

**Discovery, structure-function characterization and assessment of
polyoxometalates as modulators of the DNA binding activity of
the Sox-HMG family of transcription factors**

KAMESH NARASIMHAN

NATIONALUNIVERSITY OF SINGAPORE

2012

**Discovery, structure-function characterization and assessment of
polyoxometalates as modulators of the DNA binding activity of
the Sox-HMG family of transcription factors**

KAMESH NARASIMHAN

**(M.S BIOTECHNOLOGY, INDIAN INSTITUTE OF TECHNOLOGY, MADRAS,
INDIA)**

**A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

**DEPARTMENT OF BIOLOGICAL SCIENCES
NATIONAL UNIVERSITY OF SINGAPORE**

2012

ACKNOWLEDGEMENT

Research at the post-graduate level in the field of natural sciences is as much a journey of self-discovery as it is about deciphering the mechanisms that govern observed phenomena. The success of a research dissertation, to a large extent depends on the scientific question that is being asked and on what outcome would be deemed as a success. In graduate schools, research projects that are amenable to being recognized as a complete success are rare, while research projects that place a premium on autonomy and independent thinking are even rarer. I was immensely fortunate to have the independence and mentorship in equal measure to follow through on a scientific question of my choosing at the Genome Institute of Singapore (GIS).

Firstly, I would like to thank my supervisor **Dr. Prasanna R. Kolatkar** (PK) for readily agreeing to support me for a graduate studentship in his lab after administrative negotiations with **Dr. Philippa Melamed**. PK provided me independence and was absolutely willing to encourage me to pursue my ideas throughout the course of my research work. His support and suggestions have been valuable at many different stages of my graduate work. I reserve my most important acknowledgement to my mentor and co-supervisor **Dr. Ralf Jauch** (Ralf) for making an everlasting impact on my scientific temperament, inner confidence and maturity as a graduate student. Ralf's natural inquisitiveness, scientific ability, attitude, openness and his mode of interaction with scientific colleagues are qualities I hope I can imbibe to some extent in my life. I am especially grateful for his periods of guidance and patience through phases of my capriciousness, procrastination and whims. While Ralf was first and foremost a mentor at work, off-work I will remember him as a brother and friend for a life-time. I am grateful for his encouragement, suggestions, critical comments, for helping me progress in my scientific career and finally for giving me the freedom to explore. I should thank my collaborators **Dr. Konstantin Pervushin**, NTU, **Li Yan** and **Shubadra Pillay** for introducing me to the rich and vast field of biomolecular NMR. **Shubadra** recorded TROSY-NMR measurements and along with **Li Yan** carried out the backbone assignment. I would like to thank **Dr. Christopher**

Wong for providing access to the high-throughput screening facilities in GIS. **Rizal** contributed in the development and execution of the fluorescence anisotropy based high-throughput screening. **Dr. Bernold Hasenknopf** provided a number of Dawson-POMs and contributed to the development of the project. I would also like to thank **Dr. Zsolt Bikadi** and **Dr. Eszter Hazai** for providing autodock executables for Sox2-POM studies and helping with the docking analysis. The friendship and collegiality that I enjoyed in PK's lab will stay as an integral part of my memory and experience in GIS. I am especially thankful to the post-docs in my lab **Dr. Liang Yu, Dr. Bala, Dr. Venthan and Dr. Pugal** for their support at many different stages of my research work. Core support and friendship especially from (in no particular order) **Nithya, Calista, Siew Hua, Serene and Saran** is something I will cherish for a long time to come. The conversations and shared jokes that I have had with them on topics ranging from, biology, physics and life in general are moments that have sustained my graduate life. Most certainly, I should also thank **Marie, Sizun** and the intern students who worked with me namely **Clara, Sriram, Bharath, Tonio and Saranya** for helping me at different stages of my graduate work and from whom I have learnt a lot. I was also extremely lucky to have friends from my undergraduate studies (in no particular order) **Vishnu, Karthik, Ayshwarya and Vignesh** who were at the same time with me in Singapore doing their PhD. Their presence and continued support were central in the initiation of my graduate life in Singapore. I would also like to thank **Sravanthy**, aunt **Chandrika**, uncle **Nagaraj, Arvind** and **Madhavi** for all their affection, love and support during my stay in Singapore. Nuclear and quintessential to my life and purpose is the unconditional love and affection I get from my parents and siblings. I am eternally thankful to them for their understanding and support towards my career in life of research and exploration.

Finally, I would like to acknowledge **NUS** and **DBS** for my graduate research scholarship. This work is supported by the Agency for Science, Technology and Research (A*STAR) and the Genome Institute of Singapore. The DTP, NIH kindly provided the mechanistic and challenge diversity libraries employed in the study.

TABLE OF CONTENTS

Title	Page No
ACKNOWLEDGEMENT	I
TABLE OF CONTENTS	III
ABSTRACT	IX
LIST OF FIGURES	X
LIST OF TABLES	XVII
ABBREVIATIONS	XVIII
CHAPTER 1 INTRODUCTION	1
Transcription factors and their modulation by small molecules	
1.1 Introduction to transcription factors	2
1.2 Diversity in transcription factor binding architectures	3
1.3 Transcription factors as attractive and challenging targets for small molecule modulation	4
13.1 Targeting the ligand-binding, activation and protein-protein interaction domain of transcription factors	5
1.3.2 Targeting the DNA binding domain of transcription factors	6
1.3.3 Alternative non-small molecule based strategies for targeting the DNA binding domain of transcription factors	10
1.4 The Sox-HMG family of transcription factors as attractive targets for small molecule modulation	11
1.4.1 Role of Sox family of proteins in stem-cell biology	13

1.4.2	Role of Sox family of proteins in mammalian cellular development	13
1.4.3	Role of Sox family of proteins in neural development	14
1.4.4	Role of Sox family of proteins in cancer	15
1.5	Overall Scope of the Research Project	16
1.5.1	Sox-HMG inhibitors as potential therapeutic agents against Cancer	16
1.5.2	Sox-HMG inhibitors as tools to direct differentiation for in vitro tissue engineering	16
1.5.3	Sox-HMG inhibitors as chemical reverse genetic tools	17
1.6	Sox2 as a prototypical candidate for high throughput screening to identify inhibitors of the Sox-HMG domain	19
1.7	High throughput screening techniques	19
1.7.1	Small Molecule Libraries	20
1.7.1.1	The Chembridge libraries	21
1.7.1.2	The Chemdiv libraries	21
1.7.1.3	MayBridge screen	21
1.7.1.4	The Natural Products library from MerLion	21
1.7.1.5	NCI Chemical libraries	22
1.7.1.5.1	NCI Challenge diversity set II	23
1.7.1.5.2	NCI mechanistic diversity set	23
1.7.2	The role of academic high-throughput screening in addressing unconventional biological targets	24
1.8	The chemistry of Polyoxometalates (POMs)	25
1.8.1	General structure of the Dawson-Polyoxometalate $[P_2M_{18}O_{62}]^{n-}$	26

1.8.2	General structure of the Keggin-Polyoxometalate $[PM_{12}O_{40}]^{n-}$	26
1.8.3	The stability of Polyoxometalates	28
1.8.4	Functionalization of Dawson polyoxometalates	29
1.8.5	Biological activities of polyoxometalates	31
1.8.5.1	Anti-viral activities of Polyoxometalates	31
1.8.5.2	Anti-tumor activities of Polyoxometalates	32
1.8.5.3	Polyoxometalates as competitive inhibitors of the DNA binding activity of HIV-1 RT and Rad51	32
1.8.5.4	Polyoxometalates as non-competitive inhibitors of CK2 and Kinesin	33
1.8.5.5	Polyoxometalates as inhibitors of HIV-1 protease, Neuraminidase and HDACs	33
1.8.6	Delivery of Polyoxometalates inside cells	34
1.9	Aims of the research project	35
CHAPTER 2	MATERIALS & METHODS	36
2.1	Protein expression and purification	37
2.1.1	Sox2-HMG domain expression and purification	37
2.1.2	Sox-Homologs, REST C2H2, FoxA1 and AP2 purification	37
2.1.3	Pax6-paired domain expression and purification	38
2.2	Annealing DNA duplexes	38
2.3	Fluorescence anisotropy measurements	39
2.3.1	High-throughput fluorescence anisotropy screening using the Sciclone ALH-3000 workstation	39

2.3.2	HTP fluorescence anisotropy screening data analysis	44
2.3.3.	IC ₅₀ determination	46
2.3.4	Selectivity of various POMs using residual DNA binding activity measurement	47
2.4.	Limited proteolysis	47
2.5	Thermofluor assay	48
2.6	NMR sample preparation	48
2.6.1	NMR spectroscopy and data processing	49
2.7	Docking study of the Dawson-POM [P ₂ Mo ₁₈ O ₆₂] ⁶⁻ with Sox2-HMG	49
2.8	Polyoxometalates	50
CHAPTER 3	RESULTS I	51
	Sox2-HMG: Primary high-throughput screening and secondary validation assays	
3.1	Primary screening and identification of a polyoxometalate hit	52
3.2	Active Dawson phosphomolybdate K ₆ [P ₂ Mo ₁₈ O ₆₂] species responsible for inhibition of Sox2-HMG DNA binding activity	56
3.3	Preliminary selectivity studies on the Dawson-POM K ₆ [P ₂ Mo ₁₈ O ₆₂]	58
3.4	The Dawson-POM K ₆ [P ₂ Mo ₁₈ O ₆₂] physically interacts with the Sox2-HMG domain	60
CHAPTER 4	RESULTS II	63
	Sox2-HMG K₆[P₂Mo₁₈O₆₂] interaction: Structure-Function Relationship	
4.1	Studies of Sox2 Dawson-POM K ₆ [P ₂ Mo ₁₈ O ₆₂] interaction by NMR	64

4.2	Preferential binding site of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ on the Sox2-HMG surface	68
CHAPTER 5	RESULTS III	75
	Selectivity of polyoxometalates	
5.1	Selectivity studies on the inhibition of Sox-HMG family of TFs by polyoxometalates	76
5.1.1	Unmodified Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ and Decavandate $H_3V_{10}O_{28}$ are relatively selective TF inhibitors	79
CHAPTER 6	CONCLUSION AND FUTURE DIRECTION	87
6.1	Summary of Results	88
6.2	Mechanism of Sox-HMG inhibition by the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$	89
6.3	Future experiments to test the mechanism of inhibition of Sox2-HMG by $K_6[P_2Mo_{18}O_{62}]$	91
6.4	Assessment of the selectivity of Dawson-polyoxometalates	92
6.5	Potential Strategies that can be adapted from polyoxometalate based inhibition chemistry to target the DBDs of Transcription factors	93
	REFERENCES	95

APPENDIX

APPENDIX A	The protein sequences used in TF binding experiments like fluorescence anisotropy and EMSA	109
APPENDIX B	Representative 12% SDS gel images of the purified Sox2-HMG and Pax6 protein used in the study	111
APPENDIX C	The DNA duplexes used in TF binding experiments like fluorescence anisotropy and EMSA	112
APPENDIX D	A saturated complex of 1nM DNA and 50 nM Sox2 was competed by addition of unlabeled CCND1 in the presence of varying concentrations of DMSO. The assay shows tolerance even at high DMSO concentrations (>10% DMSO)	113
APPENDIX E	Primary hits identified from the screening	114
APPENDIX F	8nM of Pax6 was added to 1nM of a fluorescently labeled Pax6 consensus DNA sequence. Addition of unlabeled Pax6 consensus DNA sequence (100nM DNA element) serves as the positive control for complete inhibition of the Pax6-fluorescein labeled DNA complex. Addition of 200nM of Dawson POM to the previously bound Pax6-DNA complex has no effect on Pax6-DNA binding	116
APPENDIX G	An expanded snapshot of the docked configuration of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ with Sox2. Hydrogen bonds and electrostatic interactions less than 3.5Å are shown as red dots. Residue numbering is based on the PDB structure 1GT0.	117
APPENDIX H	The inhibition of Pax6 by 15 different polyoxometalates studied using EMSA. DNA binding activity was estimated from maximally bound Pax6-DNA (no POM) and free DNA gel-shift intensities (Pax6 DNA alone).	118
APPENDIX I	P-value of two-tailed, unpaired T-test (assuming equal variance) on residual DNA binding activities of 15 TFs upon $K_6[P_2Mo_{18}O_{62}]$ treatment. P-values less than 0.05 were taken as being statistically significant	119

Discovery, structure-function characterization and assessment of polyoxometalates as modulators of the DNA binding activity of the Sox-HMG family of transcription factors

Abstract

Aberrant expression of transcription factors is a frequent cause of disease yet drugs that modulate transcription factor protein-DNA interactions are presently unavailable. To this end the chemical tractability of the DNA binding domain of the stem cell inducer and oncogene Sox2 was explored in a high-throughput fluorescence anisotropy screen. The screening revealed a Dawson polyoxometalate ($K_6[Mo_{18}O_{62}P_2]$) as a direct and nanomolar inhibitor of the DNA binding activity of Sox2. [$^{15}N, ^1H$]-Transverse relaxation optimized spectroscopy (TROSY) experiments coupled with docking studies suggest an interaction site of the Dawson polyoxometalate ($K_6[P_2Mo_{18}O_{62}]$) on the Sox2 surface that enabled the rationalization of its inhibitory activity. Detailed investigation on a panel of different transcription factors against an expanded set of various polyoxometalates revealed that the Keggin and Dawson class of polyoxometalates exhibit a marked dichotomy in their selectivity and inhibition potential of the Sox-HMG family of transcription factors. Dawson polyoxometalates modified with organic moieties were found to invariably amplify the inhibitory potency of the pristine “Dawson” scaffold against Sox-HMG members, while a commensurate change in selective discrimination of the HMG family members could not be observed. The functionalization effect of the Dawson scaffold in inhibiting the Sox-HMG family merits investigation in the future. Taken in its entirety, the polyoxometalates have expanded the repertoire of molecular scaffolds that render transcription factors chemically tractable and provide strategies for the development of drugs that modulate transcription factors.

