Exploring the Role of Glutamine 50 in the Homeodomain−DNA Interface: Crystal Structure of Engrailed (Gln50 → Ala) Complex at 2.0 Å[†,‡]

Robert A. Grant,[§,||] Mark A. Rould,[§,||,⊥] Juli D. Klemm,[||,#] and Carl O. Pabo*[,§,||]

Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received January 12, 2000; Revised Manuscript Received April 17, 2000

ABSTRACT: We have determined the crystal structure of a complex containing the engrailed homeodomain Gln50 → Ala variant (QA50) bound to the wild-type optimal DNA site (TAATTA) at 2.0 Å resolution. Biochemical and genetic studies by other groups have suggested that residue 50 is an important determinant of differential DNA-binding specificity among homeodomains (distinguishing among various sites of the general form TAATNN). However, biochemical studies of the QA50 variant had revealed that it binds almost as tightly as the wild-type protein and with only modest changes in specificity. We have now determined the crystal structure of the QA50 variant to help understand the role of residue 50 in site-specific recognition. Our cocrystal structure shows some interesting changes in the water structure at the site of the substitution and shows some changes in the conformations of neighboring side chains. However, the structure, like the QA50 biochemical data, suggests that Gln50 plays a relatively modest role in determining the affinity and specificity of the engrailed homeodomain.

The homeodomain is a 60 amino acid DNA-binding motif found in a large number of eukaryotic transcription factors. There is already a wealth of genetic, biochemical, and structural information about homeodomains, but there still are intriguing questions about specificity and structure−function relationships at the protein−DNA interface. One interesting set of questions focuses on the role of residue 50, which is near the center of the recognition helix and which projects directly into the major groove. Biochemical and genetic studies of homeodomains have been interpreted to indicate that residue 50 is the central determinant of the differential specificity among homeodomains, determining which base pairs are preferred at positions 5 and 6 of the TAATNN site (*1*−*4*). However, structural studies have suggested a more complicated picture (*5*−*8*).

Crystal structures of several different homeodomains in complexes with DNA have revealed that the overall fold and DNA docking arrangements are well-conserved (*9*). These crystallographic studies are in good agreement with NMR results (*5*, *8*, *10*, *11*). The homeodomain contains an N-terminal arm that fits in the minor groove and a globular domain with three α-helices. Helix 3 (the recognition helix) fits directly in the major groove with the side chains of residues 47, 50, 51, and 54 projecting into the major groove near the TAATNN site. Structural studies have established the roles that most of these side chains play in the recognition of the TAAT subsite, with Asn51 having an especially important role in contacting the adenine at position 3 (TAATNN). However, as noted above, there has been some debate about the role of residue 50, which is near the variable bases TAATNN. A variety of side chains can occupy this position in the homeodomain family, and structural studies have failed to reveal any simple pattern of contacts. Glutamine is the most common residue at position 50, but lysine, serine, histidine, isoleucine, and cysteine are also observed in natural sequences (*12*).

Although studies of variants with changes at position 50 had suggested a key role for this residue in differential specificity, detailed structural and biochemical studies have raised questions about the role of Gln50. Its role seems slightly different in different complexes, but Gln50 often makes water-mediated contacts (*5*−*8*) and has multiple conformations (*5*, *6*) in the NMR and crystal structures. Strong, direct interactions—like the canonical glutamine−adenine contacts with a pair of hydrogen bonds (*13*)—are not observed. For example, the 2.0 Å crystal structure of the wild-type engrailed homeodomain−DNA complex shows that the glutamine side chain's interactions with the DNA are limited to a single van der Waals contact with the thymine of base pair 6 and water-mediated contacts to base pairs 4, 5, and 7 (TAATTAC) (*7*).

Mutational analysis has also been used to study the role of Gln50 of the engrailed homeodomain. Thus, QA50, an engrailed mutant with an alanine at this position, binds to the wild-type DNA site with only about 2-fold lower affinity

[†] This project was supported by an NIH grant (GM31471) to C.O.P. and by funding from the Howard Hughes Medical Institute. We also are grateful to the PEW Charitable Trusts, which provided funding for some of the initial equipment purchases.

