

A method for in silico identification of SNAIL/SLUG DNA binding potentials to the E-box sequence using molecular dynamics and evolutionary conserved amino acids

Jeremy W. Prokop · Yuanjie Liu · Amy Milsted ·
Hongzhuang Peng · Frank J. Rauscher III

Received: 15 March 2013 / Accepted: 30 April 2013 / Published online: 25 May 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Binding of transcription factors to DNA is a dynamic process allowing for spatial- and sequence-specificity. Many methods for determination of DNA–protein structures do not allow for identification of dynamics of the search process, but provide only a single snapshot of the most stable binding. In order to better understand the dynamics of DNA binding as a protein encounters its cognate site, we have created a computer-based DNA scanning array macro that sequentially inserts a high affinity DNA consensus binding site at all possible locations in a predicted protein–DNA interface. We show, using short molecular dynamic simulations at each location in the interface, that energy minimized states and decreased movement of evolutionary conserved amino acids can be readily observed and used to predict the consensus binding site. The macro was applied to SNAIL class C2H2 zinc finger family proteins. The analysis suggests that (1) SNAIL binds to the E-box in multiple states during its encounter with its cognate site; (2) several different amino acids contribute to the E-box binding in each state; (3) the linear array of zinc fingers contributes differentially to overall folding and base-pair recognition; and (4) each finger may be specialized for stability and sequence specificity. Moreover, the macromolecular movement observed using this dynamic approach may allow the NH₂-terminal finger to bind without sequence specificity yet result in higher binding energy. This macro and overall approach could be applicable to many

evolutionary conserved transcription factor families and should help to better elucidate the varied mechanisms used for DNA sequence-specific binding.

Keywords SNAIL · Zinc finger recognition · E-box binding · Transcription factor binding · Protein-DNA dynamic

Introduction

How eukaryotic transcription factors find their cognate DNA binding sites in the promoters/enhancers of target genes is critical to understanding how these factors regulate gene expression and hence cellular phenotype. This function is defined in the context of chromatin and the large excess of non-specific DNA binding sites in the nucleus. Further complicating our understand are the expression patterns of transcription factors [1], disorder of the protein structure [2], and the complicated interactions each protein makes [3, 4] in a network that may influence DNA binding specificity. With many proteins we have the power to predict protein structure and function [5, 6]; however, less work has been done on the use of computer algorithms to predict how transcription factors find and regulate a high affinity recognition site. A short DNA sequence for transcription factor binding has a high probability of being ubiquitous in the genome (with hundreds of thousands of sites), yet only a fraction of these sites are occupied by the transcription factor. Even fewer of these occupied sites yield regulatory roles [7]. This suggests that binding is a dynamic process, and that recruitment and chromatin state (spatial components) play a large role. Transcription factors have the ability to “bind and search” in a scanning mechanism in the nucleus/DNA to identify DNA sequence, and the proteins

Electronic supplementary material The online version of this article (doi:10.1007/s00894-013-1876-y) contains supplementary material, which is available to authorized users.

J. W. Prokop (✉) · A. Milsted
Program of Integrated Bioscience, The University of Akron,
Akron, OH 44325, USA
e-mail: jwp7@zips.uakron.edu

Y. Liu · H. Peng · F. J. Rauscher III
The Wistar Institute, Philadelphia, PA 19104, USA

exist in multiple states of DNA interaction [8–11]. These factors all point to a very dynamic role for proteins in the ways in which they interact with and identify their consensus sequence—an aspect that is undervalued in solid state or solution structures of DNA–protein complexes.

In this study, we employed *in silico* biology to develop a new DNA scanning array program of a consensus DNA sequence through a model of a protein–DNA structure to identify dynamic binding properties. We have elected to use the zinc finger transcription factor SNAIL and its homologous family to address the potential mechanisms involved in DNA binding to the E-box sequence CAGGTG [12]. SNAIL contains four canonical zinc fingers—each independently folded into a helix—and an antiparallel beta sheet when complexed to zinc. SNAIL contains three C2H2 domains, which have two cysteines (C) and two histidines (H) that directly coordinate zinc. A fourth finger is an atypical C3H domain. While the cognate E-Box site bound by SNAIL has been known for more than 15 years, remarkably little is understood about how SNAIL interacts with DNA, which specific amino acids give it sequence specificity, and what the structure of the complex is. Herein, we propose the first molecular model for SNAIL–E-box binding using a molecular dynamics (MD)-based approach. This modeling approach will allow many future bench-top studies to be performed on the SNAIL–E-box complex, allowing a better understanding of how SNAIL is involved in cancer and cancer metastasis. Although SNAIL was chosen for this study, the approach can be applied to many other transcription factor families in the hope of elucidating mechanisms and networks of transcriptional activation and repression.