LIST OF FIGURES

Figure	Chapter 1	Page No
1.1	Gallery of small molecule inhibitors of transcription factors namely Estrogen receptor alpha (Disulfide benzamide), STAT3 (Stattic,, galliellalactone), AIF (Aurin tricarboxylic acid), HIF-1 (Echinomycin), p53 (Nutlin 3) and a bZIP protein inhibitor (NSC 13778)	9
1.2	A) The HMG box of the Sox family of proteins is primarily a three-helix bundle that binds to the minor groove of the DNA inducing a $\sim 70^\circ$ bending as observed in the case of Sox2 (PDB:1GT0). B) An unrooted neighbor joining phylogenetic tree generated using MAFFT and visualized using splitstree showing the different groupings of representative human Sox-HMG domain sequences (1-3).	12
1.3	A) Conventional chemotherapeutic agents target the somatic cells that form the bulk of the tumor but do not target the CSCs which would form a small fraction of the tumor mass causing a relapse of the tumor. By selectively targeting CSCs using novel drugs that target aberrant transcription factors involved in cancer stem cell formation, it would be possible to treat aggressive tumors. B) Sox2 expression universally marks embryonic and neural progenitor stem cells. A hypothetical Sox2 inhibitor either alone or in combination with other small molecules may have applications in directed differentiation of stem cells into specific cell fates.	18
1.4	Ball and stick representation of some canonical polyoxometalate structures namely	27

- A) Dawson $[P_2M_{18}O_{62}]^{n-}$
- B) Keggin $[PM_{12}O_{40}]^{n-}$
- C) Preyssler $[MP_5W_{30}O_{110}]^{n-}$
- D) Lindqvist $[M_6O_{19}]^{n-}$

Where “M” is the transition metal atom and “n” is the number of counter-ions.

Figure	Chapter 2	Page No
---------------	------------------	----------------

- | | | |
|-----|---|----|
| 2.1 | The principle of the fluorescence anisotropy based screening is that the fluorescently labeled DNA emits mostly depolarized light owing to its lower molecular weight as compared to the protein-DNA complex that emits partially polarized light because of its higher molecular weight. In a small molecule screening set-up, the difference in the intensity of polarized light can be measured to determine the bound and unbound states of the labeled DNA element with Sox2-HMG in the presence of the small molecules. | 40 |
| 2.2 | Overview of the fluorescence anisotropy based high-throughput screening for inhibitors of Sox2-HMG DNA interaction | 43 |

Figure	Chapter 3	Page No
---------------	------------------	----------------

- | | | |
|-----|--|----|
| 3.1 | <ul style="list-style-type: none"> A) Binding isotherm of Sox2-HMG with 1nM (FAM) <i>CCND1</i>. Increasing concentrations of Sox2-HMG increases the fluorescence anisotropy indicating Sox2-HMG DNA complex formation. Data represents the average of 3 independent titrations for a given concentration of Sox2-HMG in each of the reaction volumes. B) 80 nM Sox2-HMG and 1nM (FAM)-<i>CCND1</i> complex was allowed to reach equilibrium. The saturated complex was displaced using competing unlabeled <i>CCND1</i>. | 53 |
|-----|--|----|

- 3.2 A) Assays are carried out in a 384-well microplate depicted schematically as a heatmap (color coded by anisotropy values. Red color indicates higher inhibition while yellow indicates relatively lesser inhibition). Compounds are added to each well, while the positive and negative controls were alternately added to the peripheral columns (column 2 and 23) (4). 55
- B) Duplicate screens (Screen1 and Screen2) of the challenge and mechanistic diversity libraries revealed a Z' factor larger than 0.6 indicating a sufficiently large signal window for robust hit identification.
- C) Z-scores from duplicate screens correlate well highlighting the reproducibility of the assay.
- D) Screening results are shown as histograms of composite Z-scores. Primary hits were defined as having a composite Z-score ≤ 3 and reproducibility < -0.98 .
- 3.3 A) Fluorescence anisotropy assay showing that degradation products of $K_6[P_2Mo_{18}O_{62}]$, namely, the phosphate ($[HPO_4]^{2-}$), the molybdate ($[MoO_4]^{2-}$), and the Keggin phosphomolybdate ($[PMo_{12}O_{40}]^{3-}$), do not disrupt a half-saturated Sox2-HMG DNA complex. Ball and stick representation of the Dawson and Keggin POMs. 57
- B) Small light gray spheres are oxygen atoms and the bigger dark spheres are transition metals like Mo and W. The central phosphate atoms are labeled.
- 3.4 A) Gel-shift assays using varying POM concentrations shows that $K_6[P_2Mo_{18}O_{62}]$ selectively inhibits Sox2-HMG with an IC_{50} value of 98.6 ± 22.1 nM. 59
- B) Representative EMSA experiment showing dose-dependent titrations of $K_6[P_2Mo_{18}O_{62}]$ with 40nM Sox2-HMG and 1nM *CCND1*

(~50-70%)fraction bound) and 0.5nM Pax6 and 1nM Pax6 DNA element (~50% fraction bound) reveals selective inhibition of the HMG-domain

- 3.5 A) Limited proteolysis reveals that interaction of Sox2 with $K_6[P_2Mo_{18}O_{62}]$ confers resistance to proteolytic digestion by trypsin. Sox2-HMG was incubated with trypsin in the presence (lanes 3-6) and absence (lanes 8-11) of the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$. Reactions were stopped after different time points and analyzed by 4-12% SDS-PAGE. Molecular weight markers are added in lanes 1 and 7. Lane 2 contains the Sox2-HMG incubated for 60min but not subjected to trypsin digestion.
- B) Thermal melting profiles of Sox2-HMG monitored in the presence of Sypro-orange with and without increasing concentrations of the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$.

Figure	Chapter 4	Page No
4.1	Superimposed spectra of two-dimensional TROSY spectra of free Sox2 (pink) and Sox2 bound to POM (blue). Each cross-peak represents a bonded N-H pair. The axes correspond to the chemical shifts of N and H atoms in ppm (parts per million). The peaks that undergo significant shifts upon complex formation namely Glu66, Asp69 and His42 are highlighted.	65
4.2	The weighted change in chemical shift perturbations ($\Delta\delta = [\Delta\delta^2H_N + (0.2\Delta\delta N)^2]^{1/2}$) obtained from the ^{15}N 1H TROSY experiments are mapped on the entire Sox2-HMG surface (1GT0). Residues which are significantly shifted are depicted. The colored spectrum bar displays the extent of NMR chemical shift perturbations in ppm. Unassigned residues are colored in gray	66
4.3	Changes in chemical shift ($\Delta\delta = [\Delta\delta^2H_N +$	67

$(0.2\Delta\delta N)^2]^{1/2}$) upon POM binding is plotted against the Sox2 amino acid sequence (numbered according to Sox2: 1GT0). Threshold windows indicating Significant (S), Moderate (M) and Low (L) chemical shift perturbations are depicted as straight lines. Green colored bars indicate residues which have been implicated in direct binding to POM based on docking studies. Residues that are unchanged in TROSY are indicated with an asterisk (*). Unassigned residues are colored in light blue and given arbitrary negative chemical shift values solely for data visualization purposes. Sox2-HMG residue sequences involved in DNA binding are colored in blue. Secondary structural elements like alpha helices are named and colored to distinguish whether they belong to the major or minor wing of the HMG domain.

- 4.4 A) The lowest energy Sox2-HMG-POM complex structure from the Autodock searches shows the POM positioned within a pocket in the minor wing of the Sox2-HMG structure. Lys4, Arg5, Arg15, His63 and His67 are potentially involved in electrostatic or hydrogen bond interactions. Glu66 can donate hydrogen bond in a protonated form. Leu59, Leu62, Met7 and Val3 contribute to shaping the hydrophobic cavity. The Sox2 structure is shown as cartoon and the interacting amino-acids are shown as sticks. Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ is also shown in stick representation. 70
- B) Comparison of this docked model with the Sox2-DNA complex X-ray structure (1GT0) reveals that binding of POM to this site would directly interfere with DNA binding due to charge repulsion. The Sox2-DNA complex structure is shown as cartoon and the interacting amino-acids of Sox2 is depicted as sticks. The Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ is shown in stick representation. 71
- 4.5 A) Solvent accessibility per residue of DNA free 73

Sox2-HMG structure (PDB:1GT0) is plotted against the Sox2-HMG sequence and bars colored to depict their corresponding chemical shift perturbation category (NA-backbone unassigned residue).

- B) Interaction surface of Sox2-HMG colored by PM6 partial charges without and with docked ligand (blue – positive, red – negative).

Figure	Chapter 5	Page No
5.1	The regioselective ($\alpha 1/\alpha 2$) organic side chains of tin substituted Dawson POMs used in the study.	77
5.2	A heatmap of the average residual DNA binding activity ('value' in %) of 15 different TFs against a panel of 15 POMs, clustered by their inhibition profiles (Red color indicates higher inhibition, while yellow color indicates relatively lesser inhibition). Keggin POMs exert lowered inhibition on the Sox-HMG members leading to the observation that the size of polyoxometalate is an important consideration in the inhibition of the Sox-HMG family. Inhibitor compounds are color coded according to their polyoxometalate class as indicated in Table 5.1 (Dawson, Keggin or other simpler POMs like decavandate and sodium metatungstate).	80
5.3	A 3D bar plot depiction of the selectivity study of 15 TFs against a panel of 15 POMs from three-five independent experiments	82
5.4	A 2D bar plot extract of figure 5.3 depicting the diverse and relatively selective inhibition effect of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ and KM633 ($H_3V_{10}O_{28}$) in inhibiting a panel of 15 TFs from three-five independent experiments	85
5.5	Multiple sequence alignment of the core HMG-domain of representative sox proteins reveals differences between the Group F	86

members (Sox7 and Sox18) and Sox2 in 6 out of 10 amino-acid positions proposed to be involved in $K_6[P_2Mo_{18}O_{62}]$ binding with Sox2. Residues potentially involved in $K_6[P_2Mo_{18}O_{62}]$ binding based on docking studies with Sox2 are indicated by red dots. Homologous Sox-HMG residue positions involved in POM binding that exhibit consistent differences between the Group F members (Sox7 and Sox18) and Sox2 across the sequence alignment are indicated by an arrow. Numbering is based on Sox2 structure from PDB: 1GT0.

Figure	Chapter 6	Page No
6.1	A schematic representation of Sox2-HMG bound DNA complex inhibited by the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$	90

LIST OF TABLES

Table	Chapter 5	Page No
5.1	Panel of 15 POMs screened for inhibition of DNA binding activity of 15 transcription factors	78
5.2	Residual DNA binding activity (in %) of 15 TFs against 15 different POMs from three-five independent experiments expressed as mean \pm standard deviation	81

ABBREVIATIONS

AEG syndrome	Anophthalmia esophageal genital syndrome
B1H	Bacterial one hybrid
bHLH	Basic helix loop helix
ChIP-seq	Chromatin immunoprecipitation sequencing
DC5	Delta crystallin 5
DIBA	Disulfide benzamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribo nucleic acid
EMSA	Electrophoretic mobility shift assay
ERalpha	Estrogen receptor alpha
ESC	Embryonic stem cells
FAM	Fluorescein amidite
FDA	Food and drug administration
FGF	Fibroblast growth factor
FP	Fluorescence polarization
FRET	Fluorescence resonance energy transfer
Hits-flip	High throughput sequencing fluorescent ligand interaction profiling
HIV-1RT	Human immuno deficiency virus reverse transcriptase
HMG	High mobility group
HTH	Helix turn helix
HTP	High throughput
HT-SELEX	High throughput systematic evolution of ligands by exponential

	enrichment
LEP	Liposome encapsulated polyoxometalate
MITOMI	Mechanically induced trapping of molecular interactions
MLSV	Murine leukemia simian virus
NAD	Nicotinamide adenine dinucleotide
NCI	National cancer institute
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
PBM	protein binding microarray
PCR	Polymerase chain reaction
PDB	protein data bank
POM	Polyoxometalate
PPM	Parts per million
RMSD	Root mean square deviation
SCOP	Structural classification of proteins
SiRNA	Silencing RNA
Sry	Sex determining region Y
TEV	Tobacco mosaic etched virus
TF	Transcription factor
TROSY	Transverse relaxation optimized spectroscopy

CHAPTER 1

INTRODUCTION

Transcription factors and their modulation by small molecules

The Earth is a mote of dust, suspended in a sunbeam, a very small stage in a vast cosmic arena. Think of the rivers of blood spilled by all those generals and emperors so that, in glory and triumph, they could become the momentary masters of one corner of this pixel, scarcely distinguishable from others. How frequent their misunderstandings, and how fervent their hatreds. Our posturings, our imagined self-importance, the delusion that we have of some privileged position in the universe, are all challenged by this point of pale blue light.¹

¹ Sagan, Carl, *Pale Blue Dot: A Vision of the Human Future in Space*, 1994

1.1 Introduction to transcription factors

Transcription refers to the process in which information encoded in genomic DNA is faithfully transcribed into a complementary RNA sequence by the enzyme RNA polymerase (5). Transcription factors (TFs) are proteins that bind to DNA regulatory sequences either upstream or downstream of the transcription start site and control the transcription level of genes in association with the RNA polymerase through a variety of macro-molecular interactions (6). Transcriptional regulation is a complex event especially in eukaryotes as that involves the assembly of multi-protein complexes on core/proximal promoter or upstream enhancer modules that requires the co-operative assembly of co-activators, mediator, general and sequence specific transcription factors complexes (7). The eukaryotic core promoter is typically a TATA-box of (~25% of eukaryotic genes have a TATA-box) the consensus sequence TATAAA that is found in the -25 region. RNA polymerase II and general transcription factors like TFIIA, TFIIB, TFIID, TFIIE and TFIIH constitute the basal transcriptional machinery responsible for transcription from the core promoter region(7). However, the level of transcription by RNAP II and the general TFs alone is usually low. In addition to regulation by core promoters, transcription can be regulated by enhancer sites that can be as far as a few kb or Mb from the transcription start site. Enhancers are part of the non-coding matter of the eukaryotic genome and insofar as in humans as many as 110,000 gene enhancer sequences have been identified(8). Enhancers exert spatial and temporal control over gene expression programs in specific tissues resulting in a regulated pattern of gene expression (8). Sequence specific transcription factors bind to the proximal promoter or enhancer module and serve to enhance the rate of the transcription of genes under its control. The specific recognition of *cis*-regulatory DNA elements by transcription factors is achieved by a multitude of factors like *in vivo* TF (transcription-factor)

concentration, the relative affinity of the TF towards its specific and non-specific sites, its cooperativity with other protein-complexes, accessibility of nucleosomal DNA and not the least, like aspects of the presence/absence of epigenetic marks such as DNA methylation. Access to transcriptional templates in eukaryotes is also contingent on the displacement of nucleosomes from the promoter region of genes by chromatin remodellers and histone acetyltransferases (9). Biophysically, the readout mechanisms of TF's can be thought to be a sum contribution of 'direct' and 'indirect' readouts where the direct recognition by TF's refers to major/minor groove base interactions characterized by hydrogen-bonding, hydrophobic and water-mediated interactions. By contrast, the indirect readout by TF's refers to the global and local shape readout of the DNA characterized by bends, kinks, the groove widths and electrostatic potentials(10). Transcription factor-DNA recognition interfaces are characterized by a complex, interdependent network of bonding interactions, subtle alterations of which lead to substantially diverse binding preferences (11-13). A number of high-throughput measurements of protein-DNA interactions have significantly contributed towards a better understanding of the interaction landscape of a number of TF DNA binding structural classes(14-17).

