Presence of pKa Perturbations Among Homeodomain Residues Facilitates DNA Binding

Christopher M. Frenz* and Philippe P. Lefebvre

1Department of Otolaryngology
University of Liege
CHU de Liege
4000 Liege, Belgium

*Corresponding Author: cfrenz@gmail.com

Abstract - Homeodomain containing proteins are a broad class of DNA binding proteins that are believed to primarily function as transcription factors. Electrostatics interactions have been demonstrated to be critical for the binding of the homeodomain to DNA. An examination of the electrostatic state of homeodomain residues involved in DNA phosphate binding has demonstrated the conserved presence of upward shifted pKa values among the basic residue of lysine and arginine. It is believed that these pKa perturbation work to facilitate binding to DNA since they ensure that the basic residues always retain a positive charge.

Keywords: Computational Biology, Structural Biology, Protein Electrostatics, Protein Evolution

1 Introduction

Homeobox proteins are characterized by a 60 amino acid helix-turn-helix motif that is involved in the binding of DNA. Homeobox proteins have been characterized as transcription factors that regulate gene expression during the development and differentiation of neural systems within mammals, although they are also present in a wide range of eukaryotes and in bacteria as repressor proteins [1,2]. Recent examinations of the forces involved in the binding of the homeodomains of Antp, NK2, Engrailed, and Mata2 to DNA have demonstrated that DNA binding is largely an electrostatic phenomena. Unbound homeodomains were demonstrated to have chloride ions bound to the residues that will coordinate with the phosphates of DNA upon binding. During the binding process these chloride ions are displaced along with calcium ions that were bound to the DNA phosphates, and the ion interactions replaced by electrostatic coordination between homeodomain residues and the DNA phosphates [2].

Electrostatics playing a key role in ligand binding is not limited to homeodomain containing proteins, however, and electrostatic perturbations, as demonstrated by a pKa shift of greater than 2 pK units, have been recently demonstrated among residues involved in ligand binding. Within protein-protein complexes 80% of acidic residues have been demonstrated to have a negatively shifted pKa value that allows residues to maintain their negatively charged state within the complex [3]. Within the HCV NS3 helicase, the glutamate residue at position 493 was demonstrated to have a positive shifted perturbation which enabled the glutamate to retain a neutral charge and to coordinate with DNA. Perturbations were also demonstrated within the site of ATP binding and hydolysis [4] for the HCV helicase, a DEXH motif containing region of the protein, while another study has demonstrated that the related DEAD-box motif exhibited conserved electrostatic properties among multiple DEAD-box proteins [5].

Moreover, recent work has suggested that protein electrostatic states are evolutionary conserved. Conserved electrostatic interaction networks have been demonstrated in TIM-barrel proteins and the electrostatic surfaces of proteins from 4 distinct protein families and 1 superfamily have been demonstrated to be evolutionary conserved as well [6,7]. Studies of the HIV reverse transcriptase and the HCV helicase and polymerase have further demonstrated that the greatest number and magnitude of electrostatic perturbations exist in the residues that are most highly conserved [8]. Additionally, an electrostatic investigation of nucleoside monophosphate kinase proteins determined that the pKa values of highly conserved titratable groups within the protein families tended to also have highly conserved pKa values [9]. Moreover, the electrostatic potential maps of lysozyme proteins from a variety of organism sources were demonstrated to have a high degree of similarity, suggesting the evolutionary conservation of the electrostatic state of the protein [10].

Given the importance of electrostatics to the DNA binding of homeodomain containing proteins, this study seeks to determine if any electrostatic perturbations are present among the DNA phosphate binding residues of homeobox proteins and to determine whether pKa perturbations are conserved among a diversity of homeobox proteins.

2 Methods

2.1 Sequence and Structure Selection

Proteins for inclusion in this study were selected by performing a BLAST search for the homeodomain of the Oct-
1 transcription factor against the PDB sequence database [11] to ensure that only proteins with crystal structures, a requirement for the electrostatic calculations, were identified. Trying the BLAST search with other homeodoma containing sequences yielded similar BLAST results. Sixty three protein sequences were returned and sequence alignment was performed with MAFFT [12]. Of the 63 proteins available, 15 proteins bound to DNA were chosen for electrostatic calculations and pKa determination (Table 1). The few DNA bound proteins not selected were due to missing residues being present in the structures, which the H++ server lacks the capability of processing.