[‡]The coordinates of this structure have been deposited at the Protein Data Bank (PDB ID: 1DU0).

* Correspondence should be addressed to this author. Email: pabo@mit.edu. Phone: (617) 253-8865. Fax: (617) 253-8728.

[§] Howard Hughes Medical Institute.

[||] Department of Biology.

[⊥] Current address: Department of Molecular Physiology and Biophysics, University of Vermont College of Medicine, Burlington, VT 05405.

[#] Current address: Incyte Pharmaceuticals, Inc., 3160 Porter Dr., Palo Alto, CA 94304.

10.1021/bi000071a CCC: $19.00 © 2000 American Chemical Society
Published on Web 06/20/2000

Table 1: Data Collection and Refinement Statistics

(A) Data Collection	
resolution	20−2.0 Å
measured reflections	115772
unique reflections	23204
completeness (overall)	95.5%
completeness (2.07−2.0 Å shell)	70.3%
average $I/\sigma I$ (overall)	17.9
average $I/\sigma I$ (2.07−2.0 Å shell)	2.2
R_{merge}[a]	0.039
(B) Refinement	
R_{cryst}[b]	0.234
R_{free}[c]	0.269
rms deviation of bond lengths	0.007 Å
rms deviation of bond angles	1.20°
number of non-hydrogen atoms	1890
number of water molecules	104
rms ΔB[d]	2.67 Å^2

[a] $R_{merge} = \sum|I - (I)|/\sum I$, where I = observed intensity and (I) = average intensity of multiple observations of symmetry-related reflections. [b] The crystallographic R factor uses all data from 20 to 2 Å excluding 1606 reflections with $F < 2\sigma F$. [c] R_{free} is calculated from 10% of the reflections which were set aside for cross-validation prior to refinement. [d] rms ΔB is the root-mean-squared difference between temperature factors of covalently bonded atoms.

than the wild-type homeodomain [$K_d = 0.79 \times 10^{-10}$ M for the wild-type; $K_d = 1.9 \times 10^{-10}$ M for the QA50 mutant (*14*)]. However, there also is some reduction in specificity, and we thought that a crystal structure of this variant might help elucidate the role of Gln50 in recognition. We therefore have solved the structure of the QA50 variant bound to the wild-type TAATTA site at 2.0 Å resolution. Our refined model shows some interesting changes in the water structure at the protein−DNA interface [a topic that has received intense attention in the study of homeodomain−DNA interactions (*7, 8, 10, 15*−*18, 27*)], and there also are some interesting changes in the conformations of side chains near Ala50. However, there is no major rearrangement at the protein−DNA interface that might compensate for the loss of a favorable contribution to the binding energy by Gln50. (In principle, the small loss of binding energy in the QA50 variant might have been explained by rearrangements in the variant complex that produce favorable new contacts and thus replace energetically important contacts that had been made by the glutamine.) Our structure thus confirms that Gln50 makes only a minor contribution to the affinity of the engrailed homeodomain−DNA interactions.

METHODS

Crystallization and Data Collection. Purified protein and DNA were prepared essentially as described (*14*), and the DNA site was the same as that used to produce crystals of wild-type engrailed (*19*). The QA50−DNA complex was prepared by mixing 7 μL of QA50 engrailed (10 mg/mL in 10 mM bis-Tris-propane, pH 7.0) with an equimolar amount of duplex DNA dissolved in 3.5 μL of 3 M ammonium acetate. Crystals were grown by vapor diffusion in hanging drops that were prepared by mixing 1 μL of complex with an equal volume of well solution (1% PEG-400, 0.25 M ammonium acetate) and then equilibrating against 1 mL of well solution. The space group of the crystals was *C2*, with unit cell parameters of $a = 126.90$ Å, $b = 45.55$ Å, $c = 71.62$ Å, and $\beta = 119.9°$. Diffraction data (Table 1) were collected from a single crystal at −145 °C. The crystal was