1.2 Diversity in transcription factor binding architectures

A census of the human genome reveals the presence of about 2000-3000 sequence specific DNA binding transcription factors (~10% of the genome) belonging to many different structural classes (18). The eukaryotic genomes exhibit a diversity of TF structural classes with an array of diverse folds that have evolved strategies for interactions with DNA in a sequence specific manner. Based on available PDB structures of protein-DNA complexes, the SCOP family database identifies 70 SCOP superfamilies of DNA binding domains (10,19). For the sake of convenience, Remo Rohs et.al classify these superfamilies into four major groups based on their

overall secondary structure into mainly alpha, mainly beta, mixed alpha/beta and other multi-domain proteins(10). Since some of the structure based classifications were arbitrary in classifying the mixed alpha/beta group of proteins, a recent classification pegs sequence-specific TFs into four major superclasses namely the *basic*, *zinc-coordinating*, *Helix-turn-Helix* and *β -scaffold* based on phylogenetic analysis(20). Sequence specific TFs which don't fall into the above three categories are referred to as the unclassified *Others* family(20). The largest metazoan transcription factor families are the C2H2 zinc fingers, the homeodomains, and the bHLH, while the plant kingdom is dominated by AP2, MADS box, WRKY and B3 families. In humans, the *basic* superfamily consists of TFs which includes notable families like the bZIP (53 members) and the bHLH (110 members). Zinc-coordinating superfamily has prominent members like the C2H2 zinc finger (largest TF family in humans with 600 members), nuclear receptors (C4 zinc fingers), and the GATA family of TFs in humans(20). The helix-turn helix family in humans is most notably characterized by the homeodomain class of proteins which form a major chunk that utilizes the HTH structure. Examples of families that make use of HTH structures are the Hox, POU, Fox, IRF, Ets, RFX, HSF and E2F. The *β -scaffold* superfamily consists of TFs families like p53, RHR, NF- κ B and the STAT family(20). *Other* human transcription factors which are conspicuous by their biological activities include the CBF/NF-Y and the simple tri-helical Sox-HMG family.

1.3 Transcription factors as attractive and challenging targets for small molecule modulation

Transcription factors constitute an attractive subset of proteins for therapeutic intervention by small molecules, as aberrant expression of transcription factors is involved in a number of diseased states (21,22). Approximately 10% of the best selling drugs approved by the FDA

(Food and drug administration) are known to target transcription factors(23). At the structural level, transcription factors are modular proteins with usually a transcriptional activation domain, an oligomerization domain and a DNA binding domain. Modulator compounds targeted against transcription factors would have to prevent them from binding to one of their two primary molecular targets: protein or DNA (21). Traditional methods of targeting proteins rely heavily on non-high-throughput-screening approaches like structure-based drug design, *in silico* screening studies and high-throughput *in-vitro* assays. Small molecules targeted against ligand-binding/co-factor binding/dimerization interfaces of transcription factors have been notably successful in the past(24). On the other hand the disruption of protein-DNA interfaces by targeting the DNA binding surfaces of the TFs poses a key challenge because of the highly electrostatic nature of protein-DNA interface, the lack of binding pockets, and the largely unstructured nature of the DNA binding domain in the absence of being bound to DNA(21,24). Targeting the DNA-binding surface of TFs can be a unique way of inhibition, in that it would not decouple the regulatory interactions of the other domains in the transcription factor. Such a targeting strategy cannot be accomplished by siRNA and morpholino based knock-down experiments. Several examples are abound in literature where small-molecule inhibitors have been successfully employed to physically interact and deactivate transcription factors by either targeting the activation domains, the ligand binding domains or the DNA binding domains (24-27).

1.3.1 Targeting the ligand binding, activation and protein-protein interaction domain of transcription factors using small molecules

The nuclear receptor family consisting of approximately 50 members is one of the most favorite targets of small molecule modulation as it intrinsically has a ligand binding activation domain(28). Notable nuclear receptor modulators are the selective estrogen receptor modulators

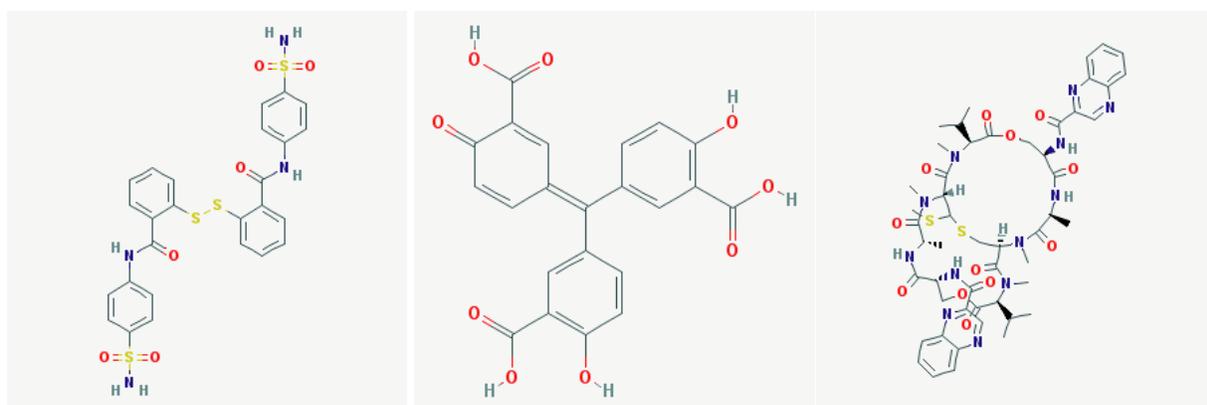
tamoxifen and fulvestrant, which are known to bring about tissue-specific ER agonistic or antagonistic activity(29). Small molecules like pyrimidine scaffolded molecules, benzenes and guanylylhydrazones also target the ligand binding domain of Estrogen receptor and are known to affect subsequent SRC (steroid receptor co-activator) interactions(30-32). In the case of thyroid receptors, vinyl-aryl ketones covalently react with their activation domain and abrogate their binding interaction with thyroid receptor co-activators (33). β -aminoketones were also found to inhibit steroid receptor co-activator SRC2 from binding to thyroid receptor (34). HNF4 α is a nuclear receptor that regulates hepatic lipid metabolism and is implicated in diabetes and atherosclerosis (35). Nitronaphofuran compounds were found to directly bind to HNF4 α activating transcription in HepG2C3A cells better than its natural ligand linoleic acid, leading to the development of newer routes for targeting these receptors in cells (36). Small molecule modulators have also been successfully employed to target protein homo/hetero-dimerization. Examples include small molecule modulators for c-Myc/Max, STAT3, p53 and Hap3p transcription factors (26,27,37-40). c-Myc is a bHLH protein involved in cell proliferation and differentiation suppression. Using high-throughput dimerization inhibitor assays, naphthols were identified as capable of disrupting c-Myc/Max interaction (40). Dichlorocarbazole 2 was also identified as capable of inhibiting c-Myc/Max in a yeast-hybrid library screen (41). STATs are TFs that dimerize via their phosphorylated SH2 domain and drive transcriptional responses that govern inflammation, apoptosis and immune response(42). Stattic disrupts STAT3 dimerization and translocation of STAT3 into the nucleus of HepG2 carcinoma cells rendering it very effective against STAT3 dependent cancer cell-lines (Figure 1.1)(39). p53 is a TF involved in cell-cycle control and in apoptosis pathways (43,44). P53 levels are tightly controlled by direct interaction with MDM2, a RING finger domain protein. MDM2 interaction with P53 could be

modulated by cis-imidazolines, benzodiazepine 11 and Nutlin3 (Figure 1.2) (37,38). All three of these p53 modulator compounds were also found to be active in cellular assays. Yeast HAP3 and its eukaryotic homologue NF-Y are known to bind as heterotrimers to consensus CCAAT DNA sequences to activate transcription. Using a novel high-throughput screening small molecule based micro-array dihydropyran 4 was successfully identified to bind to the yeast transcription factor Hap3 and its eukaryotic homologue NF-Y leading to a disruption in the formation of a Hap3 based hetero-trimeric complex(26).

1.3.2 Targeting the DNA binding domain of transcription factors using small molecules

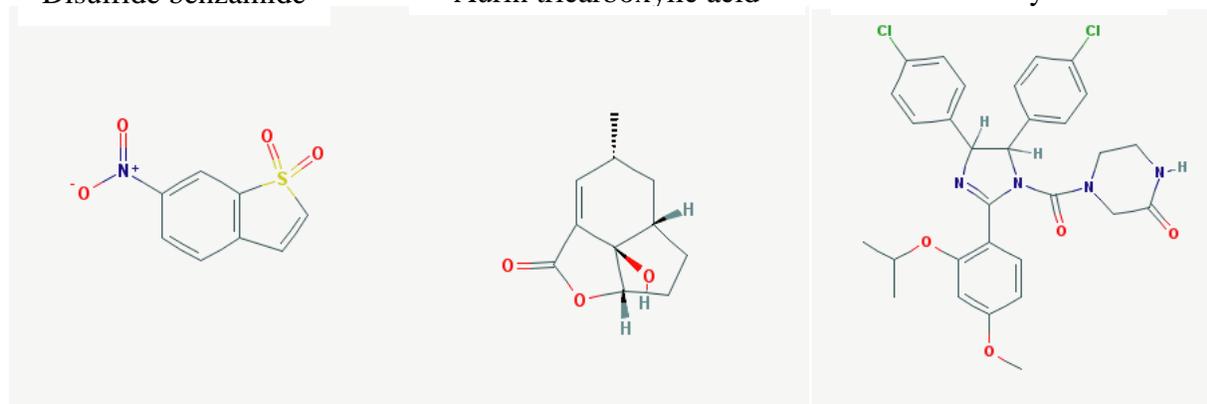
Unlike the ligand binding domains of TFs, the TF DNA binding domains lack “druggable” pockets and are largely unstructured in the absence of DNA. Additionally, the DNA binding domain is highly positively charged and unlike the ligand binding domain of engages in numerous contacts with DNA (24,45). High-throughput screening approaches have resulted in the identification of small molecules that disrupt specific protein-DNA complexes, by direct interaction with the DNA binding domains as in the case of transcription factors like ER- α , B-zip proteins, AIF, NF- κ B, HOXA13, HIF-1, STAT3 and HIVNCp7(39,46-52). Among DBDs, the zinc finger binding domains appear to be the best candidates for small molecule based inhibition of the DNA-binding activity. A platinated purine nucleobase compound was found to inhibit the HIV NCp7 zinc finger domains while the electrophilic disulfide benzamide and benzisothiazolone derivatives were found to be effective against the Estrogen receptor zinc fingers (Figure 1.2)(52,53). The small molecules that target these Zinc finger domains are usually electrophilic compounds that cause a Zn ejection from the DNA-binding domain resulting in a loss of the tertiary structure of the protein and subsequent abolishment the DNA

binding activity(52,53). The inhibition of the ER α zinc fingers by DIBA, an electrophilic compound gained much attention and provided a proof of principle for a new strategy to inhibit breast cancer at the DNA binding level, rather than the classical antagonism of estrogen binding(53). Recently, a high-throughput fluorescence anisotropy screen to identify inhibitors of the homeodomain protein Hoxa13 resulted in the discovery of a stereochemically complex lactam carboxamide that inhibited its DNA binding activity. The lactam carboxamide compound was prepared using diversity oriented synthesis and was also shown to be biologically active in Hoxa13 reporter assays(54). High-throughput fluorescence anisotropy assays have also been effective in identifying inhibitors of bZIP proteins and ER α (48). A high-throughput screening process aimed at identifying inhibitors that selectively bind to the DNA-binding interfaces of B-zip transcription factors like CREB, C/EBP β , VBP, and FOS|JUND was performed by Vinson C et.al., 2005 (48). The study identified inhibitors NSC13778 and NSC146443 that could inhibit all of the four B-zip transcription factors. Surprisingly NSC13778, was found to be able to discriminate between C/EBP β and C/EBP α in selectivity assays (Figure 1.1) (48). Similarly, a high-throughput screen of over 11,690 compounds identified a small molecule theophylline that inhibited ER α from binding to its cognate DNA binding elements. ChIP and reporter assays further confirmed that theophylline is a powerful inhibitor of the DNA binding activity of ER α (47). Using a novel technique based on photonic crystal biosensors, researchers were able to identify aurin tricarboxylic-acid as a low micromolar inhibitor of AIF, a non-specific chromatin binding protein(Figure 1.1)(49). Other examples include small molecules like galiellalactone and the platinum (IV) complex IS3 295 that directly block the DNA binding activity of STAT3 (Figure 1.1)(55,56). NF- κ B is a TF that plays a key role in immune system regulation in response to infection and has been linked to cancer and auto-immune diseases. Small molecules like



Disulfide benzamide

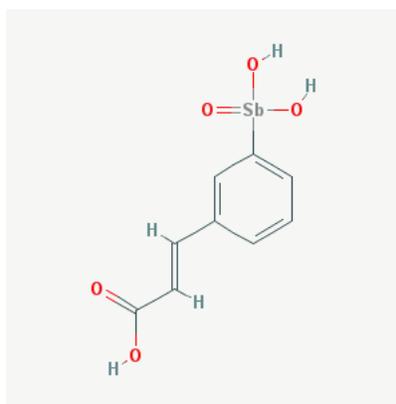
Aurin tricarboxylic acid



Stattic

Galliellalactone

Nutlin3



NSC 13778

Figure 1.1

Gallery of small molecule inhibitors of transcription factors namely Estrogen receptor alpha (Disulfide benzamide), STAT3 (Stattic, galliellalactone), AIF (Aurin tricarboxylic acid), HIF-1 (Echinomycin), p53 (Nutlin 3) and a bZIP protein inhibitor (NSC 13778)

sesquiterpene lactones and dimeric procynaidins have been proposed to inhibit NF- κ B DNA binding by direct interaction with the NF- κ B DNA binding surface (51,57). Hypoxia inducible factor is a bHLH TF that is critical for responses to changes in oxygen level, especially hypoxia and plays an important role in vascular development and cancer tumors. Echinomycin is a DNA intercalator that is known to disrupt HIF-1 DNA binding to DNA through PAS domain (Figure 1.1) (46,58). A comprehensive survey of the current literature reveals that the pharmacological potential of transcription factor DNA binding domains remains nascent and largely unexplored (21). Other than inhibitors of Zinc finger DNA binding domains, very few competitive inhibitors have been identified to target other structural classes of TF DNA binding domains.