Methods

Modeling of SNAIL protein with DNA

A model of SNAIL containing 4 Zn fingers bound to DNA was created using PDB structures 1tf3 (TFIIIA protein, for zinc fingers 1 and 2 of SNAIL) and 2i13 (an artificial Zn finger, for Zn fingers 3 and 4 of SNAIL) based on the most homologous sequence of known zinc fingers in the pdb to each one of the four zinc fingers of SNAIL. To create the homology model of SNAIL, fingers 1 and 2 of 1tf3 were parsed and amino acids changed to those present in fingers 1 and 2 of SNAIL based on CLUSTAL [13] sequence alignment. These fingers (from 1tf3) were then aligned to the 2nd and 3rd fingers of the structure 2i13 using Mustang [14]. Zn fingers 1, 2, 3 and 6 of 2i13 were then deleted. The remaining Zn fingers (4 and 5) were mutated *in silico* to the amino acids matching fingers 3 and 4 of SNAIL as determined by alignments. A bond was created, joining the 2nd finger's carboxyl terminus (originally from structure

1tf3) to the amino terminus of finger 3 (from 2i13). This resulted in a structure contain four fingers, each with its own Zn ion, and a 20-mer DNA sequence (the original recognition sequence from 2i13). The linker domain between fingers 1 and 2 was corrected for SNAIL, which contains a slightly shortened linker domain compared to most Zn fingers. The final SNAIL sequence (Fig. 1), including the deletion of two amino acids from structure 1tf3, was energy minimized to relieve bond stress. All atoms were then freed and energy minimizations performed multiple times using AMBER03 force field [15] with 0.997 g/mL water. ConSurf analysis [16] for SNAIL, SNAIL like, Slug and Scratch proteins was done using ClustalW [13] alignment with default settings and Bayesian method with the T92 model for phylogenetic analysis.

Creation and MD simulations of the DNA scanning array

The DNA sequence in the modeled protein–DNA structure above was changed four times to one of four DNA bases (A, T, C, or G) on the first DNA strand of the structure, with the complementary base placed on the opposite strand using the “6bp Scanning Array” macro (Supplemental file pages 11–68) in YASARA [17], so that all the DNA of one strand is the same. This creates a background with all possible DNA bases at each site (four backgrounds), which serves as a relative starting point for stability when scanning the consensus binding sequence. The macro then inserts the 6-bp consensus sequence (CAGGTG, used in this study) at multiple sites on each DNA strand with all four DNA base backgrounds, generating four YASARA scene files for each location. Changes to only the first lines of the macro are required for changing the sequence of the consensus site, to allow for the use with other proteins. At each site the “md and analysis for scanning array” macro (supplemental file pages 69–121) was run. Details of the conditions for all MD simulations can be found in the macro scripts. The macro runs a 500 ps MD simulation, which then calculates multiple root mean squared deviations (RMSD). The score for each amino acid from the ConSurf analysis (ConSurf, on a scale of 1–9 with 9 being highly conserved and 1 having no conservation, Figure S1) was divided by the averaged heavy atom RMSD (movement of residue, M_R) of each amino acid in the four DNA base pair backgrounds at each location of the consensus (Fig. 2, Eq. 1) yielding X_{RZ} (amino acid R's conserved movement, at the Z position of the consensus sequence). This quantity was then averaged for all the amino acids in SNAIL (Fig. 2, Eq. 2) and normalized to the value from the control DNA sequences without a consensus (Fig. 2, Eq. 3) yielding the $ConSurf_{relZ}$ (The ConSurf total relative to the Z position of the consensus sequence). For the DNA base pair movement, each of the bases' RMSD from consensus sequence (M_{DZ}) was divided by the same location on the control DNA sequence (M_{D0}) to