| PDB ID | Protein                          | Source                  |
|--------|----------------------------------|-------------------------|
| 1AHD   | ANTENNAPEA HOMEODOMAIN-DNA COMPLEX | Billeter et al., 1993   |
| 1AKH   | MAT A1/ALPHA2/DNA TERNARY COMPLEX | Li et al., 1998         |
| 1OCT   | OCT-1                            | Klemm et al., 1994      |
| 1YRN   | MATA1/MATALPHA2 HOMEODOMAIN      | Li et al., 1995         |
| 1ZQ3   | BICOID HOMEODOMAIN               | Baird-Titus et al., 2006|
| 2HOs   | PHAGE-SELECTED HOMEODOMAIN       | Shokat et al., 2006     |
| 9ANT   | ANTENNAPEA HOMEODOMAIN-DNA COMPLEX | Fraenkel & Pabo, 1998   |
| 1FJL   | DROSOPHILA PAIRED PROTEIN        | Wilson et al., 1995     |
| 1NK3   | VND/NK-2 HOMEODOMAIN/DNA COMPLEX | Gruschus et al., 1997   |
| 1DU0   | ENGRAILED HOMEODOMAIN            | Grant et al., 2000      |
| 1LE8   | MATA1/MATalpha2-3A HETERODIMER   | Ke et al., 2002         |
| 1HDD   | ENGRAILED HOMEODOMAIN            | Kissinger et al., 1990  |
| 2H1K   | Pdx1 HOMEODOMAIN                 | Longo et al., 2007      |
| 1PUF   | HoxA9 HOMEODOMAIN                | Laronde-Leblanc & Wolberger, 2003|
| 1LFU   | PBX HOMEODOMAIN                  | Sprules et al., 2003    |

Based on some preliminary findings, non-DNA bound homeodomains were not considered accurate electrostatic representations since chloride ion coordination has been demonstrated to be crucial to the unbound forms of the proteins. The explicit placement of these ions is not available in the structural data and the ions are thus not feasible to accurately account for their effects in the pKa calculations. The importance of negative charges to the electrostatics of homeodomains is further supported by calculations where the bound DNA ligands were removed, which resulted in a loss of perturbation among the DNA binding residues.

### 2.2 Electrostatic Calculations

pKa values were calculated using the H++ Web Server, available at http://biophysics.cs.vt.edu/H++ [13]. pKa predictions are begun by adding missing atoms and assigning partial charges to the uploaded protein structure using the parm99 force field and the AMBER molecular modeling package [14]. Positions of these added protons are optimized using 100 steps of conjugate gradient descent minimization and 500 steps of Molecular Dynamics simulation at 300K. The Poisson-Boltzmann equation in the program package MEAD is used to compute the free energies of the protonation microstates [15]. Titration curves and pKa values are then determined using the clustering approach described by Gilson [16]. Perturbations present on the Lysine and Arginine residues involved in DNA phosphate binding were computed by subtracting the standard pKa value of an unperturbed form of the residue (K=10.52, R=12.48) from the calculated value.

### 2.3 Plots and Statistics

Plots and statistics were computed using the program Graph Pad Prism version 4.01.

### 3 Results and Discussion

An amino acid sequence alignment for the homeodomain containing protein sequences was performed and the sequence region identified by Dragan et al. [2] as participating in coordinating with the DNA phosphates located within the alignment identified (Figure 1), as indicated by the residues aligned with the FCNRRQKEKR in the top sequence. This region contains several positions for which the presence of basic amino acid residues (K and R) is conserved. It is these basic residues which have been demonstrated to have interaction with the DNA phosphate groups.
Figure 1: BLAST alignment of homeodomain DNA binding sequences.
Figure 2: pKa perturbations associated with the DNA binding sequences of homeodomain proteins.
Calculation of the pKa values of the various homeodomain proteins revealed several notable similarities (Figure 2). One such similarity is that in all of the homeobox proteins examined, upward shifted pKa perturbations of at least 2 pK units are present in at least one of the lysine or arginine residues involved in coordination with the DNA phosphates. Given the high degree of residue conservation among this sequence region of homeobox proteins and the uniformity of the data computed it is hypothesized that these upwards shifted pKa perturbations are also likely conserved among all homeodomain proteins. This hypothesis is further supported by findings that indicate that pKa perturbations tend to only be present amongst residues that are highly conserved [8]. Moreover, the presence of these perturbations makes functional sense, since the upwards shift in pKa, some to pH levels that could never be reached in a physiological system, work to ensure that the lysine and arginine residues always remain in a protonated state, and hence always retain a positive charge. Given the negative charge of DNA the ability to maintain a strong positive charge would work to enhance the ability of these residues to bind DNA. Charge distributions such as this have been demonstrated to be conserved within protein families and for the CuZnSOD family it was suggested that a conserved charge distribution present around the catalytic sites of the enzymes served as a cationic funnel that helped steer the ligand into the catalytic binding site [5]. These findings further support the idea that a conserved electrostatic perturbation may exist in homeodomain structures to facilitate the binding of DNA.