1.3.3 Alternative non-small molecule based strategies for targeting the DNA binding domain of transcription factors

An alternative approach for targeting the DNA binding domains of transcription factors with a core inhibition strategy is by using DNA decoys or decoy-like aptamers. The aptamers utilize the natural propensity of transcription factors to bind nucleic acids and mimic the target sequence of the proteins thereby inhibiting transcription (59). Aptamers could either be single or double stranded RNA/DNA molecules although oligonucleotides with modified phosphorothioate DNA backbone are preferred as they are relatively resistant to nucleases within cells. Decoy aptamers have been used successfully to target a number of transcription factors like NF- κ B, E2F, STAT-3, c-Myc and Ets1 (60-64). A 14bp double stranded phosphorothioate aptamer Edifoligide, targeted against the cell-cycle transcription factor E2F, showed promise in initial clinical trials for treatment against smooth-muscle cell proliferation during surgical vein bypass but failed to show efficacy in Phase III trials(62). Similarly, a dsDNA decoy targeted against NF- κ B DNA binding activity from Averina is in Phase I/II clinical trial for the treatment of eczema (65).

1.4 The Sox-HMG family of transcription factors as attractive targets for small molecule modulation

The DNA binding activity of the Sox family of transcription factors is characterized by a ~80 residue high mobility group (HMG) domain family that bind to a consensus C(T/A)TTG(T/A)(T/A) motif (66,67). Its angular inner surface binds to the minor groove of the DNA and inserts a hydrophobic phenylalanine-methionine wedge into TT/AA DNA base pairs inducing a ~70° kink (Figure 1.2A) (68-71). It has been suggested that the induction of Sox specific kinks affects the gene regulatory outcome by initiating the assembly of specific regulatory complexes or enhanceosomes that crucially depends on the local shape of the DNA (72,73). The HMG domain consists of a three-helix bundle exhibiting an L-shaped structure composed of flexible major and minor wings that are subject to some structural rearrangements upon DNA binding (68,70,71,74). The functional and tissue specific gene expression programs of the Sox proteins are largely contingent on its differential partnership with other transcriptional regulators. Sox proteins are well known to physically interact with other transcription factors such as POU or Pax proteins contributing to the regulation of specific sets of genes involved in functions like eye lens development and stem cell pluripotency (70,75-79). Most of the HMG family members are key regulators of mammalian cell development patterning and are critical for cellular differentiation (79,80). Structurally, the Sox-HMG family belongs to a larger group of HMG proteins that could be classified into the HMG nucleosome binding family (HMGN), the HMG AT-hook family (HMGA) and the HMG-box family (HMGB) (81,82). Furthermore the HMGB family could be classified into the non-sequence specific and sequence-specific HMG families (82). The Sox-HMG family is a sequence specific HMG box and there are ≥ 20 members of the Sox family of TFs in vertebrates (1). A neighbor joining tree generated using

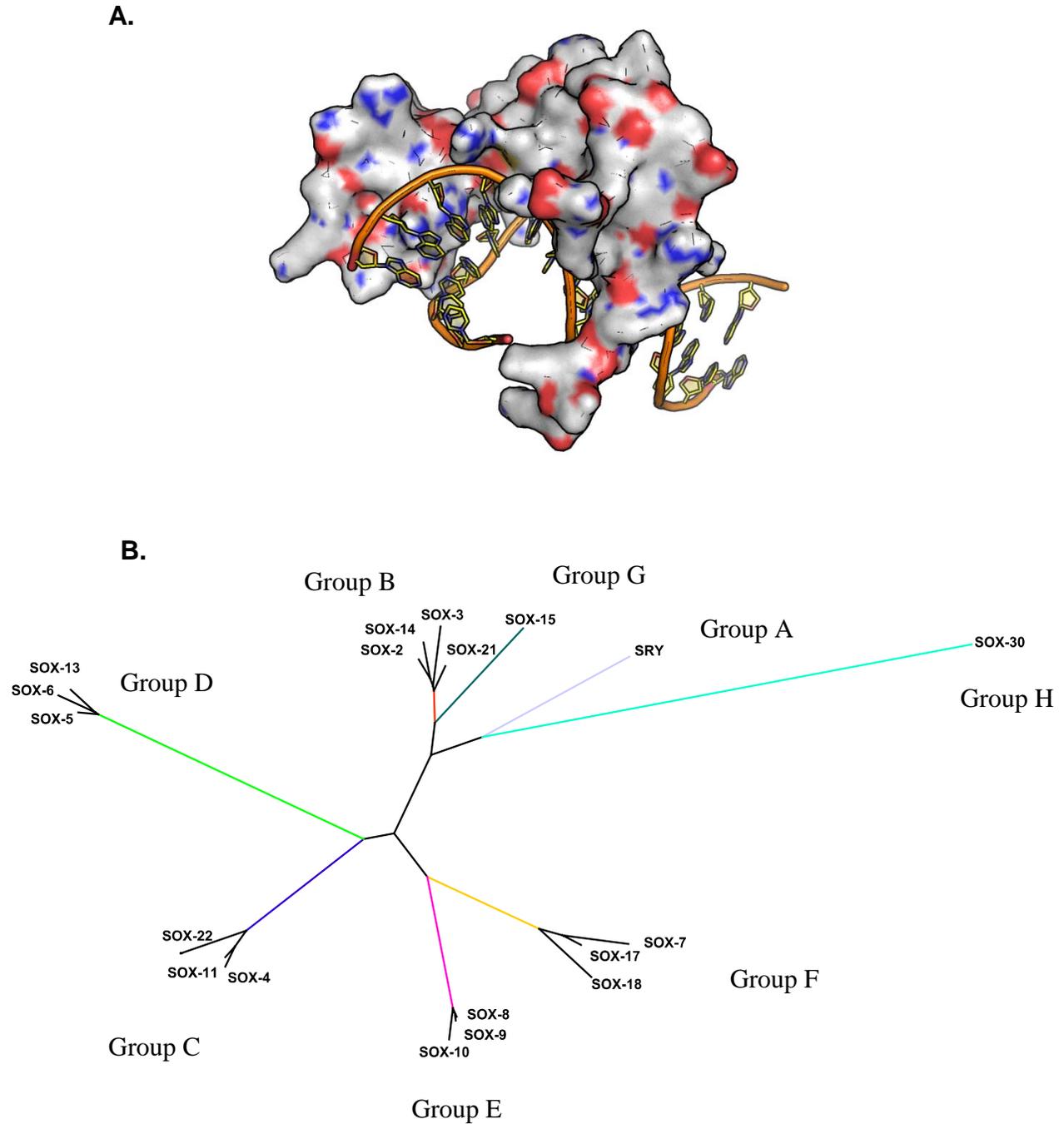


Figure 1.2

- A) The HMG box of the Sox family of proteins is primarily a three-helix bundle that binds to the minor groove of the DNA inducing a $\sim 70^\circ$ bending as observed in the case of Sox2 (PDB:1GT0).
- B) An unrooted neighbor joining phylogenetic tree generated using MAFFT and visualized using splitstree showing the different groupings of representative human Sox-HMG domain sequences (1-3).

The online multiple sequence alignment program MAFFT and visualized using splitstree shows the different phylogenetic groupings (Groups A-H) of the Sox-HMG domains. (Figure 1.2B) (1-3). The prototypical sox gene Sry, belongs to Group A, Group B1 consists of Sox 1, 2, 3, Group B2 consists of 14, 21, 25, Group C consists of Sox 4, 11, 12, 22, 24, Group D consists of Sox 5, 6,13, 23, Group E consists of Sox 8, 9, 10, Group F consists of Sox 7, 17, 18, Group G consists of Sox 15, 16, 20, Group H consists of Sox 30, Group I consists of Sox 31 and finally Group J consists of Sox 32, and Sox33(1).

1.4.1 Role of Sox family of proteins in stem-cell biology

The Group B1 Sox HMG Sox2 is normally expressed in pluripotent mammalian cells and plays a key role in the maintenance of cellular pluripotency (67). Sox2 is also required for self-renewal of embryonic stem (ES) cells (83). By featuring in a cocktail of four transcription factors required for generating induced pluripotent stem (iPS) cells Sox2 gained a lot prominence (84). Consistently, knockdown of Sox2 results in the loss of the undifferentiated state (85,86). Sox2 directly interacts with different members of the Oct family of proteins when bound to its DNA targets like *UTF1* and *Fgf4* and engages in a number of gene expression programs pertaining to stem-cell pluripotency and cell development(87,88). Rational mutation of a Sox2 interface residue with Oct4 based on the *Fgf4* crystal structure is known to perturb the ability of the interface-mutant Sox2 to form pluripotent stem cells (89).

1.4.2 Role of Sox family of proteins in mammalian cellular development

The Group A Sox HMG Sry is critical for testis development in mammals. Ectopic expression of this gene is known to induce testis formation in XX transgenic mice while its deletion or mutation has been known to cause female genitalia in XY humans (90-93). The GroupB1 member Sox2 is known to bind co-operatively with Pax6 to form a ternary complex on the DC5

enhancer eliciting eye lens placode formation while mutations in its Sox2-HMG domain have been associated with microphthalmia (77,94-97). Mutations in Group B2 Sox14 gene have been associated with limb development defects like Mobius Syndrome while Sox21 knockout mice display hair loss and is hypothesized to be responsible for hair-loss condition in humans (98,99). The Group F member Sox7 is involved in parietal endoderm differentiation and is known to be critical for the induction of Gata4 and Gata6 (100). Sox17 and Oct4 have been known to co-express in endoderm cells and is presumed to cooperatively bind on a “compressed” Sox-Oct motif (80,89). Sox18 acts as a molecular switch to induce differentiation of lymphatic endothelial cells by activating *Prox1* and mutations in the *Sox18* gene have been associated with lymphatic obstruction and vascular lesions (101,102). The Group G Sox-HMG member Sox15, plays a crucial role in skeletal muscle regeneration, while Sox9 is known to be critical for cartilage development. Mutations in the Sox9 gene have been known to cause campomelic dysplasia, a skeletal malformation syndrome (103,104).

1.4.3 Role of Sox family of proteins in neural biology

The Group B1 HMG Sox3 is known to be highly expressed in ventral diencephalon and mutations of Sox3 is known to have a significant effect on cognitive activities (105). Sox2 interacts with Brn2 on a nestin enhancer element and this partnership is known to play a key role in brain development (106). Sox1 is known to have roles in post-mitosis controlling neuronal cell-specific differentiation of ventral striatum neurons (107,108). The Group B2 Sox HMG proteins Sox14 and Sox21 have been known to antagonize Group B1 Sox-HMG proteins (109). For example, Sox21 promotes neuronal cell differentiation by competing with the Group B1 interaction partner like Pax6, while Group B1 proteins are known to be required for maintaining the undifferentiated state of neural cells (110,111). The Group C Sox-HMG representatives Sox4

and Sox11, are expressed in post mitotic neuro blasts resulting in activation of Tuj1 and MAP2 markers for neuronal differentiation (112). The Group D members Sox5 and Sox6 are known to antagonize the SoxE family in oligodendrocyte development by competing with SoxE proteins for the same binding sites as is observed in competition between Sox5 and Sox10 to bind to myelin gene promoters (113-115). The Group E Sox-HMG members Sox8 and Sox10 are known to have important roles in oligodendrocyte differentiation and is also known to be essential for neural crest and peripheral nervous system development in (116) (117,118).

1.4.4 Role of Sox family of proteins in Cancer

Elevated expression levels of the Group B1 Sox2 is known to have been associated with a large number of tumor types *in vivo* and it is hypothesized that the up regulation of Sox2 expression in carcinomas may have important pathological relevance (119,120). Sox2 has recently been implicated in the transcriptional regulation of the oncogene *CCND1* in breast cancer (121). In another study, it was found that patients with the milder monoclonal gammopathy whose immune system developed anti-Sox2 antibodies in the earlier stages, showed better prognosis than patients with a full fledged myeloma, who fail to develop spontaneous immunity to Sox2 (122). Sox2 has also been recently identified as a potential target for therapy in malignant glioma(123). The oncogenic potential of Sox2 received further recognition after elevated expression levels were detected in several tumors such as squamous cell carcinomas, lung cancer, gastric carcinoma, malignant glioma and in breast cancer (119,121,124) . Overexpression of the Group C Sox members namely Sox4 has been reported in adenoid cystic carcinomas and in breast cancer cell lines while Sox11 overexpression has been reported in anaplastic oligodendroglioma (125,126). These observations of the oncogenic potential of Sox proteins lends support to the “Cancer-stem cell” hypothesis that states that adult stem cells give rise to

cancer cells and that aberrant upregulation of Sox proteins promotes self-renewal, de-differentiation, proliferation and cell-survival reminiscent of their role in stem cell biology (127).