Fig. 1 Sequence of SNAIL zinc fingers used to construct the model. The four zinc fingers are highlighted in yellow and the amino acids interacting with zinc shown in red. Numbers are based on human SNAIL

give the relative movement (M_{DNAZ}) (Fig. 2, Eq. 4). This was then averaged for the six bases of the consensus sequence (Fig. 2, Eq. 5) yielding DNA_{relz} . The compiled movement (R_z) was calculated by dividing the $ConSurf_{relz}$ by the DNA_{relz} (Fig. 2, Eq. 6) and the positions ± 1 (Fig. 2, Eq. 7) or ± 2 (Fig. 2, Eq. 8) of the R_z were added to address dynamic binding and transitions. The models were analyzed (z-score) using the YASARA2 force field and model quality (normality of dihedrals and packing).

Results

The overall approach was to create a program that could help illuminate the dynamic mechanism(s) by which a transcription factor binds a well-established DNA consensus sequence using a computational approach that could be advanced to high throughput methodology. This method should be able to confirm or identify the energy stable binding state, while also addressing the dynamic energy landscape in which that binding is directed. On the energy-minimized structure of SNAIL, Zn finger 1 was found around DNA base pairs C13–C16, finger 2 at C9–C12, finger 3 at C6–C8, and finger 4 at C3–C5. The combination of the two macros used to study the Zn finger SNAIL binding to the E-box sequence CAGGTG, allowed for placement of the consensus sequence at each location throughout the DNA strand and studied the energetics of the DNA-protein interaction at each location, with the consensus sequence (C of CAGGTG) starting at C1 (C1–C6),

C2 (C2–C7),..., C15 (C15–C20) or on the D strand of DNA. Potential energies of the modeled SNAIL protein from the multiple consensus locations showed few sites (D32 and D33) to have lower values, suggesting most of the SNAIL structures are stabilized and similar in all consensus locations (Figure S2A). Binding energy of SNAIL to DNA were not altered much for the multiple locations, with the highest value at the location C3 (Fig. 3). The relative conserved movement (R_z) was highest at the C4 location (1.12581), with a high value also found at C13 (1.114211, Fig. 4a). A value of 1 represents the R_z calculated for the DNA that did not contain a consensus sequence. The value of $Z_{\pm 1}$ (which allows a single metric for multiple binding locations to identify energy profiles with multiple sites in close proximity with high binding energies) enabled identification of two areas of probable binding, with the consensus sequence starting around C3–C5 and again around C12 to C13 (Fig. 4b). The $Z_{\pm 2}$ values are similar to the $Z_{\pm 1}$ values (Figure S2b,c), and therefore do not allow further identification of possible sites. Future experiments should allow optimization of the $Z_{\pm 1}$ value for quick identification of probable binding sites.

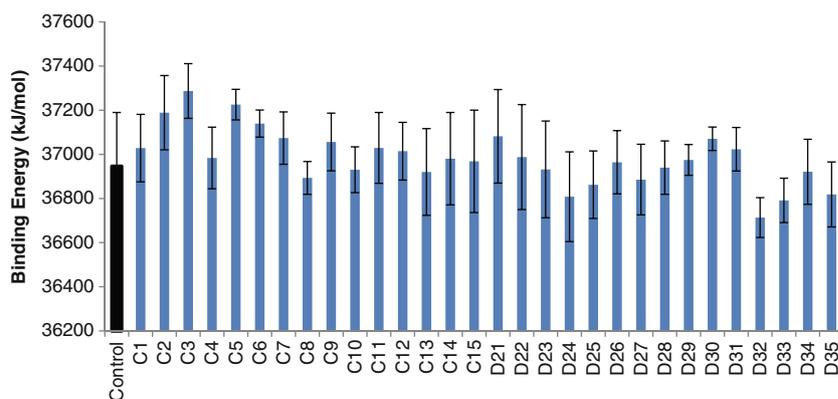
To address the specifics of why these two locations (C4 and C13) are highly favored, individual components of our calculations were studied. The DNA movement (DNA_{relz}) of the consensus sequence, which shows the stability of the consensus base pairs in simulations, was lowest when the consensus started at both C4 and the C13 locations (Figure S3), suggesting increased stability due to the protein environment at that location. Next, the X_{RZ} scores were used to address the

