Structure Meets Function: Quantum Prediction of Arsenic-Binding Zinc Finger Targets in DNA Repair Pathways

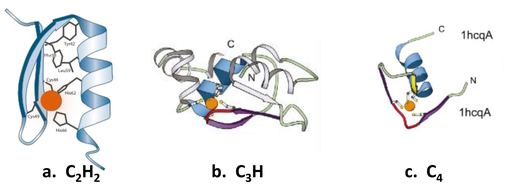
Qufei Gu,[1] Daniel Felker,[2] Daniel Chee,[3] Samuel H. Smith,[4] Karen Cooper,[5] Laurie G. Hudson,[5,6] and Susan R. Atlas[1,6]

Departments of [1]Physics and Astronomy, [2]Biology, and [3]Computer Science; [4]UNM School of Medicine;
[5]UNM College of Pharmacy; and [6]UNM Cancer Center, University of New Mexico, Albuquerque, NM 87131

Abstract

Arsenic toxicity deriving from environmental factors represents an urgent international health issue. Acting alone, arsenic is an established human carcinogen, and it can further amplify the carcinogenicity of other DNA damaging agents, e.g. UV radiation [1], thus acting as a co-carcinogen. One proposed mechanism for the carcinogenic and co-carcinogenic actions of arsenic is inhibition of DNA repair. This mechanism has therefore been suggested as a strategy for augmenting chemotherapeutic regimens. This exciting avenue of clinical research is under active investigation in several cancer drug trials. Two DNA repair proteins, PARP-1 and XPA, have been found to be direct arsenic targets based on arsenic interaction with their C3H and C4 zinc finger (ZNF) domains [2]. Using a combined bioinformatics [3] and structural [4] data analysis approach, we have developed a multiscale machine learning framework for quantitatively predicting relative arsenic binding affinities of DNA repair zinc fingers based on a combination of amino acid sequence and local atomic structure. Our approach utilizes first-principles quantum mechanical structure refinement of metal-atom binding motifs derived from experimental structures deposited in the Protein Database (PDB). We present quantitative predictions for a set of experimentally-studied C2H2, C3H, and C4-type zinc finger proteins including PARP1, XPA, SP1, and LIG3, demonstrating excellent concordance between prediction and experiment. The methodology is readily generalizable to other protein families and pathways.

Introduction



a. C_2H_2 b. C_3H c. C_4

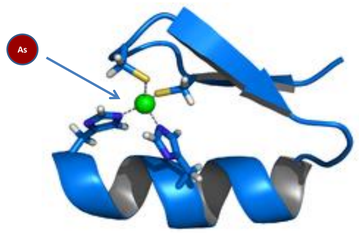
Figure 1: **Zinc fingers** are relatively small protein domains that involve a tetrahedral interaction between at least one zinc ion (orange sphere in figure) complexed with cysteine and histidine residues. (a) from Ref. [5]. (b),(c) from Ref. [4].

Arsenic and DNA repair

- Arsenic is an established human carcinogen.
- It can also amplify the carcinogenicity of other DNA damaging agents, e.g. UV radiation [1], thus acting as a *co-carcinogen*.
- One proposed mechanism for carcinogenic and co-carcinogenic actions of arsenic is inhibition of DNA repair.
- This also suggests a strategy to augment chemotherapeutic regimens via As inhibition of DNA repair.
- Two DNA repair proteins, PARP-1 and XPA, have been found to be direct arsenic targets based on arsenic interaction with their C_3H_1 and C_4 zinc finger (ZNF) domains [2].

A relationship between number of cysteines and arsenic binding affinity?

- ZNFs are highly prevalent in human proteins, where a single zinc atom is held in position by C_2H_2, C_4, or C_3H residues.
- Experimental results suggest that As interacts selectively with zinc finger motifs containing three or more cysteine residues [6].

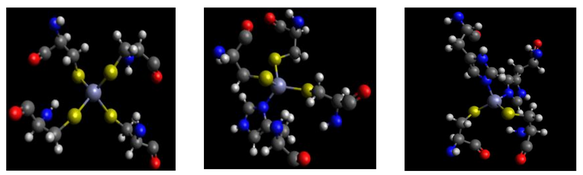


The challenge: How can we quantitatively model the binding affinity between zinc fingers and arsenic at the molecular level?