1.5 Overall scope of the research project

In the light of the importance of Sox-HMG proteins in stem cell, cancer and developmental biology, it can be envisaged that small molecule inhibitors of Sox-HMG proteins will have three areas of application: (i) as potential therapeutic agents against cancer (Figure 1.3A) (ii) as tools to direct differentiation for *in vitro* tissue engineering (Figure 1.3B) (iii) and as a useful chemical reverse-genetic tool to mimic mutational/developmental disorders.

1.5.1 Sox-HMG inhibitors as potential therapeutic agents against Cancer

Because of their special properties associated with self-renewal - some of the Sox transcription factors are also known to drive analogous gene expression programs in both pluripotent cells and tumor cells play oncogenic roles in the maintenance and propagation of germ cell tumors (128) (67,129). The current anti-tumour drugs in the market target metabolic pathways active in mature and differentiated cancer cells but not pathways unique to cancer stem cells. As a consequence while the differentiated and mature cancer cells are eliminated by the conventional anti-tumour drugs, the cancer stem cells may survive and self-renew leading to new tumours (127). Therefore it can be expected that inhibitors designed to target pathways unique to cancer stem-cells like the Sox-HMG mediated pathways could potentially hold the key to more effective cures in tumorigenesis (119).

1.5.2 Sox-HMG inhibitors as tools to direct differentiation for *in vitro* tissue engineering

Embryonic stem cells (ESCs) hold promise in the field of regenerative medicine and therefore targeting key transcription factors like the Sox-HMG family with small molecules has the

potential to control stem cell fate and in directing differentiation processes for many practical applications towards stem-cell based therapies (130) (131,132). For example, Sox2 is a key-factor responsible for the maintenance of the stem-cell like property of neural stem cells and there is a potential that Sox2 inhibitors either alone or in combination with other growth factors may be used to direct the neural stem cells *in-vitro* into special neural cell types like astrocytes or oligodendrocytes (86).

1.5.3 Sox-HMG inhibitors as chemical reverse genetic tools

Sox-HMG inhibitors will also serve as useful chemical reverse-genetic tools to mimic mutational disorders that affect the biochemical function of Sox proteins. For example, mutations in the Sox2-HMG domain that abolish DNA binding, have been associated with genetic disorders like Anophthalmia-Esophageal-Genital (AEG) syndrome (96). Thus, animal models can potentially be induced by Sox2-HMG inhibitors to mimic these genetic disorders in a cheap and reversible manner unlike RNAi based techniques or knock-out methods, which do not have stable dose response and lack reversibility.

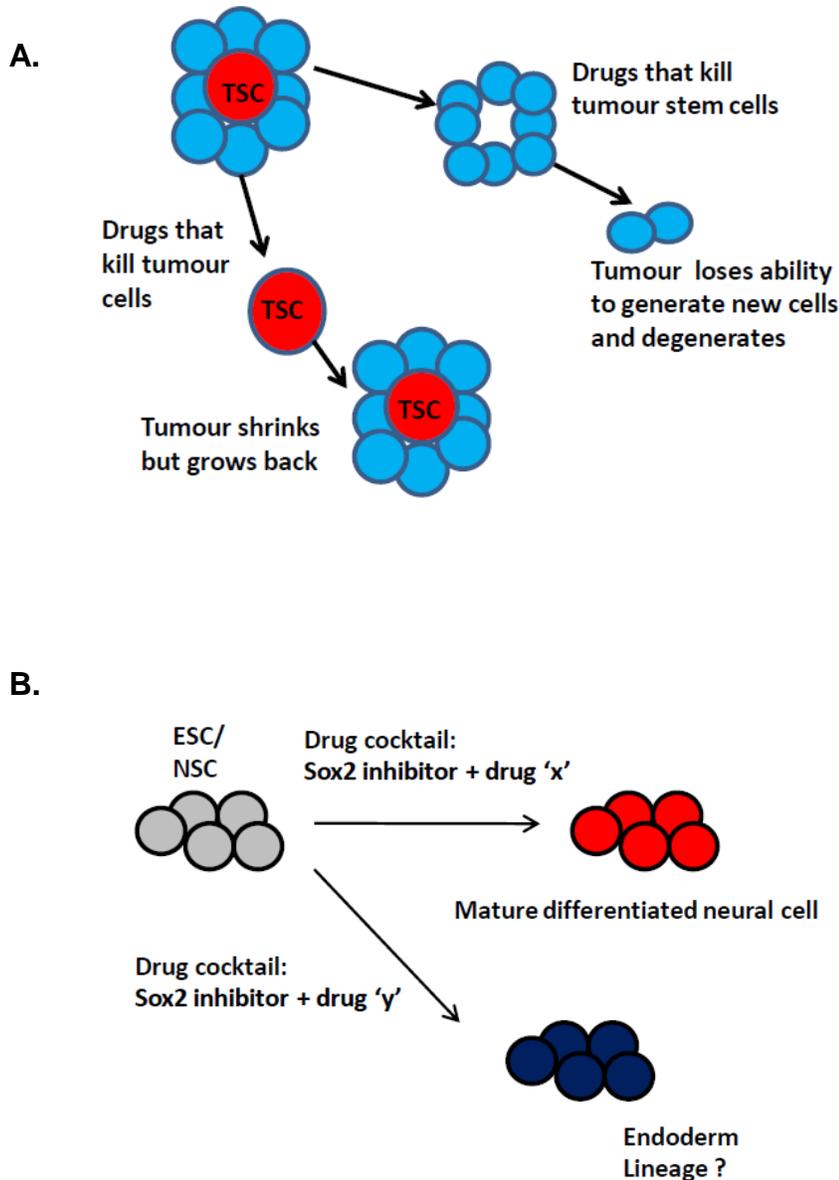


Figure 1.3

- A) Conventional chemotherapeutic agents target the somatic cells that form the bulk of the tumor but do not target the CSCs which would form a small fraction of the tumor mass causing a relapse of the tumor. By selectively targeting CSCs using novel drugs that target aberrant transcription factors involved in cancer stem cell formation, it would be possible to treat aggressive tumors.
- B) Sox2 expression universally marks embryonic and neural progenitor stem cells. A hypothetical Sox2 inhibitor either alone or in combination with other small molecules may have applications in directed differentiation of stem cells into specific cell fates.

1.6. Sox2 as a prototypical candidate for high throughput screening to identify inhibitors of the Sox-HMG domain

Sox2 is a prototypical representative of the Sox-HMG domain family that binds to a consensus (A/T) (A/T) CAA (A/T) G core and is approximately 40% identical to the rest of the family members (67,111). The DNA binding interface of Sox2 is predominantly cationic, extended, with no binding pockets and bends DNA at an angle of ~70 degrees comparable to other Sox proteins like Sox4 and Sox17 (13,70,71). Sox2 is known to play a critical role in the maintenance of stem cell pluripotency and self-renewal of ES cells(67,83) . Since Sox2 is known to confer the property of “Stemness” it can be conceived that Sox2 inhibitors may be used to enable directed differentiation of stem-cells into mature phenotypes. Sox2 has also been reported to have oncogenic potential and therefore tumors with deregulated expression of Sox proteins might be treatable with small molecule inhibitors (Figure 1.3A) (119,124).Overall, the biological properties of Sox2 in stem cell and cancer biology makes it desirable for small molecule modulation and the identification of a potential inhibitor of Sox2-HMG by high-throughput screening methods could in principle be extended to target other Sox-HMG proteins. In this context, a brief introduction to high-throughput screening techniques and small molecule libraries will be provided below.

1.7 High throughput screening techniques

HTS assays can either be biochemical or cell based and come in many formats (24, 96, 384 or 1536 wells). HTS can be either carried out on live cells to study certain processes like apoptosis or it can be a biochemical assay using purified proteins to target a specific protein (like a protein kinase) (133). Cell based HTS assays are usually reporter gene assays (luciferase, GFP), cell proliferation assays and secondary messenger assays (calcium ion, NADH concentration). Cell

based assays can also be high-content in nature (High-content screening) where live cells are analyzed for multiple parameters like multiple fluorescent reporters, phenotypical changes, cellular apoptosis, cytoskeletal rearrangements, nuclear DNA content etc., during small molecule screening. Unlike HTS which usually measures one or two parameters, high content screening yields more temporal, spatial information and generally analyses multiple parameters leading to more efficient primary hit optimization (133). Biochemical assays predominantly require volume miniaturization and involve read-out techniques like fluorescence, luminescence, radioactivity or UV absorption, depending on the scope and design of the assay. In particular luminescence based screens like the Alpha screen, scintillation proximity assays that make use of radioactivity and fluorescence based assays that utilize FRET, fluorescence anisotropy, fluorescence correlation spectroscopy, fluorescence intensity distribution analysis and time-resolved fluorescence are most popular as biochemical assays as fluorescence\luminescence based assays are typically sensitive even at low reagent concentrations and can be scaled up well for high throughput measurements (133).

1.7.1 Small Molecule Libraries

An important component of high-throughput screening is the choice and size of small molecule libraries. A number of small molecule libraries are commercially available that cover a wide range of pharmacological chemical space. The pharmacophore chemical space essentially refers to descriptors like hydrogen bond acceptor, hydrogen bond donor, hydrophobic, charged and aromatic groups that could be used to theoretically classify compounds(134). Compounds with very different structures might also have similar chemical descriptors and hence libraries can be constructed to cover different subsets of pharmacophore space (134). Some of the most recognized chemical library providers are the Chembridge corporation, Chemdiv Inc., the

Maybridge screen, the Natural products libraries from MerLion and the NCI repository of small molecule libraries.

1.7.1.1 The Chembridge libraries (<http://www.chembridge.com>)

The library has more than 500,000 compounds in its repositories although screeners typically prefer more mid-sized libraries like the ChemBridge DIVERSet, which have about 50,000 compounds that cover about 60-65% of the entire chembridge pharmacophore space. ChemBridge also offers investigators to pick smaller subsets of 5,000-10,000 compounds from the DIVERSet depending on the requirements of the biological target for screening.

1.7.1.2 The Chemdiv libraries (<http://www.chemdiv.com>)

The ChemDiv library is one of the largest libraries with a collection of over million compounds with about ten percent of the total collection dedicated to specific targets like kinase inhibitors, GPCRs, ion-channels and developmental pathway modulators. Overall, the ChemDiv library is represented by approximately 10,000 uniquely diverse scaffolds with as much as 700 compounds per scaffold population having greater than 90% purity by NMR.

1.7.1.3 MayBridge screen (<http://www.maybridge.com>)

The MayBridge collection of 60,000 compounds covers ~87% of the 400,000 pharmacophores defined by the world drug index and are known to obey Lipink's "rule of five" thereby generally demonstrating good absorption, distribution, metabolism and excretion profiles. MayBridge also offers a reduced version known as the preplated HitFinder with a collection of 16,000 compounds selected to represent the diversity of their 60,000 compound collection.

1.7.1.4 The Natural Products library from MerLion (<http://www.merlionpharma.com>)

The Singapore based MerLion Pharmaceuticals is known for its collection of about 1,800 purified natural compounds from Fungi, Plants and Actinomycetes, out of which about 300

compounds are completely new structures that are not available in any other natural products library. The MerLion natural products collection consists of compounds of the class like tannins, terpenes, steroids, alkaloids, polyketides, aminoacids, aromatics, carbohydrates, glycosides, flavonoids, peptides and polypyrroles.

1.7.1.5 NCI Chemical libraries (<http://dtp.nci.nih.gov>)

The NCI DTP repositories houses a uniquely diverse set of about 200,000 compounds, that are shipped annually to research investigators around the world free of cost except for shipping charges. The NCI also makes available four reduced set of libraries from their entire collection using different selection processes resulting in what is known as the approved oncology drugs set (97 compounds), the challenge diversity set II (1364 compounds), the mechanistic set (879 compounds) and the natural products set (120 compounds).

A survey of more than 2.6 million compounds from over 32 different chemical providers including the ChemBridge, ChemDiv, MayBridge and the NCI chemical libraries were carried out for assessing properties like drug-likeness, lead-likeness, fingerprint based diversity and frameworks (134). The assessment revealed that the NCI database is the top most representative of the diversity in chemical space covering about 59% of the chemical space of the whole database of 2.6 million compounds. Even in selection processes that rank for “lead-like” drugs, the NCI compound collection emerged as the top-most diverse database among all of the 32 chemical providers (134). The NCI library collection also have the additional incentive that these libraries could be procured free of cost.

Hence the reduced NCI library sets namely the NCI Challenge diversity set II and the NCI Mechanistic diversity set were chosen for ideally beginning the campaign screen. The NCI

libraries were also chosen to make an initial assessment of the success in terms of the number of “primary” hits that could be obtained given the chemical diversity of the library collection.

1.7.1.5.1 NCI Challenge diversity set II

The challenge diversity set consists of 1364 compounds derived from almost 200,000 compounds available with the DTP repository. The pharmacophoric space of these 200,000 compounds was reduced using the ChemX (Oxford Molecular group) and Catalyst (Accelrys Inc.) programs creating a diverse set of compounds that were amenable for structure-based hypotheses. The final set consisted of 1364 compounds that are more than 90% pure and consists of molecules that are relatively rigid, planar, having at most one chiral centre and does not contain undesirable pharmacophore groups like organometallics, polycyclic aromatic hydrocarbons and weakly bonded heteroatoms.