Amino Acid Movement	DNA Movement	Compiled Movement
1). $X_{Rz} = \frac{ConsR}{M_R}$	4). $M_{DNAZ} = \frac{M_{DZ}}{M_{DO}}$	6). $R_z = \frac{ConSurf_{relz}}{DNA_{relz}}$
2). $\bar{X}_{Rz} = \frac{1}{n} \sum_{i=1}^n X_{Rz}$	5). $DNA_{relz} = \bar{M}_{DNAZ} = \frac{1}{n} \sum_{i=1}^n M_{DNAZ}$	7). $Z_{\pm 1} = R_{(Z-1)} + R_{(Z)} + R_{(Z+1)}$
3). $ConSurf_{relz} = \frac{\bar{X}_{Rz}}{X_{RO}}$		8). $Z_{\pm 2} = R_{(Z-2)} + R_{(Z-1)} + R_{(Z)} + R_{(Z+1)} + R_{(Z+2)}$

Fig. 2 Math equations used in the DNA scanning array analysis. $ConsR$ Conservation score of each residue (based on ConSurf results), M_R heavy atom RMSD of each residue, X_{Rz} conserved movement for amino acid R with consensus at the Z location, X_{RO} conserved movement for amino acid R with no consensus sequence, $ConSurf_{relz}$ relative conserved movement of the protein with the consensus at the Z location, M_{DZ} movement of one of the bases of the consensus site of

the DNA, M_{DO} movement of the base M_{DZ} in the background with no consensus sequence, M_{DNAZ} relative movement of the consensus DNA sequence., DNA_{relz} average movement relative to the background of the consensus sequence at each position (Z), R_z score for conserved movement of the protein and the DNA at position Z of the consensus sequence, $Z_{\pm 1(2)}$ additive scores for positions plus or minus one site (or 2) around the Z site of the consensus location

Fig. 3 Binding energy of SNAIL to DNA from the energy minimized structure for each background at various locations of the two strands (C1–C15 and D21–D35) as determined in YASARA with AMBER03 force field



amino acids contacting the consensus sequence at each of the two highly favored sites. The score was averaged for all of the consensus locations in the array for each amino acid (Fig. 5a) to understand the intrinsic variation of the assay. A difference from each location of the array to this average allows for identification of potential amino acid leading to specificity of binding and stability of the DNA base pair movement. For the consensus found with the E-box sequence at the C4 location, the largest variation was found at amino acids 246 and 247 with a difference greater than 2 (Fig. 5b). Two times the standard error of the average for these amino acids was only 0.322957 and 0.295983, suggesting a strong significance of the difference seen at the C4 location from all the other locations. Both these amino acids are found in the fourth zinc finger with a highly conserved Ser at 246 and an Arg at 247. The hydroxyl group on the Ser likely hydrogen bonds with the phosphate backbone, putting the Arg side chain in the perfect location to hydrogen bond with the first or second G of the

consensus sequence (CAGGTG, Fig. 6). The average structures for each of the backgrounds with the C4 location confirm the proper distance is maintained in all simulations (Figure S4a). Mutation of these two amino acids significantly altered the ability to shift the E-box sequence (Figure S4b). Tracking the score of Ser 246 and Arg 247 over the entire consensus locations shows that they are higher at locations flanking the E-box consensus at the C4 location (C1–C5) and no other consensus locations (Figure S4c).

Besides Ser 246 and Arg 247 of finger 4, several amino acids are found in the normal positions for DNA specificity of a zinc finger. These amino acids include Met 171 of finger 1, Arg 191/Trp 193 of finger 2, Ser 221 of finger 3, and Met 248/Ser 249 of finger 4. Met 171, Arg 191, and Met 248 have no increased X_{RZ} score for any of the consensus locations (Figure S5), thus suggesting they do not contribute to specificity of binding. Trp 193 has an increased X_{RZ} score around the E-box sequence at the C9 location with

Fig. 4a–d Calculated conserved movement scores. **a** Relative conserved movement (R_z) for each site of the consensus on the C DNA strand (left) or the D strand (right). **b** Adding the R_z to the consensus shifted to the left or right by one shows two highly conserved favorable binding sites on the C strand (left) while none on the D strand (strand). A value of 1 for **a** or 3 for **b** would be the score of the control