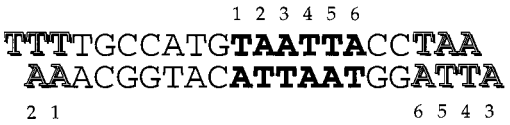

FIGURE 1: Sequence of the DNA in the cocrystal. The asymmetric unit of the cocrystal contains a 21 base pair DNA duplex with two homeodomain binding sites. The optimal site (TAATTA) is highlighted in boldface type. The suboptimal site (AAATTA, shadow highlighting) is formed at the junction between different DNA molecules, where neighboring DNA duplexes stack to form a pseudo-continuous helix in the crystal. The numbering convention used in the text to describe the position of base pairs within each site is indicated above (optimal site) and below (suboptimal site) the respective sequences. (Note that both sites are numbered in the 5′ → 3′direction.)

cryoprotected by addition of glycerol directly to the hanging drop. After a 3 min incubation, the crystal was picked up with a cryo-loop and frozen in the cold gas stream. Data were collected from a rotating anode source equipped with Yale/MSC mirrors and a R-Axis IIC detector. Intensities were integrated and scaled by the DENZO/SCALEPACK package (*20*) with a step added to premerge partials before scaling (Rould, unpublished experiments).

Structure Solution and Refinement. The structure was solved by molecular replacement using the 1.9 Å structure of the QK50 mutant (*21*), which forms nearly isomorphous crystals, as an initial phasing model. The asymmetric unit of both crystals consists of a single 20 base pair DNA duplex, with a one base overhang at each end, and two homeodomain monomers. One homeodomain in the QA50 cocrystal is bound to the optimal TAATTA site in the center of the DNA duplex; the other is bound to a suboptimal AAATTA site created by end-to-end stacking of DNA molecules in the crystal (Figure 1). The QK50 model was modified by deleting all the solvent atoms, substituting the correct residue at position 50 of the protein (alanine replacing lysine), and changing the base pairs at positions 5 and 6 of both the suboptimal and optimal binding sites (from TA to CC). It was then refined using X-PLOR (*22*) and manual rebuilding. Simulated annealing (SA) omit maps were used periodically throughout the process in order to monitor rebuilding, and the water structure at the protein−DNA interface was checked again with a SA omit map after the refinement had converged. The final model uses individual restrained atomic temperature factors and a bulk solvent correction. It includes residues 7−59 of the homeodomain bound to the optimal site (near the center of the DNA), residues 3−58 of the homeodomain bound to the suboptimal site (created by the stacking of neighboring duplexes), all of the DNA, and 109 ordered water molecules. Using data from 20 to 2.0 Å, the model has crystallographic and free R-factors of 0.235 and 0.269, respectively (Table 1). The DNA conformation was analyzed using the program Curves (*23, 24*).

RESULTS

The overall structure of the QA50−DNA complex is very similar to that of the wild-type and QK50 structures. Each homeodomain forms a compact globular structure with three α-helices and an extended N-terminal arm. The third (recognition) helix of the homeodomain fits directly into the major groove of the DNA, while the N-terminal arm wraps

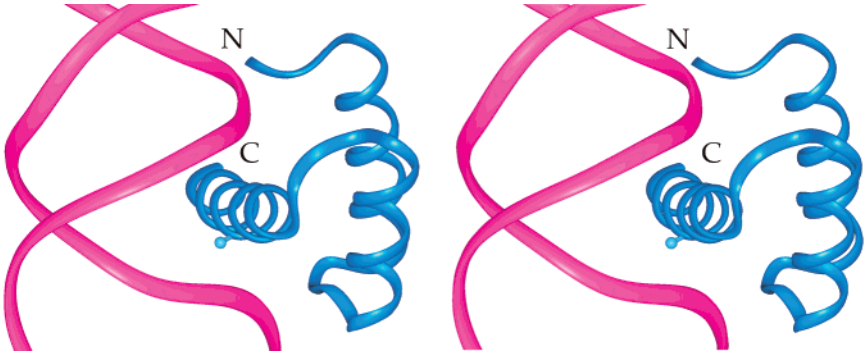

FIGURE 2: Docking of the engrailed QA50 variant homeodomain to DNA. In this stereoview, ribbons representing the DNA (magenta) and the protein (blue) demonstrate how helix 3 (the recognition helix) of the homeodomain fits into the major groove of the DNA. For reference, the methyl group of the side chain of Ala50 (near the center of the recognition helix) is represented as a ball-and-stick model. The amino (N) and carboxyl (C) terminal ends of the protein chain are labeled. The amino-terminal arm of the QA50 complex is more disordered than in the wild-type complex, so the residues that interact with the minor groove in the wild-type complex are not included in the refined model of the QA50 complex.