1.7.1.5.2 NCI Mechanistic diversity set

The mechanistic diversity set consists of 825 compounds derived from 37,836 open compounds known to have activity in NCI human tumor 60 cell line screens. In contrast to the structural diversity of the compounds that constitute the challenge library, the mechanistic diversity consists of compounds known to have a diverse growth inhibition pattern (broad GI50 range) on the human tumor 60 cell line screen. Compounds that exhibited activity in the human tumor 60 cell line were clustered using FASTCLUS resulting in 1272 clusters. 825 representative compounds for which sufficient material was available was chosen from each one the 1272 clusters to form the final mechanistic library set.

1.7.2 The role of academic high-throughput screening in addressing unconventional biological targets

The concept of high-throughput screening, once solely the forte of major pharmaceutical companies is increasingly becoming amenable for academic researchers and research institutes and this has led to a greater variety of targets being screened than would be in the pharmaceutical industry(135). Unlike pharmaceutical companies where the budgets, equipment and manpower are largely focused on targets that would fetch a billion dollar drug, academic high-throughput screening even with its limited resources is ushering in a slow and silent revolution, as the financial and research priorities are different (136). A number of broad research problems that addresses rare genetic diseases, basic cell physiology, difficult protein targets etc., can all be tackled with more freedom in an academic high-throughput screening setting. Academic screening unlike industrial screening has also resulted in an open source environment of sharing the results of high-throughput screening through academic publications or through databases like PubChem(137). Integration of small molecule structures with genomic, proteomic, crystallographic and high-throughput screening information in public databases like PubChem has enabled the academic community to identify criteria for selecting lead compounds with potential for further development.

In this context, it must be mentioned that an academically interesting class of inhibitors that arose from high throughput screening in the current study is the inorganic polyoxometalates. A Wells Dawson polyoxometalate was identified as a primary hit in the high-throughput screening for inhibitors of Sox-HMG DNA binding domain in this study. Hence, the chemistry and biological properties of polyoxometalates will be examined in detail below as this class of molecule is central to the current study.

1.8 The chemistry of polyoxometalates

Polyoxometalates are inorganic compounds built on a framework of oxyanions derived from transition metals belonging to Group 5 and 6 of the periodic table in their highest oxidation states (138). Examples of transition elements that form polyoxometalates are Vanadium, Molybdenum and Tungsten (138). The oxyanions are held together by oxygen atoms and enclose one or more central heteroatoms like phosphorus or silicon. Typically, polyoxometalates are nanometer sized aggregates, with a high negative charge density and a versatile structural architecture that is amenable to modification with organic functional groups. Polyoxometalates are synthesized from a condensation reaction of oxyanions in an acidified solution, resulting in the formation of a framework of oxyanion bridges that get repeated in a regular manner (5,139). The condensation reaction can be controlled by the choice of solvent used, pH, temperature, stoichiometry and counterions resulting in a number of different polyoxometalate structures.

For example, the Dawson polyoxometalate $K_6[P_2M_{18}O_{62}]$ can be synthesized by a condensation reaction of sodium molybdate in a phosphoric acid and potassium chloride solution (5).



(Sodium Molybdate) (Phosphoric acid)

(Dawson potassium phospho-molybdate)

Polyoxometalates find a number of applications in the field of nuclear waste treatment, (electro) catalysis, nanotechnology, material sciences and in medicine. Some common (though not limited to) structural polyoxometalate families of importance in the field of biomedicine are the 1) Keggin structure $[PM_{12}O_{40}]^{n-}$, 2) the wells-Dawson structure $[P_2M_{18}O_{62}]^{n-}$ 3) the preyssler structure $[MP_5W_{30}O_{110}]^{n-}$ and 4) the lindqvist structure $[M_6O_{19}]^{n-}$ where “M” is the transition metal atom and “n” is the number of ionic charges (Figure 1.4) (138). Many of the physical properties of polyoxometalates like redox potential, acidity, elemental composition, structure,

charge density and distribution is amenable to varying degrees of alteration making it attractive for biomedical applications. Since much of the current study, places a great emphasis on the Dawson polyoxometalates $[P_2M_{18}O_{62}]^{n-}$ and the Keggin polyoxometalates $[PM_{12}O_{40}]^{n-}$ their structural aspects will be examined in detail below.

1.8.1 General structure of the Dawson-Polyoxometalate $[P_2M_{18}O_{62}]^{n-}$

The Dawson Polyoxometalate is a prolate ellipsoid consisting of two cap centers, each made of three molybdenum or tungsten atoms forming a triad and two equatorial belt centers of six molybdenum or tungsten atoms each (Figure 1.4 A). Two internal caged phosphates co-ordinate the cap and belt centers. In total, there are 18 terminal and 44 bridging (36 two-, 6 three-, and 2-four coordinated) oxygens. Three isomeric rotational forms of the metal-oxygen unit are known to exist for the Dawson polyoxometalate namely α , β and γ .

1.8.2 General structure of the Keggin-Polyoxometalate $[PM_{12}O_{40}]^{n-}$

The Keggin polyoxometalate is spherical in shape and has a central tetrahedral phosphate arranged as units of four M_3O_{13} units, giving it a global tetrahedral structure (Figure 1.4 B). Each of the four M_3O_{13} units can be found in two possible orientations in the caged framework thus making five rotational isomeric forms possible namely α , β , γ , δ and ϵ . The Dawson structure can be thought of as being made up of two Keggin lacunary fragments with three missing octahedral metal units.

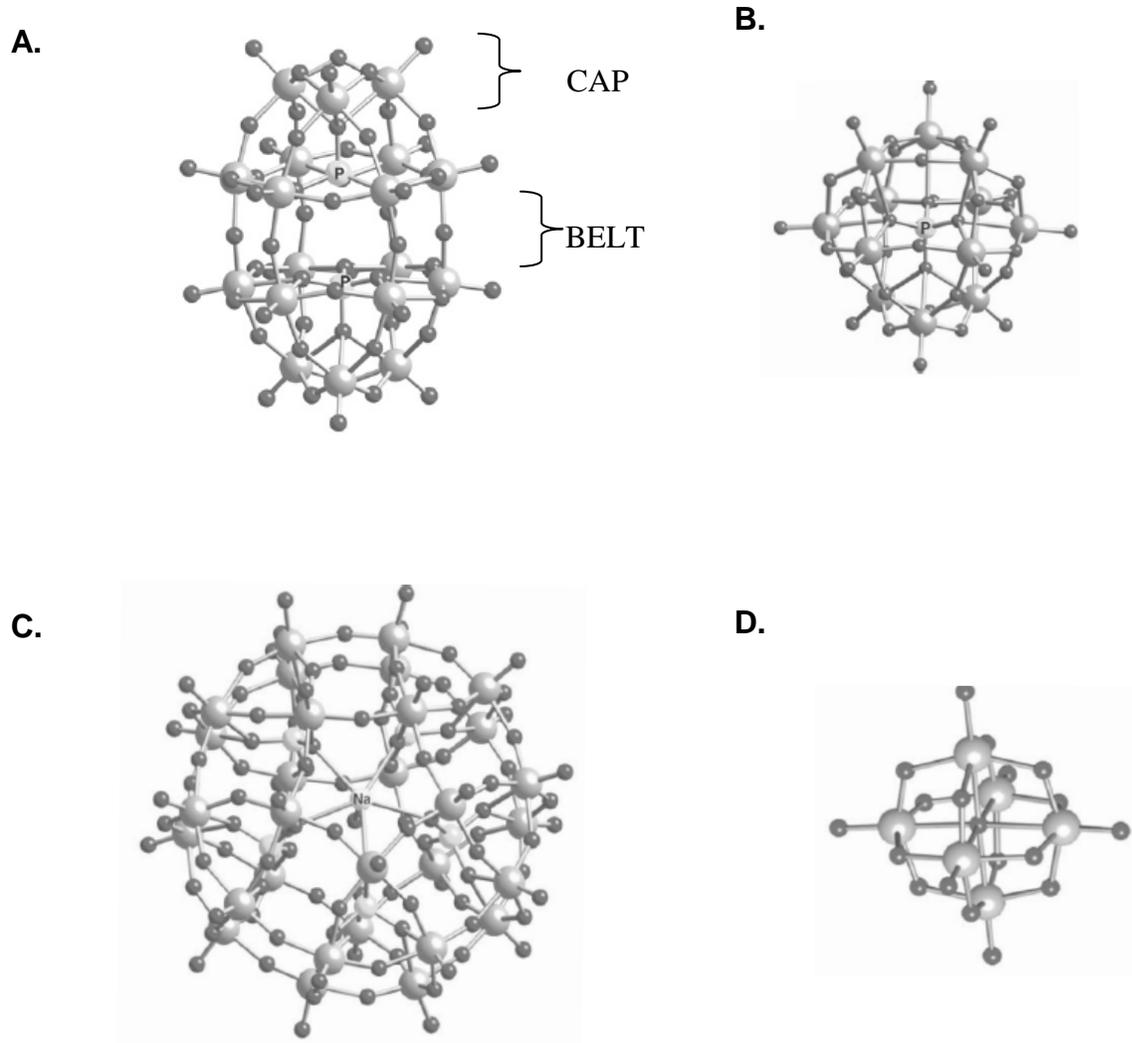


Figure 1.4

Ball and stick representation of some fundamental polyoxometalate structures namely

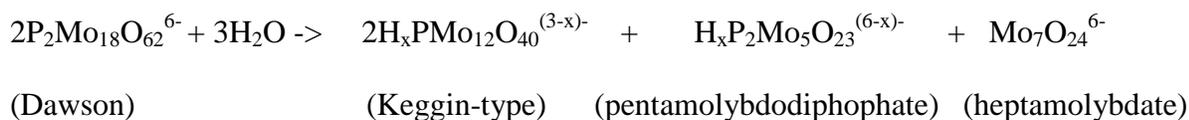
- A) Dawson $[P_2M_{18}O_{62}]^{n-}$
- B) Keggin $[PM_{12}O_{40}]^{n-}$
- C) Preyssler $[MP_5W_{30}O_{110}]^{n-}$
- D) Lindqvist $[M_6O_{19}]^{n-}$

where “M” is the transition metal atom and “n” is the number of ionic-charges. Small light gray spheres are oxygen atoms and the bigger dark spheres are transition metals like Mo and W. The central phosphate atoms are labeled.

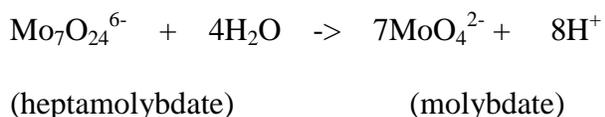
1.8.3 The stability of polyoxometalates (POMs)

POMs generally undergo multiple condensation-hydrolysis equilibria in solution depending on the pH and temperature(140). Hence one of the critical issues in all biological studies is to evaluate the stability of the POM under consideration. A number of studies have highlighted the difficulty and challenge involved in identifying the final active POM species in aqueous solutions and biological media (141,142). In the absence of NMR, kinetic or thermodynamic speciation data, a number of studies on POMs have been limited in their identification of the active hydrolytic species (139,141). For example, the classic phospho-molybdc Dawson-type POMs like $(\text{NH}_4)_6\text{P}_2\text{Mo}_{18}\text{O}_{62}\cdot 12\text{H}_2\text{O}$ decompose into the lacunar Keggin-type anion $\text{H}_x\text{PMo}_{11}\text{O}_{39}^{(7-x)-}$, pentamolybdodiphosphate $\text{H}_x\text{Mo}_5\text{P}_2\text{O}_{23}^{(6-x)-}$, phosphate and oxomolybdate regardless of the pH of the solution(140). A more detailed profile of the hydrolysis reaction of the Dawson-POM is shown below.

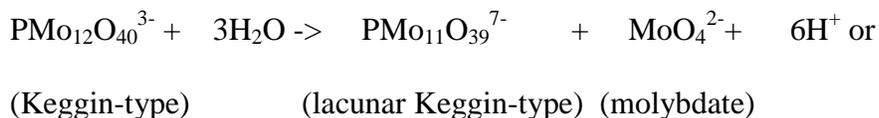
Decomposition reaction of the Dawson-anion:



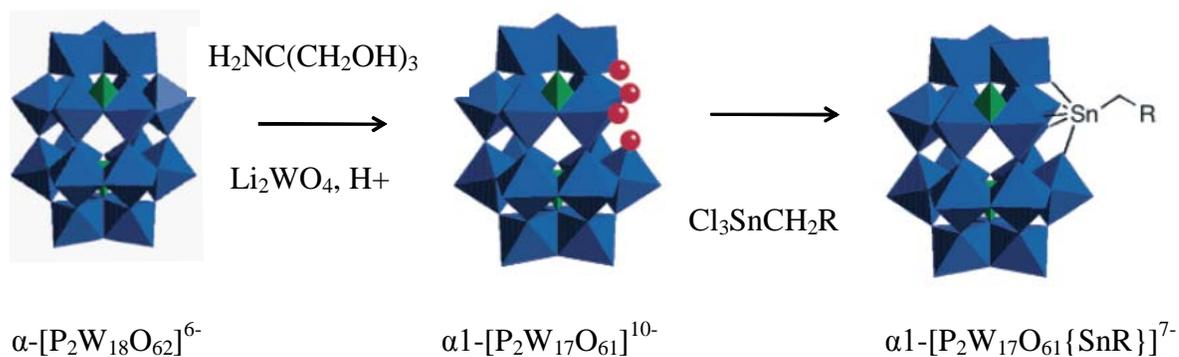
If pH ≥ 6.0 then heptamolybdate breaks down to molybdate species



If pH ≥ 2.0 then Keggin-type breaks down to lacunar Keggin-type species



A.



B.