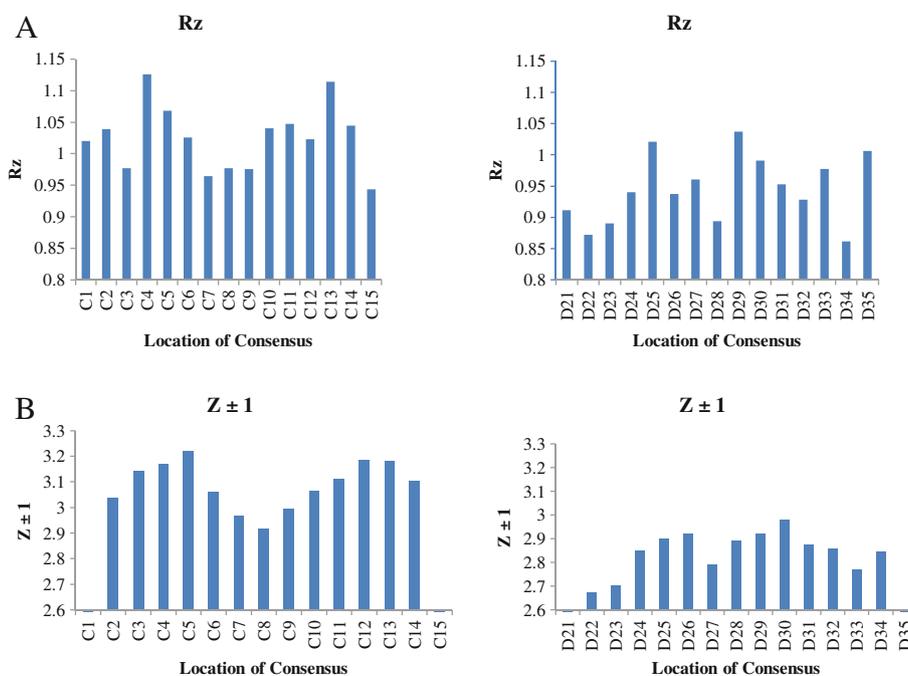
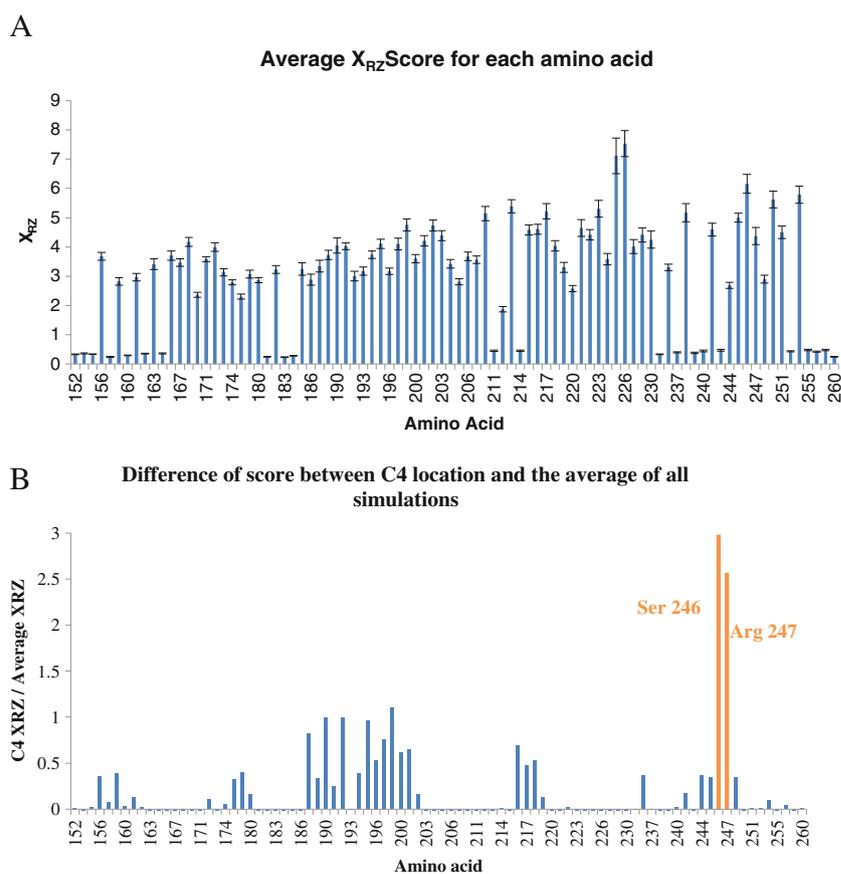


