G specificity.

Of the three DNA-contacting residues in the zinc finger binding helix, position 6 is the farthest from the DNA. This observation may be correlated with the tendency of natural Cys_2His_2 zinc finger proteins to bind DNA triplets beginning with a 5′-G through the relatively long side chains of Arg and Lys. Alternatively, zinc finger proteins incorporate relatively short side chains in position 6 that do not contact DNA, but do not prevent other contact residues from making specificity-inducing contacts.[10] We have demonstrated that this limitation can be overcome by introducing an "unnatural" amino acid that combines an appropriate functional group with an appropriate side-chain length. This is, to our knowledge, the first demonstration of specific DNA binding due to incorporation of an "unnatural" amino acid. The accessibility of zinc finger proteins by total synthesis based on peptide ligation[11] or by the semisynthetic approach that we have used will allow more complete exploration of the use of unusual amino acids to generate zinc finger DNA-binding domains with specificities not readily achievable in nature.

Acknowledgment. This work was supported by a grant from NIGMS. The authors would like to thank Gregory Gatto, Jr. for assistance with curve-fitting.

Supporting Information Available: Sample binding isotherms (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) For example: (a) Beerli, R. R.; Segal, D. J.; Dreier, B.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14628−33. (b) Beerli, R. R.; Dreier, B.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1495−500. (c) Liang, Y.; Li, X. Y.; Rebar, E. J.; Li, P.; Zhou, Y.; Chen, B.; Wolffe, A. P.; Case, C. C. *J. Biol. Chem.* **2002**, *277*, 20087−94. (d) Rebar, E. J.; Huang, Y.; Hickey, R.; Nath, A. K.; Meoli, D.; Nath, S.; Chen, B.; Xu, L.; Liang, Y.; Jamieson, A. C.; Zhang, L.; Spratt, S. K.; Case, C. C.; Wolffe, A.; Giordano, F. J. *Nat. Med.* **2002**, *4*, 4.

(2) (a) Segal, D. J.; Dreier, B.; Beerli, R. R.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2758−63. (b) Dreier, B.; Segal, D. J.; Barbas, C. F., III. *J. Mol. Biol.* **2000**, *303*, 489−502. (c) Liu, Q.; Xia, Z.; Zhong, X.; Case, C. C. *J. Biol. Chem.* **2002**, *277*, 3850−6.

(3) (a) Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. *Structure* **1996**, *4*, 1171−80. (b) Kim, C. A.; Berg, J. M. *Nat. Struct. Biol.* **1996**, *3*, 940−5. (c) Elrod-Erickson, M.; Benson, T. E.; Pabo, C. O. *Structure* **1998**, *6*, 451−64.

(4) For example: (a) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7345−9. (b) Desjarlais, J. R.; Berg, J. M. *Proteins* **1992**, *13*, 272. (c) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2256−60. (d) Jamieson, A. C.; Wang, H.; Kim, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12834−9. (e) Wolfe, S. A.; Greisman, H. A.; Ramm, E. I.; Pabo, C. O. *J. Mol. Biol.* **1999**, *285*, 1917−34. (f) Elrod-Erickson, M.; Pabo, C. O. *J. Biol. Chem.* **1999**, *274*, 19281−5.

(5) Kim, C. A.; Berg, J. M. *Nat. Struct. Biol.* **1996**, *3*, 940−5.

(6) The assay will be described in detail elsewhere. Briefly, a labeled DNA "probe" was made by annealing a pair of complimentary oligonucleotides, one of which was fluorescein-labeled on an internal thymidine base. The probe was 25 bp in length, centered around the 5′-GAGGCAGAA protein binding site. The fluorescence anisotropy of a 5 nM solution of this probe was then determined on a spectrofluorometer set with excitation and emission filters appropriate for fluorescein before and after the stepwise addition of purified zinc finger protein. As the probe was bound by the protein, the reduced tumbling rate of the resulting complex produced a concomitant increase in the anisotropy of the probe. Fitting the data from this "forward" titration produced a curve from which the K_d for the protein-probe complex could be determined. A solution of unlabeled "competitor" DNA could then be titrated into a solution of preformed protein-probe complexes, disrupting the complexes and thereby reducing the anisotropy of the probe. Fitting this "back" titration generated a curve from which the K_d for the protein-unlabeled probe complex could be determined. By carrying out 12 back-titrations in parallel using each of the point variants in the 5′-GAG triplet, the specificities conferred by each of the proteins' C-terminal fingers could be determined quantitatively.

(7) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11099−103.

(8) The C-terminal 23 amino acids of QNK-QDK-RH**X** (where X is Arg, Gln, or Cit) were synthesized using standard Fmoc chemistry. The N-terminal amino acid of this peptide was the second metal-liganding cysteine in the third finger to allow for subsequent ligation. All peptides were purified by reversed-phase HPLC and confirmed by MALDI mass spectrometry prior to ligation.

(9) (a) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705−10. (b) A region encoding amino acids 1−65 of QNK-QDK-RHR, corresponding to the N-terminal two fingers, was cloned into the PTYB2 vector (New England Biolabs) for expression as a C-terminal intein fusion. Glu65 was mutated to a glycine to avoid the possibility of racemization at the point of ligation, and a C-terminal His-tag was cloned in place of the chitin-binding domain in PTYB2 to improve expression. The fusion protein was expressed in BL21 cells and purified over a nickel column. Following elution with imidazole, purified finger 3 peptide was added in the presence of 200 mM mercaptoethanesulfonic acid, and the ligation reaction was allowed to proceed overnight at room temperature. Ligated three-finger protein was then purified by reversed-phase HPLC and quantitated spectrophotometrically. All three proteins used in this report were made in this semisynthetic fashion.

(10) Elrod-Erickson, M.; Benson, T. E.; Pabo, C. O. *Structure* **1998**, *6*, 451−64.

(11) Beligere, G. S.; Dawson, P. E. *Biopolymers* **1999**, *51*, 363−9.

JA029671I