What is also notable is that in all structures analyzed a significant (>2 pK units) perturbation was consistently present in the basic residue that aligned with the second arginine of the FCNRQRQKEKR base sequence. In many cases this perturbation was the strongest perturbation present among the DNA phosphate binding residues. One explanation for this is that for the 1NK3 and 9ANT structures it has been reported that out of all of the DNA phosphate binding residues, these residues were the ones that were buried most deeply into the bound DNA [2]. Given that pKa perturbations have been demonstrated to have an association to residue packing, where a tighter packing density has a higher magnitude of perturbation [17], the extensive burying of this residue can help to account for the magnitude of perturbation at this residue position. This suggestion is further supported by the finding that if the DNA is removed from the pKa calculations the observed perturbations diminish back into the range of values typically associated with basic residues. It is further hypothesized that the chloride ions which have been demonstrated to coordinate with the basic residues in the absence of DNA [2] help to maintain this perturbed state in the unbound structures as well, in order to facilitate the binding of DNA even further.

While there appears to be a conservation of the occurrence of an upwardly shifted pKa value for at least one basic residue in each structure it is notable that the positions and the magnitudes of each perturbation are not perfectly conserved. These differences are believed to be the result of the differences in the homeodomain sequences and along with geometrical considerations are believed to contribute to the sequence specificity associated with each type of homeodomain.

4 References

[1] Holland PW, Takahashi T (2005) The evolution of homeobox genes: implications for the study of brain development. Brain Res Bull 66: 484-90.
[2] Dragan AI, Li Z, Makeyeva EN, Milgotina EI, Liu Y, Crane-Robinson C, Privalov P (2006) Forces driving the binding of homeodomains to DNA. Biochemistry 45: 141-51.
[3] Kundrotas PJ, Alexov E (2006) Electrostatic properties of protein-protein complexes. Biophysical Journal 91: 1724-36.
[4] Frick DN, Rypma RS, Lam AM, Frenz CM (2004) Electrostatic analysis of the Hepatitis C Virus NS3 helicase reveals both active and allosteric site locations. Nucleic Acids Research 32: 5519-28.
[5] Frenz, CM (2008) The Role of Protein Electrostatics in Facilitating the Catalysis of DEAD-box Proteins. Proceedings of the 2008 International Conference on Bioinformatics and Computational Biology (BIOCOMP ’08), 708-712.
[6] Livesay DR, Jambeck P, Rojnuckarin A, Subramaniam S (2003) Conservation of electrostatic properties within enzyme families and superfamilies. Biochemistry 42: 3464-73.
[7] Livesay DR, La D (2005) The evolutionary origins and catalytic importance of conserved electrostatic networks within TIM-barrel proteins. Protein Sci. 14: 1158-70.
[8] Frenz CM. (2007) Interrelationship between protein electrostatics and evolution in HCV and HIV replicative proteins. Proceedings of the 2007 International Conference on Bioinformatics and Computational Biology (BIOCOMP ’07), 91-98.
[9] Kundrotas, P, Georgieva, P, Shosheva, A, Christova, A, Alexov, E (2007) Assessing the quality of the homology modeled 3D structures from electrostatic standpoint: test on bacterial nucleoside monophosphate kinase families, Journal of Bioinformatics and Computational Biology, 5: 693-715.
[10] Frenz, C.M. (2008) ESPSim: A JAVA Application for Calculating Electrostatic Potential Map Similarity Scores. Proceedings of the 2008 International Conference on Bioinformatics and Computational Biology (BIOCOMP ’08), 735-737.
[11] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-402.
[12] Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33: 511-8.
[13] Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Research 33: W368-71.

[14] Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. Journal of computational chemistry 26: 1668-88.

[15] Bashford, D (1997) An object oriented programming suite for electrostatic effects in biological molecules. Proceedings of the Scientific Computing in Object-Oriented Parallel Environments. Springer-Verlag.

[16] Gilson MK (1993) Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. Proteins 15: 266-82.

[17] Kim J, Mao J, Gunner MR (2005) Are acidic and basic groups in buried proteins predicted to be ionized? Journal of molecular biology 348: 1283-98.