Fig. 5a,b X_{RZ} score for each amino acid. **a** Average X_{RZ} for each amino acid over all locations of the DNA consensus sequence; *error bars* plus or minus the standard error times two of the 30 locations of the consensus as well as the no consensus control. **b** X_{RZ} score for the C4 location of the consensus minus the average for each amino acid shown in **a**. The largest differences were seen at amino acids 246 and 247 (*orange*)



additional elevated levels for C8, C11 and C12. These locations put the E-box between fingers 2 and 3. Ser 221 had an elevated X_{RZ} score at E-box location C2 and D35 while Ser 249 showed elevated values at C14. As the E-box locations C12–C13 yielded the second highest $Z \pm 1$ score, the difference between each consensus X_{RZ} score and the control X_{RZ} score was observed. None of the amino acids around finger one or two (amino acids 156–202) contributed to the elevation of the X_{RZ} score (Figure S6). However, using heavy atom RMSD values for these consensus locations showed that the linker region between fingers 1 and fingers 2 (amino acids 176–180) were lower and therefore more stable in DNA interaction, with consensus sites placed starting at C12, C13, and C14 (Figure S7, this linker region is not conserved in the SNAIL like and Scratch proteins and therefore received no scoring in the X_{RZ} scores based on ConSurf results). In summary, it appears that critical contact residues for sequence specificity are Ser 246 and Arg 247, with additional contributions from the first linker domain and amino acids Trp 193, Ser 221, and Ser 249. The dynamic approach taken here suggests a multiple state binding for SNAIL.

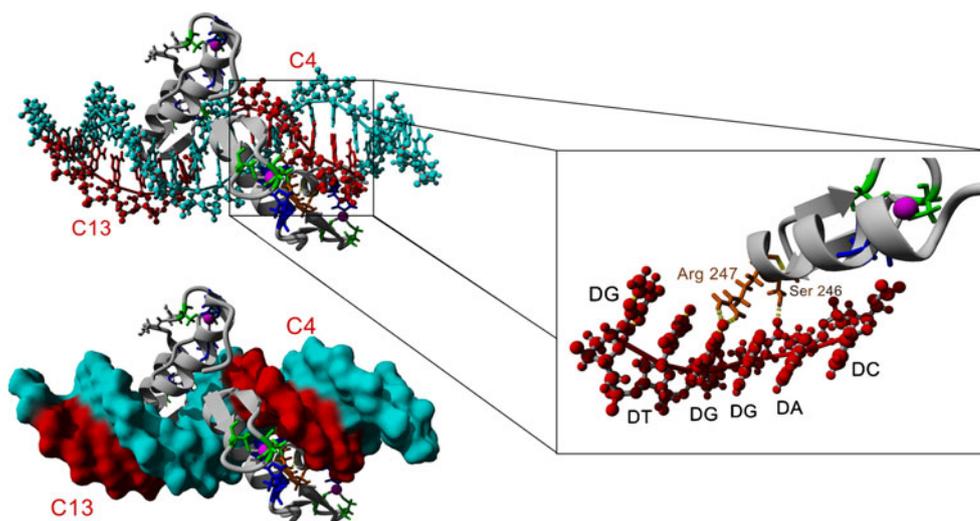
Discussion

Here we introduce a program that should be useful in localizing DNA–protein contacts in a cognate transcription

factor–DNA complex. Many structure determination methods are based on static, non-biological conditions, yet biology functions in dynamic processes in aqueous environments. Thus, new approaches are needed to address how proteins bind to DNA. Recent evidence has supported a role of zinc fingers binding and shuffling on DNA at a nanosecond timescale in an asymmetrical role [11]. Many solid state structures lack this dynamic component, thus there is a clear need for tools to help determine these dynamic processes. In silico experiments provide low-cost preliminary data generation relative to expensive NMR and protein purification, allowing scientists to screen a larger dataset of proteins for their potential role in these dynamic processes. By using either known or modeled structures of proteins interacting with DNA, it is possible to elucidate dynamic binding mechanisms.