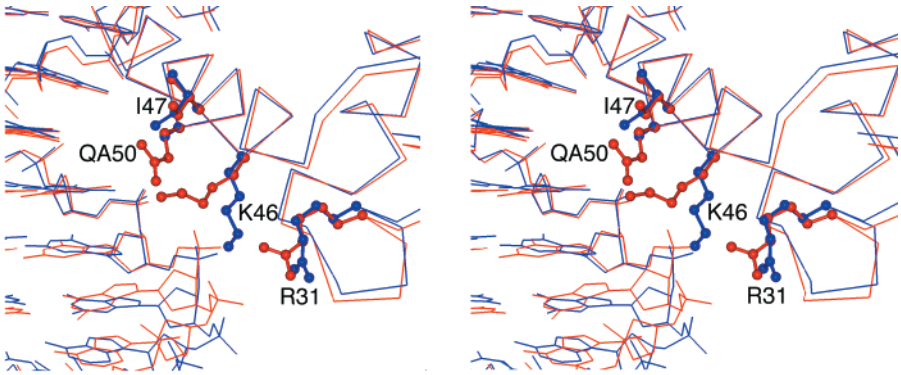

FIGURE 3: Differences in protein conformation at the protein−DNA interface. In this stereoview of the interface between the homeodomain recognition helix and the major groove of the DNA, the wild-type and QA50 structures have been aligned by superposition of the C_α's of the recognition helices. The wild-type structure is red, and the QA50 variant structure is blue. The side chain atoms for residues 31, 46, 47, and 50 of each structure are represented by ball-and-stick models. For simplicity, only the C_α trace is shown for the rest of the protein structure in this view. The figure highlights the differences in the conformations of Arg31, Lys46, and Ile47, as well as the differences in DNA backbone conformations near the Arg31 and Lys46 side chains.

around to interact with the minor groove (Figure 2). The two monomers in the asymmetric unit of the crystal are very similar to each other and share many of the same interactions with the two closely related DNA sites. Because the secondary site is created by the stacking of DNA duplexes in the crystal and has a slightly different sequence, our analysis will focus on the intact, optimal site. The C_α positions of the monomer that is bound to this optimal site align quite well with the corresponding monomers in the wild-type and QK50 structures (rms differences of 0.39 and 0.22 Å, respectively, when superimposing residues 7−59). Most of the interactions between the protein and the DNA are virtually identical to those in the wild-type engrailed complex. Critical conserved contacts include (1) the pair of hydrogen bonds between the side chain of Asn51 and adenine 3 of the binding site (TAATTA), (2) the van der Waals contact between $C^{\gamma 2}$ of Ile47 and the thymine at base pair 3, and (3) the interactions of Tyr25 and Arg53 with the DNA backbone. In addition, many of the water-mediated contacts between the protein and DNA are conserved. For example, the QA50 complex has a clathrate cage of water molecules around the side chain of Ala54 that is almost identical to the water structure in the corresponding regions of the wild-type and QK50 complexes.

Despite these overall similarities, however, there are several clear differences between the wild-type and QA50 complexes. In the QA50 structure, the N-terminal arm of the homeodomain is more disordered than in the wild-type complex, so the interactions of Arg5 and Thr6 with the minor groove are not seen. This disorder does not appear to be related in any way to the side chain substitution at position 50, and we note that comparing different homeodomain−DNA complexes shows a wide variation in the degree of ordering of the N-terminal arm. However, there also are conformational changes in three side chains (Arg31, Lys46, and Ile47) which are located at the protein−DNA interface (Figure 3) near Ala50. Two of these changes in side chain conformation (for Arg31 and Lys46) appear to be correlated with the changes in the conformation of the DNA near positions 7 and 8 of the extended binding site (TAATTACC), where a widening of the major groove pulls the DNA backbone away from the protein in the variant complex. In the QA50 variant, there also are three additional water molecules at the protein−DNA interface which help fill the gap created by the mutation of Gln50 to alanine. These differences in side chain structure, water structure, and DNA structure are discussed in more detail below.