Methodology

Method

- Select "exemplars" from a filtered list of 126 ZNFs [3] identified via an amino acid pattern-based search methodology and annotated as DNA repair proteins.
- Use bioinformatic filtering of Protein Data Bank (PDB) [11] atomic structure files to exclude ZNFs that contain no Zn ligand.
- Manually filter the remaining 71 ZNFs to exclude those with > 1 zinc ligand.
- Extract PDB structure IDs for each candidate ZNF using automated scripting.
- Select candidate PDB X-ray or NMR structure files for quantum mechanical calculations based on analysis of reported resolution and date of publication.
- Extract corresponding atomic coordinates for the first-level amino acid shell surrounding the metal ion from PDB structure file.
- Use Avogadro visualization tool [12] to replace zinc atom with arsenic for each ZNF motif and perform geometry optimization.



1D4U (XPA,C4) 2L30 (PARP-1,C3H) 1SP1 (SP1,C2H2)

Figure 2: **Molecular structures of ZNFs** in different DNA repair proteins. Designation "2L30 (PARP-1, C3H)" indicates "PDB accession ID 2L30 for ZNF PARP-1, C3H-type motif";

Relative Binding Free Energies

$$\mathrm{As} + \mathrm{M} \xrightarrow{\Delta G_a} \mathrm{As{\cdot}M} \quad (1)$$

$$\mathrm{Zn} + \mathrm{M} \xrightarrow{\Delta G_b} \mathrm{Zn{\cdot}M} \quad (2)$$

$$\mathrm{As} + \mathrm{Zn{\cdot}M} \xrightarrow{\Delta\Delta G_{bind}} \mathrm{Zn} + \mathrm{As{\cdot}M} \quad (3)$$

Zn·M (M denotes the ZNF motif without metal atom) is the ZNF motif; As·M is the corresponding motif in which an arsenic atom replaces the zinc atom. Eqs. (1) and (2) represent the experimentally-accessible real processes with binding free energies ΔG_a and ΔG_b corresponding to As and Zn, respectively. Eq. (3) represents the displacement reaction between Zn and As.

The standard binding free energy is the free-energy difference between products and reactants:

$$\begin{aligned}\Delta G &= \sum G(\text{product}) - \sum G(\text{reactant}) \\ &= \sum(\epsilon_0 + G_{\text{corr}})_{\text{product}} - \sum(\epsilon_0 + G_{\text{corr}})_{reactant}\end{aligned} \quad (4)$$

G(X) denotes the Gibbs free energy (the sum of the total electronic energy ε_0 and the thermal correction G_{corr} of component X).

The relative binding free energy $\Delta\Delta G_{bind}$ in Eq. (3) can be defined as follows [7]:

$$\begin{aligned}\Delta\Delta G_{\text{bind}} &= G(\text{As·M}) + G(\text{Zn}) - G(\text{Zn·M}) - G(\text{As}) \\ &= [G(\text{As·M}) - G(\text{As}) - G(\text{M})] - [G(\text{Zn·M}) - G(\text{Zn}) - G(\text{M})] \\ &= \Delta G_a - \Delta G_b\end{aligned} \quad (5)$$

Thus, the relative binding selectivity of As versus Zn for a given protein is given by $\Delta\Delta G_{bind}$ [8] which is simply related to the free energy difference from Eqs. (1) and (2).

In practice, we evaluate the free energy contribution to the binding affinities by using the expression from the first line of Eq. (5).

Computational Details

- All calculations were performed using the Gaussian 09 [13] electronic structure code with the B3LYP exchange-correlation functional (B3LYP) [9,10] which gives a high average accuracy of order 2 kcal/mol for thermochemistry.
- Geometries for all exemplar proteins were first optimized on a 6-31+G* basis set, then on a 6-311+G* basis set, and finally on a 6-311++G** basis set.
- During the optimization process, bond length constraints were imposed to preserve the rigidity of the protein backbone and prevent unrealistic geometry changes.
- The bonds between Zn/As and its four directly-connected atoms in Figure 1 were all held fixed.
- Vibrational analysis was performed for each fully-optimized structure at the 6-311++G** level and imaginary frequencies were not found, indicating that the optimized structures correspond to energy minima.