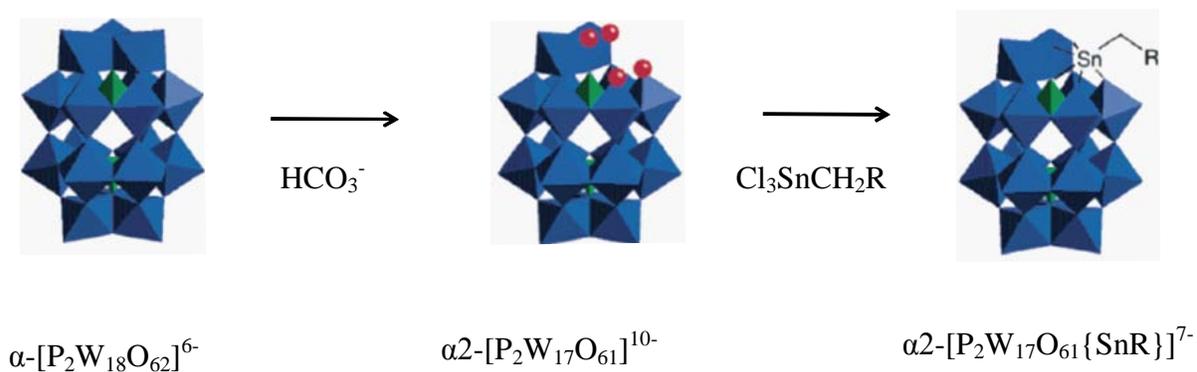


Figure 1.5

- A) Treatment of $\alpha\text{-[P}_2\text{W}_{18}\text{O}_{62}]^{6-}$ in a buffered solution of Tris and Lithium tungstate results in lacunary $\alpha 1\text{-[P}_2\text{W}_{17}\text{O}_{61}]^{10-}$ which upon treatment with a trichlorostannane Cl_3SnR yields the polyanion of the form $\alpha 1\text{-[P}_2\text{W}_{17}\text{O}_{61}\{\text{SnR}\}]^{7-}$ that can be precipitated by TBABr.
- B) Treatment of $\alpha\text{-[P}_2\text{W}_{18}\text{O}_{62}]^{6-}$ in a buffered solution of bicarbonate results in lacunary $\alpha 2\text{-[P}_2\text{W}_{17}\text{O}_{61}]^{10-}$ which upon treatment with a trichlorostannane Cl_3SnR yields the polyanion of the form $\alpha 2\text{-[P}_2\text{W}_{17}\text{O}_{61}\{\text{SnR}\}]^{7-}$ that can be precipitated by TBABr.

polyanion of the form $\alpha 2\text{-[P}_2\text{W}_{17}\text{O}_{61}\{\text{SnR}\}]^{7-}$ that can be precipitated by TBABr (Figure 1.5 B) (144). The tin atom is coordinated by being bound to one of the central phosphate groups, with the overall Dawson polyoxometalate structure acting as a pentadentate ligand. The sixth position of the tin atom is typically attached to organic moieties and open for further derivatization. Compared to the $\alpha 2$ isomer, the $\alpha 1$ isomer is chiral and therefore has a great potential in biological applications. The tin substituted $\alpha 1$ and $\alpha 2$ wells Dawson polyoxometalates thus have a flexible chemistry enabling amines, peptides and alcohols to be attached to the side chains with promising prospects in the field of biomedicine (144). In recent times, click chemistry based methods (copper-catalyzed azide/alkyne cycloaddition) have further expanded the range of organic substrates that could be attached to the polyoxometalate framework(145).

1.8.5 Biological activities of polyoxometalates

1.8.5.1 Anti-viral activities of polyoxometalates

A number of *invitro* studies on POMs revealed that they are effective against a broad spectrum of viruses like the Rhabdovirus, HIV, Rabies, polioviruses and Murine Leukemia sarcoma (MLSV) (146,147). One of the most notable anti-viral POMs was the tungstoantimonate HPA-23 that was under study for the treatment of HIV(148). However, clinical trials of HPA-23 revealed toxicity and low anti-viral activity rendering the drug unacceptable(149). Since then a number of efforts have been made to chemically render POMs benign and to reduce their toxicity (139). Large scale systematic studies on POMs have revealed that the anti-viral activity of polyoxometalates is dependent not only on the structural class, charge, composition, counter-ion and size of the POM but also on the cell-line and virus under observation(138). A number of polyoxometalates tested *invivo* on mouse models infected with Scrapie virus, Human Simian Virus and Flavivirus infections have also shown encouraging results upon polyoxometalate treatment(138).

1.8.5.2 Anti-tumor activities of polyoxometalates

One of the earliest studies on the anti-tumor activity of polyoxometalates was an *in vivo* study of human patients with carcinoma of the intestinal tract. A combination of phosphotungstic acid, phospho-molybdic acid and caffeine was used on patients and after 2-4 weeks the cancerous growth was observed to have been ameliorated (150). Mice implanted with subcutaneous tumor were found to have a substantial reduction in tumor volume after injection with Anderson-type polyoxomolybdates and heptamolybdates (138). The mechanism behind the anti-tumoral activity of heptamolybdates is proposed to be derived from the toxicity of the reduced form of the heptamolybdate $[\text{Mo}_7\text{O}_{23}(\text{OH})]^{6-}$ as compared to its oxidized state $[\text{Mo}_7\text{O}_{24}]^{6-}$. The $[\text{Mo}_7\text{O}_{24}]^{6-}$ reduction to $[\text{Mo}_7\text{O}_{23}(\text{OH})]^{6-}$ is brought about by charge transfer from oxygen-to-molybdenum by flavinmononucleotide, a natural reductant in tumor cells (138). A number of studies of derivatized polyoxotungstates against different cancer cell-line *in vitro* have been able to correlate cytotoxicity with reduction potential, the size and composition of the POMs (151). Starch-encapsulated POMs have also been found to be very effective in enhancing the anti-tumor effect of $[\text{CoW}_{11}\text{TiO}_{40}]^{8-}$ in mice, albeit with much lower toxicity as compared to drugs like cyclophosphamide and fluouracil (138,152).

1.8.5.3 Polyoxometalates as competitive inhibitors of the DNA binding activity of HIV-1 RT and Rad 51

One of the earliest studies of the effect of POM on HIV-1 RT showed that the POM $[(\text{O}_3\text{POPO}_3)_4\text{W}_{12}\text{O}_{36}]^{16-}$ acts as a competitive inhibitor of RT DNA binding, due to a strong POM-polycationic enzyme interface interaction (153,154). The anti-nucleic acid binding activity of polyoxometalates was demonstrated in another example where sodium-metatungstate was reported to act as a competitive inhibitor of the archaeal protein Rad51(155). A crystal structure

was reported by Luo et.al in 2009 that had serendipitously captured Rad51 bound to sodium-metatungstate, an ingredient used in the crystallization condition of Rad51(155). The structure revealed that the POM bound to the negatively charged DNA binding surface of Rad51, explaining its competitive DNA binding inhibitory activity.

1.8.5.4 Polyoxometalates as non-competitive inhibitors of CK2 and Kinesin

CK2 is a protein kinase critical for cell proliferation and survival and is known to become overactive in a number of cancers like prostate cancer and acute myeloid leukemia(156,157).The Dawson polyoxometalate $K_6[P_2Mo_{18}O_{62}]$ was found to be a nano-molar, non-competitive inhibitor of CK2 with a relatively high degree of selectivity over 29 other serine/threonine and tyrosine kinases (141). Kinesins are proteins that interact with microtubules and are essential for the formation of bipolar mitotic spindle (158). Small molecule inhibitors of Kinesins have immense applications as anti-cancer agents. The Dawson polyoxometalate $K_6[P_2Mo_{18}O_{62}]$ was also shown to inhibit kinesin-5, disrupting its binding to microtubules (159). This inhibitory activity was observed to be non-competitive with monastrol binding to kinesin-5.

1.8.5.5 Polyoxometalates as inhibitors of HIV-1 protease, Neuraminidase and HDACs

Studies on HIV-1 protease revealed that the wells-Dawson polyoxometalate $K_7[P_2W_{17}(NbO_2)O_{61}]$ inhibits HIV-1 protease with an IC_{50} of 2.0 μM (154). Based on molecular modeling, two different binding regions for the POM were proposed on the HIV-1 protease. The positively charged “hinge” region of the HIV-1 flaps was proposed to be the most probable binding site for POM with the active site in its “flaps-open form” being the second most probable binding site (154,160). In another study on RNA viral proteins, titanium containing polyoxotungstates were found to exhibit strong anti-influenza activity (161). Molecular modeling

and dynamic studies later revealed that the POMs exhibit a very strong electrostatic interaction with the viral protein Neuraminidase, highlighting its potential as a promising candidate against the ever increasing mutant forms of influenza Neuraminidase (162). Most recently a polyoxometalate PAC-320, a tri-organic-tin-substitute germanotungstate was found to be a very strong inhibitor of HDAC intra-cellular activity in cancer cell lines (163).

1.8.6 Delivery of polyoxometalates inside cells

As polyoxometalates are large and highly negatively charged their penetration of cellular membranes poses a serious challenge. Surprisingly, a number of studies have revealed that POMs do cross the semi-permeable membrane as in the instance where using raman spectroscopy and photonic microscopy, HPA-23 was found to penetrate through C3HBI fibroblast cells (139,164). Polyoxometalates loaded into starch nanoparticles were tested in mice and this encapsulation was found to extend its stability, anti-tumoral activity and survival rate (165). Liposome encapsulation (LEP) of POMs $K_6SiW_{11}TiO_{40}$ {(SiW₁₁Ti) LEP} was found to help retain the parent structure of the POMs resulting in enhanced stability, anti-tumoral activity, cellular penetration and lowered toxicity in HeLa and KB cancer cell lines (165).

1.9 Aims of the research project

The specific objectives of this research project are

1. To show that a prototypical Sox-HMG TF DNA binding domain is chemically tractable and therefore expand the boundaries of TF drug targeting. Demonstrating chemical tractability of a TF DNA binding interface can be expected to falsify the underlying presumption that transcription factor-DNA interactions are chemically “undruggable”.
2. The project will utilize structurally and mechanistically diverse chemical libraries in a high-throughput screening approach using fluorescence anisotropy to identify small molecule inhibitors that interfere with Sox2-HMG –DNA interaction under equilibrium conditions. More precisely, the small molecule would inhibit the binding of Sox2 to the *CCND1* (an oncogene) regulatory DNA element that was identified to be occupied by Sox2 in both hES cells and breast cancer cells (67,129).
3. The mechanism of interaction of the inhibitory candidate small molecule with Sox2 will be deciphered through biophysical and structural biology studies.
4. Whenever deemed appropriate, sufficient chemical modifications of the candidate hit molecule would be made and its selectivity would be assessed with a panel of Sox-HMG members and members of unrelated structural TF classes.

CHAPTER 2

METHODS AND MATERIALS

2.1 Protein expression and purification

2.1.1 Sox2-HMG domain expression and purification

The HMG domain of Sox2, spanning residues 33-141 (Sox2HMG) of the full length mouse protein was cloned into pETG20A expression vector (166). A single colony of chemically competent BL21(DE3) *E.coli* cells transformed with the expression plasmid was initially cultured overnight in 100ml Luria Bertani (LB) containing 100ug/ml ampicillin at 37°C shaken at 220 r.p.m. The following day, 10ml of culture was added to 1 liter of LB containing ampicillin and incubated with shaking at 37°C until the OD₆₀₀ was ~0.5. Protein expression was induced by addition of IPTG (Invitrogen) to a final concentration of 0.5mM. After a 4 hour shaking incubation at 30°C, cells were harvested by centrifugation at 10,000 r.p.m for 5min (sorrall SLA 3000). Cells were resuspended in Buffer C (50mM Tris-HCl, pH 8.0, 150mM NaCl), lysed by sonication and the lysate was clarified by centrifugation at 20,000 rpm for one hour (Sorrall SS34). The supernatant was purified by metal affinity using Ni-NTA (Qiagen) resin equilibrated with Buffer C. Protein was eluted in Buffer C containing 300mM imidazole. Eluted His₆thrx-Sox2-HMG protein was desalted with Buffer A (50mM Tris-HCl pH 8, 100mM NaCl) on a PD-10 desalting column in order to remove imidazole prior to Tobacco Etched Virus (TEV) protease digestion. The histidine-thioredoxin tag was removed by TEV digestion (substrate-to-enzyme ratio of 100:1) at 4°C for approximately 16 hours. Sox2-HMG was purified by cation-exchange chromatography (Resource S column equilibrated with Buffer A) in order to remove the his₆thrx tag and TEV from the digestion mixture. Sox2-HMG was eluted using a linear gradient ranging from 100mM NaCl to 1.0 M NaCl in Buffer A. Sox2-HMG was subjected to a final purification step using HiPrep Superdex-200 gel filtration column in Buffer A. Fractions containing Sox2-HMG were pooled and concentrated to ~10mg/ml. The purity of the collected fraction was

verified in a SDSPAGE gel. ^{13}C , ^{15}N isotope labeled Sox2HMG was obtained by growing *Escherichia coli* cells in an overnight M9 minimal media containing isotope labeled glucose and ammonium chloride (Cambridge Isotope Laboratories) at 18°C. One litre of the M9 minimal medium contained 52.7ml 1M Na_2HPO_4 , 26.5ml 1M KH_2PO_4 , 2ml 5M NaCl, 1.2ml 2mM MgSO_4 , 1.2ml 0.1mM CaCl_2 , 1.2ml 0.5%(w/v) thiamine-HCl, 2g ^{13}C glucose (Cambridge Isotope Laboratories), 1g ^{15}N ammonium chloride. Isotope-labeled Sox2-HMG was expressed and purified using the same chromatographic procedures as described for unlabeled Sox2-HMG(71). Appendix A shows the protein sequences and appendix B images of the purified Sox2-HMG protein

2.1.2 Sox-Homologs, REST C2H2, FoxA1 and AP2 purification

The mouse REST C2H2 zinc finger protein was produced as described (167). The mouse Sox paralogs Sox4, 5, 6, 7, 8, 9, 10, 11, 17 and Sox18 were kind gifts from Calista Keow Leng Ng and were purified using established protocols(168). Full length Ap2 was a kind gift from Tan Si kee and was purified using unpublished protocols. Full length FoxA1 was a kind gift from Hong Shuzen and was purified using unpublished protocols. Protein sequences are listed in appendix A.