The difference in the conformation of Ile47 involves a change of about 120° in the χ_2 torsion angle. This moves the $C^{\delta 1}$ methyl group into a position that helps fill the gap created by the glutamine-to-alanine substitution. However, in the wild-type complex, this conformation of the Ile47 side

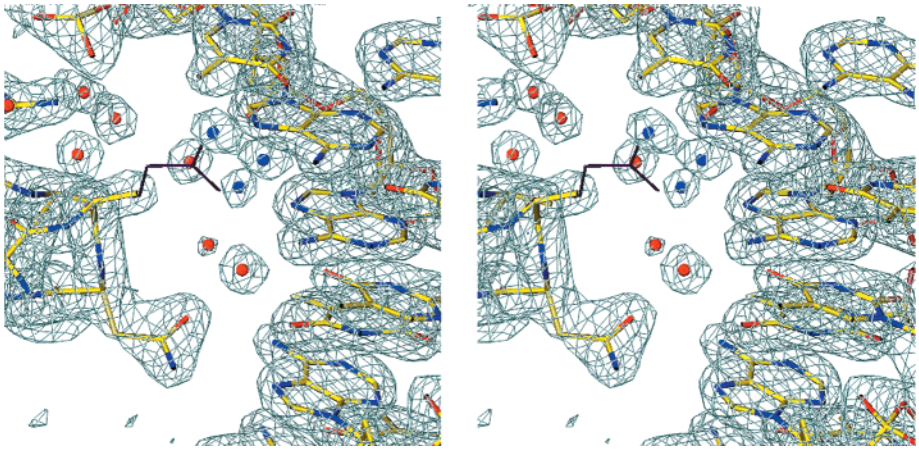

FIGURE 4: Changes in water structure adjacent to the site of the Gln50 to Ala substitution. In this stereoview, the final $2F_o - F_c$ map in the vicinity of the substitution is shown along with the refined model. The position that the glutamine side chain would occupy in the wild-type structure is also shown, but in black rather than with the atom type color coding used for the rest of the protein and DNA. The three extra waters at the interface, which help fill in the gap created by the truncation of the wild-type side chain to alanine, are colored blue. The other waters, all of which are also found in the wild-type and QK50 variant complexes, are red. The interaction between Asn51 and adenine 3 of the DNA binding site can be seen at the bottom of the figure. The map is contoured at 2σ.

chain would have produced a steric conflict with the $N^{\epsilon 2}$ of Gln50. The differences in the conformation of Lys46 may also be related to the glutamine-to-alanine change. In the wild-type complex, Lys46 makes a hydrogen bond with $O^{\epsilon 1}$ of Gln50 and a water-mediated contact with the N7 of guanine 7 (TAATTACC). The water mediating this contact also makes a hydrogen bond with the $O^{\epsilon 1}$ of Gln50. In the QA50 structure, Lys46 adopts a significantly different conformation and makes a water-mediated contacted with the phosphate of guanine 8. Finally, in the wild-type complex, the Arg31 side chain interacts with the phosphate of guanine 8, but in the variant complex it adopts a different conformation, with no direct or indirect DNA contacts.

A detailed analysis of the differences in DNA conformation in the variant and wild-type complexes revealed subtle variations in the DNA helix parameters throughout the two structures, but the most significant difference was a 2−2.5 Å widening of the major groove near the edge of the optimal site. In the QA50 complex, this widening of the major groove pulls the DNA backbone away from the protein in the region where the Arg31 and Lys46 side chains make DNA contacts at positions 7 and 8 (TAATTACC) of the wild-type complex (Figure 3). It seems clear that the resulting gap between the protein and DNA in this part of the variant complex would destabilize the interactions (as seen in the wild-type structure) involving these side chains.