Results & Future Work

Gene ID	Official full name	ZNF type	As affinity	% Zn loss 24 hr post- As [2,14]	PDB structure	Computed ΔΔG (DFT, kcal/mol)
PARP1	poly (ADP-ribose) polymerase 1	C3H	High	89.60	2L30	-37.809946
PARP1	poly (ADP-ribose) polymerase 2	C3H	High		1V9X	-71.996043
PARP1	poly (ADP-ribose) polymerase 3	C4	High		2JVN	-66.735632
XPA	xeroderma pigmentosum, complementation group A	C4	High	74.87	1XPA	-61.148037
TP53	tumor protein p53	C3H	Intermed.	53.51	2FEJ	-63.669113
LIG3	ligase III, DNA, ATP-dependent	C3H	Intermed.	60.00	1UW0	-42.424033
APTX	aprataxin	C2H2	Low	10.61	2KQE	30.088452
SP1	Sp1 transcription factor	C2H2	Low	5.84	1SP1	24.073773
POLH	polymerase (DNA directed)	C2H2			2I5O	17.986931
TRIM28	transcription intermediary factor 1-beta	One C3H + One C4			2RO1	-62.543869
MDM2	MDM2 proto-oncogene, E3 ubiquitin protein ligase	Two C2H2 + Two C4			2HDP	27.820632

Future Work

- Complex ZNF structures (C4HC3): RING, PHD, FYVE.
- Characterization of ZNF domain structures.
- Protein with multiple ZNF domains.
- Iterative refinement of As binding affinity predictions via comparison with experiment.
- Application of QM/MM methods to study ZNF domains in full protein structural context.

References

[1] Ding, W, *et al*. (2009). *Inhibition of poly(ADP-ribose) polymerase-1 by arsenite interferes with repair of oxidative DNA damage*. J. Biol.Chem. 284:6809-6817.
[2] Zhou, X, *et al*. (2011). *Arsenite interacts selectively with zinc finger proteins containing C3H1 or C4 motifs*. J. Biol. Chem. 286:22855-22863.
[3] Smith, SH (2011). *A Novel Sequence-Based Approach for Identification of Zinc Finger Motifs Involved in Arsenic-Induced Zinc Release*. B.S. honors thesis (biochemistry), University of New Mexico.
[4] Krishna, S, Majumdar, I, and Grishin, NV (2002). *Structural classification of zinc fingers*. Nucl. Acids Res. 31:532-550.
[5] Klug, A (2010). *The discovery of zinc fingers and their applications in gene regulation and genome manipulation*. In: Ann. Rev. Biochem. 79:213-231.
[6] Kitchin, KT and Wallace, K (2005). *Arsenite binding to synthetic peptides based on the Zn finger region and the estrogen binding region of the human estrogen receptor-alpha*. Toxicol. Appl. Pharmacol. 206:66-72.
[7] Tembe, BL and McCammon, JA (1984). *Ligand-receptor interactions*, Computers & Chemistry 8:281-283.
[8] Wang, L (2012). *Molecular Dynamics - Studies of Synthetic and Biological Macromolecules*, Intech.
[9] Becke, AD (1993). *Density-functional thermochemistry. III. The role of exact exchange*. J. Chem. Phys. 98:5648-5652.
[10] Lee, C, Yang, W and Parr, RG (1988). *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*. Phys. Rev. 37:785-789 .
[11] Bernstein, FC, *et al*. (1977) *The Protein Data Bank: a computer-based archival file for macromolecular structures*. M. J. Mol. Biol. 122:535-542.
[12] Hanwell, MD, *et al*. (2012) *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform*. J. Cheminform. 4:17-34.
[13] Gaussian 09 Revision D.01, Frisch MJ, *et al*. (2009) Gaussian, Inc., Wallingford CT.
[14] Hudson LG, *et al.* (2014), unpublished.

Acknowledgements: This work was supported by NIH/NIEHS grant # R21-ES021499. We are grateful to the UNM Cancer Center Shared Resource for Biostatistics and Bioinformatics (M Murphy, SM Wilson) for technical support, and the UNM Center for Advanced Research Computing for the computational resources used in this work.