SNAIL is involved in the epithelial-mesenchymal transition (EMT) involved in breast [18] and prostate [19] cancers. This involvement in cancer regulation is through an E-box sequence with a consensus of six base pairs (CAGGTG). The canonical model shows that each zinc finger domain recognizes three DNA base pairs [20]. SNAIL contains four highly conserved fingers, suggesting that only some of the fingers are involved in recognition of the E-box sequence. Until now it has been unclear which fingers contact which bases in the E-box consensus. We employed an in silico approach, creating a

Fig. 6 C4 and C13 consensus locations on the DNA. *Top* DNA in ball and stick form; *bottom* DNA molecular surface. Highlighted in the *box* is the location of the Arg 247 interacting with the DNA consensus sequence of the C4 location while Ser 246 interacts with the phosphate backbone allowing for stability of Arg 247



consensus sequence array using MD simulations and evolutionary conservation to determine a likely mechanism of binding. Our initial model before the scanning array yielded a z-score of -0.997 and with the consensus location starting at C4 of -0.891 . Z-score values between 0 and -2 are considered fair, and calculations of the z-scores for the original structures are -1.261 for 1tf3 and -0.433 for 2i13. Although more subtle variations in the structure may exist than those of our model for SNAIL, it provided an opportunity to test the macro created in this paper while providing significant hypothesis generation that can be tested in benchtop experiments for the SNAIL system. Our final models of SNAIL bound to DNA provide similar results (potential and binding energies) to other Zn fingers complexed to DNA and thus provide a strong starting point for threading the E-box consensus sequence through the DNA of our structure. Further comparisons of SNAIL structure to other Zn finger proteins binding different consensus sequences and other non-Zn finger proteins binding to the E-box sequence will be detailed in subsequent publications. In addition, the role of amino acid 246 and 247 have been confirmed through use of bench top methods.

Although the details of this method are still being actively modified for use with other proteins, the length of MD simulation (500 ps) in this manuscript appears to allow for stability of the protein in simulation, as can be seen in the representative plots for both energy and carbon alpha RMSD of the C4 location (Figure S8). This time was initially determined as RMSD values stabilized around 100 ps of SNAIL interaction with random DNA, allowing for full 400 ps stabilized simulations. We suggest that this amount of time optimizes detection of stability of critical amino acids, while reducing the computational requirements of performing hundreds of MD simulations for longer periods of time. Each protein–DNA complex will need to be adjusted and validated for the length simulation needed to allow stability of the complex.

The results from this novel *in silico* approach suggest that binding likely occurs through the fourth finger, with conserved contacts created by amino acids Ser 246 and Arg 247. Additional contacts on fingers two and three maintain protein shuffling around this consensus sequence. As this small shuffle of SNAIL on the DNA E-box sequence takes place, finger one may stabilize and bind tighter, with minimal to no sequence

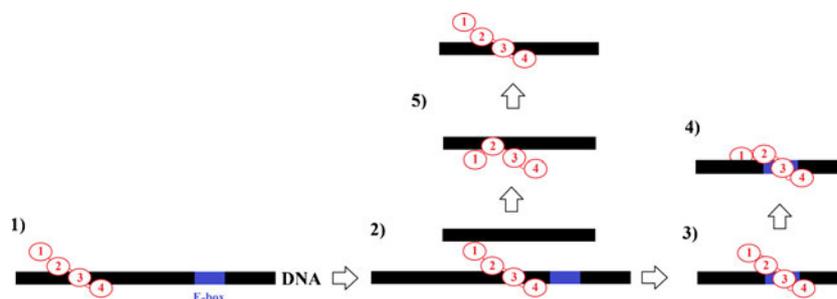


Fig. 7 Potential DNA binding mechanism of SNAIL family. 1 Fingers 2–3 of SNAIL form a weak complex with DNA with no sequence specificity. 2 SNAIL moves on the DNA. 3 SNAIL binds to the E-box sequence and stabilizes. 4 Decreased dynamics allows finger 1 to

stabilize and tightly bind the E-box sequence. 5 With finger one unbound, SNAIL may translocate to another DNA strand similar to mechanisms seen in Egr-1