Changes in the water structure at the site of the amino acid substitution suggest that water molecules can effectively fill the gap created by the mutation of glutamine to alanine. As noted above, the QA50 variant has three additional water molecules at the protein−DNA interface (Figure 4). When the QA50 and wild-type complexes are aligned by superposition of the C_α's in helix 3, we find that one of these new waters is located 0.67 Å from the position that had been occupied by the $O^{\epsilon 1}$ of Gln50 in the wild-type structure and the other new water molecules are located 1.76 and 2.34 Å away from the positions that had been occupied by the $N^{\epsilon 2}$ of Gln50. The two waters near the $N^{\epsilon 2}$ position are hydrogen bonded to the N6 of adenine 5 and to the O4 of thymine 6, while the water "replacing" the $O^{\epsilon 1}$ makes a water-mediated contact to the N7 of guanine 7 (TAATTAC). The additional waters also interact with other waters present in the wild-type complex, causing some of them to shift slightly. (Specifically, the water that mediates the interactions between guanine 7 and both the Lys46 and Gln50 side chains in the wild-type complex moves by about 1 Å.) The net effect of these changes in the water structure of the alanine variant is to produce a clathrate-like cage around alanine 50 that is similar to the one around alanine 54. The formation of this cage and the conformational change in Ile47 produce a tightly packed protein−DNA interface, in which Ala50 and the DNA are completely separated by a layer of water molecules.

Key aspects of the structure of the QA50 complex were confirmed by also examining the protein−DNA interface at the suboptimal site (AAATTA). We find that the water structure around Ala50 at the suboptimal site is identical to the structure at the optimal site, and Ile47 adopts the same rotamer as in the optimal site. Arg31, however, adopts the conformation seen in the wild-type complex and makes a corresponding phosphate contact. This is possible because the conformation of the DNA backbone is slightly different at the suboptimal site (and actually more like the wild-type complex). The electron density for Lys46 is not as clear as at the optimal site, but it is evident that the conformation of this side chain also is more like that observed in the wild-type complex than that observed at the QA50 optimal site.

In the determination of the wild-type engrailed DNA complex, data were collected at room temperature in order to avoid alterations of water structure that might occur at cryogenic temperatures. However, the QA50 variant reported here and the QK50 variant complex reported previously were solved at cryogenic temperature. A comparison of these three high-resolution engrailed cocrystal structures reveals that the structure of the ordered waters at each of the protein−DNA interfaces is remarkably conserved, except in the immediate vicinity of residue 50 (where different side chains are present). Thus, it seems that temperature differences during data collection are unlikely to have affected the water structure in a significant way.

DISCUSSION

In combination with the crystal structure of the wild-type complex and biochemical studies of the QA50 variant, our QA50 structure puts clear limits on the significance of the role of Gln50 in engrailed homeodomain−DNA interactions. Binding studies have shown that the QA50 variant has an approximately 2-fold decrease in affinity. It also has a slightly reduced specificity for T at position 5, and shows no discrimination between A and T at position 6 (where the wild-type homeodomain has a strong preference for A). These biochemical studies raised questions about the importance of Gln50—which previous studies had highlighted as a key "specificity determinant"—and we have determined the crystal structure of the QA50 variant complex to further explore the role of Gln50 in recognition. We find that removing this side chain leads to interesting changes in the water structure and in the conformations of some of the neighboring side chains. However, there is no drastic rearrangement of the interface, and we interpret this finding—in combination with the previous binding studies—as an indication that Gln50 makes only a relatively modest contribution to the affinity of binding. We also note that the most significant structural changes in the QA50 variant occur in a region adjacent to—but not within—the six base pair site normally recognized by engrailed. Thus, there is no reason to believe that these changes observed in the variant cocrystal are correlated with the altered specificity of the variant, and again this suggests a limited role for Gln50 in determining DNA binding specificity. The reasoning for these conclusions is discussed in detail below.

The most interesting conformational change in the variant structure is in the side chain of Lys46, which makes a water-mediated base contact in the wild-type complex but shifts to make a water-mediated phosphate contact in the variant complex. This altered side chain conformation could be relevant since Gln50 hydrogen bonds to Lys46 in the wild-type complex, and removing the glutamine may lead to the rearrangements observed at the protein−DNA interface of the QA50 complex. Since the conformation of Lys46 in the QA50 variant is incompatible with the wild-type conformation of Arg31, it is likely that the changes in both side chains of the variant complex are correlated with each other. The nature of these rearrangements also suggests that they are both correlated with the change in the local DNA conformation near positions 7 and 8 of the extended binding site. However, the structural differences in Lys46, Arg31, and the DNA backbone seem to affect only protein- and solvent-mediated interactions at positions 7 and 8 of the extended binding site, and in both the wild-type and QA50 complexes the protein−DNA interfaces near the base pairs at positions 5 and 6 appear to be virtually identical. Since most of the observed changes in the variant are outside of the six base pair site (TAATNN), they are probably not relevant to understanding the specificity of engrailed homeodomain−DNA interactions.

In evaluating the significance of these changes, it is also useful to consider other homeodomain sequences and structures. The consensus sequence derived from the known homeodomains includes both Lys46 and Gln50 (*12*). However, if the interaction between these two side chains (as observed in the wild-type engrailed structure) were an important determinant of homeodomain specificity, we would expect this interaction to be well conserved. Examining the structural database shows that cocrystal structures have been determined for four other homeodomain−DNA complexes that have both Lys46 and Gln50 in their sequences. In three of these four complexes (HOX-1, ultrabithorax, and antennapedia), Lys46 does not interact with Gln50 (*8*, *25*, *26*). In the cocrystal structure of the even-skipped dimer bound to DNA (*6*), one of the unique homeodomains has the two residues interacting in a manner similar to that observed in the wild-type engrailed DNA complex, but the other homeodomain in the crystal structure does not. These observations suggest that the Gln50−Lys46 interactions are not especially stable and imply that the conformational change in Lys46 observed in the QA50 variant may not be very important. This idea is also supported by the observation in the QA50 cocrystal structure that the conformation of the Lys46 side chain at the suboptimal site is similar to the conformation observed in both sites of the wild-type engrailed complex. In addition, the overall differences between the wild-type engrailed cocrystal structure and the QA50 complex involve a net loss of a direct arginine-to-phosphate contact, an observation that seems inconsistent with the relatively minor loss of DNA-binding affinity of the QA50 variant. This raises the possibility that the changes in this region of the QA50 variant may actually involve crystal packing effects that cause subtle shifts in the DNA conformation and thereby disrupt the contacts normally made by Lys36 and Arg31.

Whether the structural changes discussed above are directly related to the mutation in the QA50 variant or not, our crystal structure helps put clear bounds on the energetic significance of Gln50 in DNA binding. Despite the above complexities, the bottom line is clear: our structure proves that there are no new contacts that could possibly compensate for the loss of a significant energetic contribution by Gln50. This proves that Gln50 makes only a modest contribution to the binding energy, but does not clearly define its role in specificity. We also recognize that even small changes in affinity may be biologically significant. In the case of the QA50 variant, we infer that a small reduction in affinity and the accompanying modest loss of specificity must (together) be significant, since no known homeodomain sequence (of the hundreds analyzed) has an alanine at position 50. Even a modest improvement in specificity may be biologically relevant since the homeodomain recognizes such a short site and only has about a 100-fold preference for its specific site.

The limited energetic contribution of Gln50 contrasts with the much more favorable contacts that can occur when there is a lysine at position 50. Binding studies of a QK50 engrailed variant show that it binds tightly to the sequence TAATCC (*14*), and the crystal structure of the complex shows that Lys50 makes direct hydrogen bonds to base pairs 5 and 6 of the target sequence (*19*). As we have noted before (*19*), most of the experiments which highlight the role of residue 50 in recognition either inserted or removed a lysine from this position, so conclusions about the importance of residue 50 may have been biased by the very favorable contacts that are possible for Lys50.