specificity, through electrostatic interactions of the polar basic amino acids of SNAIL with the phosphate backbone of the DNA. As the linker domain between fingers one and two is smaller and atypical of most other zinc finger proteins, it may require more time to stabilize with the phosphate backbone. This mechanism would allow the SNAIL protein to scan the DNA and, when the DNA consensus identified, to reduce its movement to a shutter around the site and have tighter binding through stabilization of finger one (Fig. 7). Interactions of this linker domain with 14-3-3 [21] may alter the stability and binding affinity of SNAIL to the E-box sequence. Similar mechanisms may also exist to those previously shown for Egr-1 [11] allowing finger one of SNAIL to bind to another strand of DNA, translocating the other fingers to another DNA strand.

The results of this study provide the first molecular model for binding of SNAIL class zinc finger proteins to the E-box—a DNA sequence of high biological and disease relevance. This model may serve as a tool in understanding cancer progression and metastasis, allowing for drug design to specific amino acids. The data generated using this *in silico* DNA scanning array can be used to suggest which amino acids to mutate for benchtop analysis. At minimal cost (excluding time to run simulations and building computers), this allows screening of potential mechanisms that are conserved in a family of transcription factors. When combined with mutagenesis, DNA gel shift assays, NMR and other molecular/biochemical techniques, it has the potential to add strongly to our understanding of protein–DNA interactions in a dynamic, rather than static, process.

Conclusions

Use of the *in silico* DNA scanning array presented here has elucidated a potential mechanism for SNAIL–E-box sequence specificity. A multiple state binding mechanism appears likely, in which the first of four zinc fingers on SNAIL may form tight complexes when fingers 2–4 complex to the E-box. The use of these methods can be applied to all transcription factor families with hopes of further explaining DNA sequence specificity.

Acknowledgments J.W.P. is funded by the Ohio Board of Regents, American Heart Predoctoral fellowship, and The University of Akron. Work in the Rauscher laboratory is supported by National Institutes of Health grants CA129833, CA010815, CA163761, DOD-BCRP W81XWH-11-1-0494, The Samuel Waxman Cancer Research Foundation, SUSAN B. KOMEN FOR THE CURE and The Noreen O'Neill Foundation for Melanoma Research. Y.L. is supported by National Cancer Institute Training Grant: T32 CA09171.

References

- Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440:242–245
- Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45:6873–6888
- Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA* 100:12123–12128
- Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5:101–113
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protocols* 5:725–738
- Baker D, Sali A (2012) Protein structure prediction and structural genomics. *Science* 201(294):93–96
- Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmänn R et al (2012) NA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc Natl Acad Sci USA* 109:21330–21335
- Misteli T (2001) Protein dynamics: implications for nuclear architecture and gene expression. *Science* 291:843–847
- Catez F, Lim JH, Hock R, Postnikov YV, Bustin M (2003) HMGN dynamics and chromatin function. *Biochem Cell Biol* 81:113–122
- Shav-Tal Y, Darzacq X, Singer RH (2006) Gene expression within a dynamic nuclear landscape. *EMBO J* 25:3469–3479
- Zandarashvili L, Vuzman D, Esadze A, Takayama Y, Sahu D, Levy Y, Iwahara J (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc Natl Acad Sci USA* 109:E1724–E1732
- Nieto MA (2002) The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Biol* 3:155–166
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
- Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: a multiple structural alignment algorithm. *Proteins* 64:559–574
- Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W et al (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38:W529–W533
- Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* 57:678–683
- Côme C, Magnino F, Bibeau F, De Santa BP, Becker KF, Theollet C, Savagner P (2006) Snail and slug play distinct roles during breast carcinoma progression. *Clin Cancer Res* 12:5395–5402
- Smith BN, Otero-Marrah VA (2012) The role of Snail in prostate cancer. *Cell Adh Migr* 6:433–441
- Pavletich NP, Pabo CO (1991) Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* 252:809–817
- Hou Z, Peng H, White DE, Wang P, Lieberman PM, Halazonetis T, Rauscher FJ 3rd (2010) 14-3-3 binding sites in the snail protein are essential for snail-mediated transcriptional repression and epithelial–mesenchymal differentiation. *Cancer Res* 70:4385–4393