There has been considerable discussion about the potential role of ordered water molecules at the homeodomain−DNA interface (*7*, *8*, *10*, *15*−*18*, *27*), and the QA50 structure

certainly provides important new data on this topic. One key observation from comparisons with other structures is that many water molecules occupy conserved positions in closely related structures. It is interesting that the water structure around Ala50 shows some similarities to the water structure around Ala54. Also, many of the water positions seen in this complex correspond directly to waters present in the engrailed wild-type and QK50 complexes (Figure 4). This confirms that certain water molecules are a consistent part of the homeodomain−DNA interface but provides no information about the "energetic significance" of these waters or their role in recognition. (It can be hard to even pose a clear question about the role of water since energies are always related to the difference between two states and it is not clear what the appropriate reference state would be in this system. For example: would it be meaningful to consider the interface in the absence of solvent?) However, our impressions about the relatively fixed positions of flanking waters must be contrasted with the local changes in water structure that compensate for the truncation of the glutamine side chain to alanine (Figure 4). It is clear that the water structure can readjust—with relatively small changes in binding energy—to accommodate other changes at the protein−DNA interface. Further studies (and a more precise definition of the appropriate reference state) will be needed to fully understand the role of water at the homeodomain−DNA interface. However, we see that water molecules can readily fill the region of the interface that is normally occupied by the Gln50 side chain, and we know that there is relatively little change in binding energy when this side chain is removed.

The overall impression from biochemical and structural studies is that Gln50 has only a nominal effect on affinity but apparently has some subtle effects on specificity via a complex set of side-chain−water and side-chain−side-chain contacts. This is very different from the canonical role often played by glutamine, which in other structural contexts can make a pair of hydrogen bonds that specify an adenine. This also is strikingly different than the clear role played by Lys50 in the QK50 variant. In short, we believe that much of the complexity and confusion involved in analyzing the role of Gln50 merely reflects the fact that it makes limited contributions to affinity and specificity.

ACKNOWLEDGMENT

We thank Robert Sauer and Sarah Ades for their contributions to the early stages of this project and for supplying the construct used for expression of the QA50 protein.

REFERENCES

1. Hanes, S. D., and Brent, R. (1989) *Cell 57*, 1275−1283.
2. Hanes, S. D., and Brent, R. (1991) *Science 251*, 426−430.
3. Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989) *Cell 59*, 553−562.
4. Percival-Smith, A., Muller, M., Affolter, M., and Gehring, W. J. (1990) *EMBO J. 9*, 3967−3974.
5. Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1993) *J. Mol. Biol. 234*, 1084−1093.
6. Hirsch, J. A., and Aggarwal, A. K. (1995) *EMBO J. 14*, 6280−6291.
7. Fraenkel, E., Rould, M. A., Chambers, K. A., and Pabo, C. O. (1998) *J. Mol. Biol. 284*, 351−361.
8. Fraenkel, E., and Pabo, C. O. (1998) *Nat. Struct. Biol. 5*, 692−697.
9. Wolberger, C. (1996) *Curr. Opin. Struct. Biol. 6*, 62−68.
10. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989) *Cell 59*, 573−580.
11. Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Muller, M., Gehring, W. J., and Wuthrich, K. (1994) *J. Mol. Biol. 238*, 333−345.
12. Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem. 63*, 487−526.
13. Pabo, C. O., and Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293−321.
14. Ades, S. E., and Sauer, R. T. (1994) *Biochemistry 33*, 9187−9194.
15. Schwabe, J. W. (1997) *Curr. Opin. Struct. Biol. 7*, 126−134.
16. Labeots, L. A., and Weiss, M. A. (1997) *J. Mol. Biol. 269*, 113−128.
17. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell 82*, 709−719.
18. Wilson, D. S., Sheng, G., Jun, S., and Desplan, C. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 6886−6891.
19. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell 63*, 579−590.
20. Otwinowski, Z., and Minor, W. (1996) in *Methods in Enzymology* (Carter, C., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
21. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., and Pabo, C. O. (1997) *Structure 5*, 1047−1054.
22. Brunger, A. T. (1992) *X-PLOR Manual Version 3.1*, Yale University Press, New Haven, CT.
23. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
24. Stofer, E., and Lavery, R. (1994) *Biopolymers 34*, 337−346.
25. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) *Nature 397*, 714−719.
26. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999) *Cell 96*, 587−597.
27. Gruschus, J. M., Tsao, D. H. H., Wang, L., Nirenberg, M., and Ferretti, J. A. (1997) *Biochemistry 36*, 5372−5380.

BI000071A