2.1.3 Pax6 paired domain expression and purification

The mouse Pax6 paired domain spanning residues 4-136 (GI:220938183) was cloned into a pETG40A expression vector and expressed as a MBP fusion protein in BL21(DE3) cells at 18°C by addition of 0.5mM IPTG in LB (Luria Bertani) media. Cells were collected by centrifugation, resuspended in a lysis buffer containing 40mM HEPES pH 7.5, 200mM NaCl, 5mM β -mercaptoethanol and 2mM EDTA and sonicated on ice. Fusion proteins were extracted from cell lysates using an amylose column (New England Biolabs) equilibrated with the lysis buffer

and eluted with the same buffer supplemented with 10mM maltose. The fusion MBP tag was cleaved using TEV protease at 4°C for approximately 16 hours. A further purification by heparin column (GE) chromatography was performed to remove the MBP tag. Pax6 4-136 was eluted using a linear gradient ranging from 100mM NaCl to 1.0M NaCl and the purity of the proteins were found to be greater than 90%. Please refer appendix A for the protein sequences and appendix B for images of the purified Pax6 protein.

2.2 Annealing DNA duplexes

PAGE purified, labeled/unlabeled DNA elements were obtained from Sigma-ProLigo. The DNA strands were annealed in an annealing buffer with a working composition of 20 mM Tris pH 8.0, 50 mM KCl and 50 mM MgCl₂ in a PCR thermocycler by initially ramping to a temperature of 95°C for 5 mins followed by a slow cooling to 4°C at the rate of (0.5°C/sec). The sequences of all the DNA elements used in the study are provided in appendix C.

2.3 Fluorescence anisotropy measurements

The fluorescence anisotropy assay is a spectroscopic technique that measures the tumbling rate of a sample containing a fluorophore (169). A 22bp fluorescently labeled (Fluorescein) DNA element based on a *cis*-regulatory element of the *CCND1* gene that has been reported to bind to Sox2 was chosen as the DNA substrate (See Appendix C for *CCND1* sequence)(129). When polarized light excites the (FAM)-*CCND1* element (5' Fluorescein labeled *cis*-regulatory *CCND1* DNA element), the relatively small (FAM)-*CCND1* element which undergoes rotational diffusion causes depolarization of the emitted light resulting in a low anisotropy measurement. When Sox2-HMG binds to (FAM)-*CCND1*, the larger size of the protein-DNA complex causes a slower rotation, resulting in a relatively higher polarization/anisotropy of the emitted light (Figure 2.1). This anisotropy assay strategy was chosen as it can easily be scaled up to a HTS

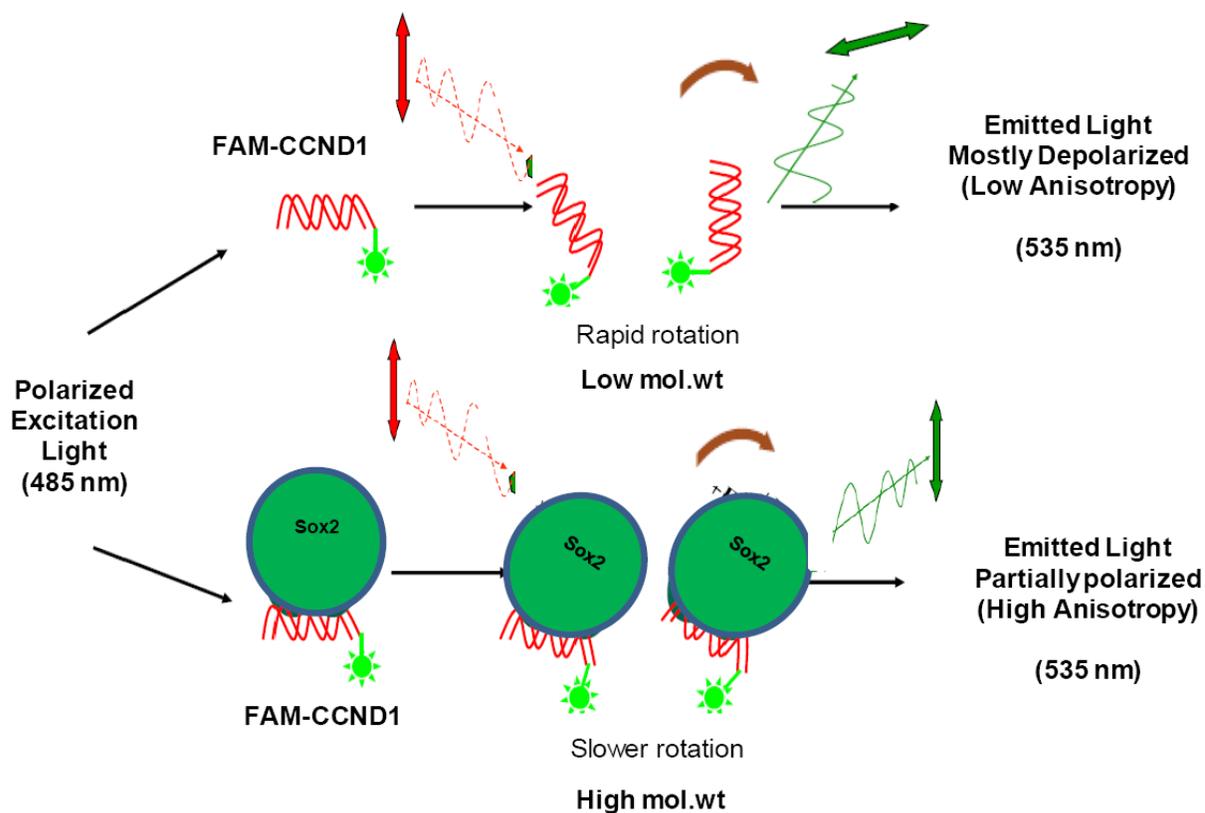


Figure 2.1

The principle of the fluorescence anisotropy based screening is that the fluorescently labeled DNA emits mostly depolarized light owing to its lower molecular weight as compared to the protein-DNA complex that emits partially polarized light because of its higher molecular weight. In a small molecule screening set-up, the difference in the intensity of polarized light can be measured to determine the bound and unbound states of the labeled DNA element with Sox2-HMG in the presence of the small molecules.

assay, to identify small molecule inhibitors that can specifically disrupt Sox2-HMG DNA complexes. Assays involving fluorescein labeled DNA and Sox2-HMG were carried out in a black 384 microplate (Corning NBS). The fluorescence anisotropy measurements from the microplates are read on a Spectramax M5 microplate reader (Molecular Devices) with excitation at 485 nm, emission at 525 nm and a cut-off filter of 515 nm. The PMT was set to a high sensitivity with number of reads per well set to 100. The buffer solutions for the fluorescence binding experiments (fluorescence anisotropy) had the final working composition of 10mM Tris pH 8.0 and 100mM KCl prepared with molecular grade water.

2.3.1 High-throughput fluorescence anisotropy screening using the Sciclone ALH-3000 workstation

All the liquid transfer steps involving assay reagents like protein, DNA and small molecules, to the 96 or the 384 well microplates, were automated using the Caliper Life Sciences, Inc.'s Sciclone ALH 3000 LiquidHandler Workstation. Two small molecule libraries, namely the Mechanistic (825 compounds, 1mM in 100% DMSO) and Challenge diversity library (1364 compounds, 10mM in 100% DMSO) were procured from DTP, NIH (http://dtp.nci.nih.gov/branches/dscb/repo_open.html). Prior to screening, a working stock of both the libraries was created in 96 well microplates such that the final concentration of the small molecule was 20 μ M in 100% DMSO. Since the chemical libraries employed for screening is dissolved in DMSO, different concentrations of DMSO was used to assess the tolerance of the assay. Increasing the DMSO concentrations seemed to lower the anisotropy of the bound Sox2-DNA complex (Appendix D). However even at 12.5% DMSO, the assay retained a good signal window of ~85mA, suggesting that the assay is tolerant even at high DMSO concentrations (Appendix D). A schematic of the HTP screening setup is presented (Figure 2.2). Initially, 20.5 μ l

of Sox2-HMG was dispensed into 384 well Corning NBS microplates, followed by addition of 4 μ l of the small molecule from the working library stock. The small-molecule protein mixture in the microplate was incubated for one hour at 25°C. This initial step involving the addition and incubation of small molecules with the Sox2-HMG domain was to ensure that the candidate small molecule inhibitors get sufficient access to the DNA binding interface of Sox2. Addition of unlabeled DNA (4 μ l of 1 μ M *CCND1*DNA element) serves as the positive control for complete inhibition of Sox2- fluorescein labeled DNA complex while addition of DMSO (4 μ l of 100% DMSO) acts as the negative control. Finally, 20.5 μ l of (FAM)-*CCND1*DNA (22bp duplex) containing a consensus Sox motif was then added to the 384 well microplate such that the total reaction volume in the microplate was 45 μ l. The assay reagents in the microplate were incubated for one hour at 25°C after which the fluorescence anisotropy was recorded. The final reaction mixture in the microplate contained 75nM of Sox2-HMG, 1.77 μ M of small molecule and 1nM (FAM)-*CCND1*DNA. Screening of each small molecule library was carried out in technical duplicates with the screens being carried out on different days and on freshly prepared working stocks of protein to ensure an unbiased assessment of the assay reproducibility.

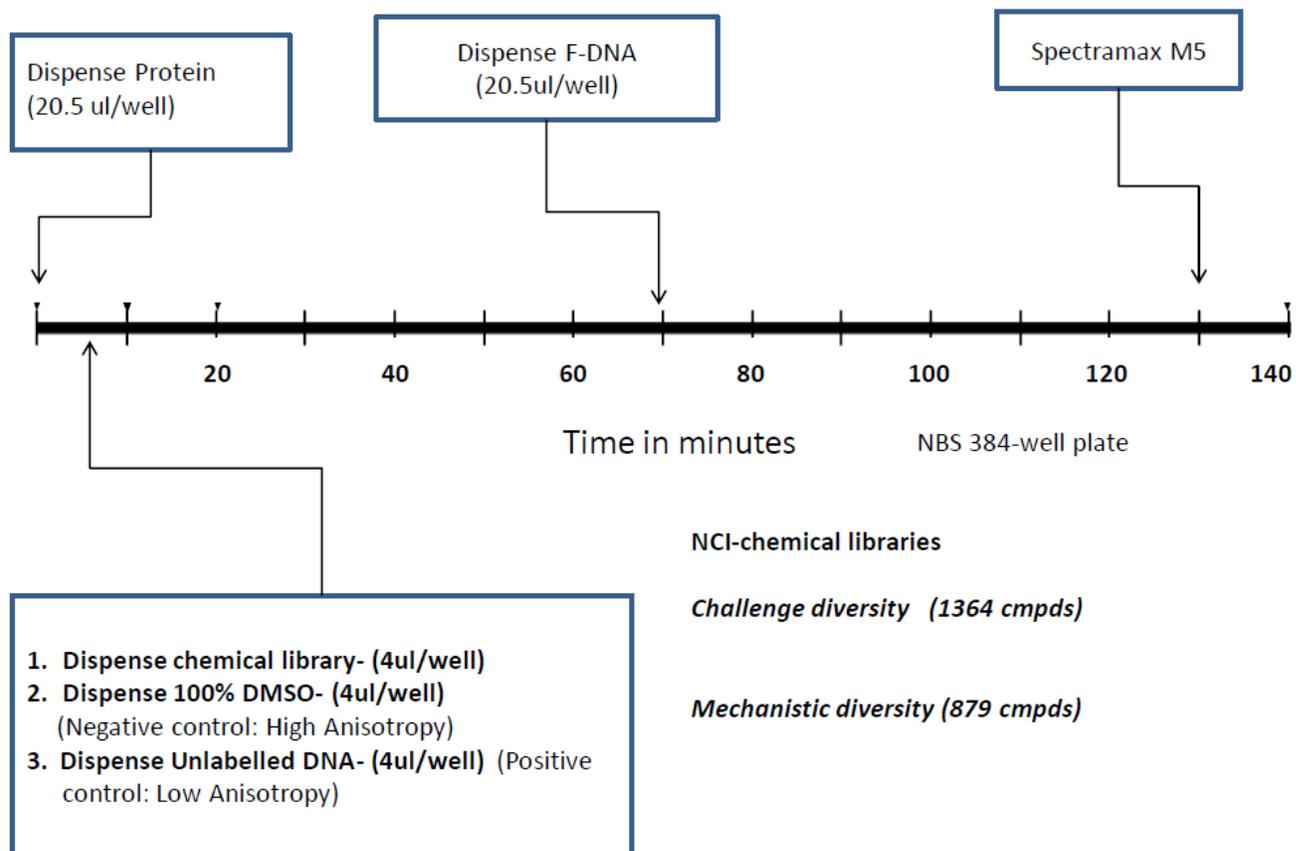


Figure 2.2

Overview of the fluorescence anisotropy based high-throughput screening for inhibitors of Sox2-HMG DNA interaction

2.3.2 HTP fluorescence anisotropy screening data analysis

The Z' factor is a dimensionless parameter routinely employed in high-throughput screening to determine the signal window and precision for accurate assay performance evaluation(170). The Z' factor for the HTP screening was calculated as

$$Z' = 1 - \frac{[3 * SD('+' control) + 3 * SD('-' control)]}{[mA('-' control) - mA('+' control)]}$$

--Eqn(1)

SD ('+' control), SD ('-' control) are the standard deviations (SD) of the positive and negative control anisotropies during the screening while mA('-' control) and mA('+' control) are the millianisotropy units (mA) of the positive and negative controls(170). A dimensionless Z-score measurement was used as the basis for internal-plate normalization. The Z-score values were calculated by subtracting the raw anisotropy value for each compound from the mean of all the measurements within the plate (including the positive and negative controls) and by subsequent normalization using the standard deviation of all the measurements in that plate(4)

$$Z = \frac{(X_i - \bar{X})}{SD_x}$$

--Eqn(2)

X_i - is the raw anisotropy measurement for the 'i'th compound

\bar{X} - is the mean of all anisotropy measurements in the plate

SD_x - is the standard deviation of all measurements in that plate

Thus, each small molecule receives an algebraically assigned “+” or “-“ Z-score depending on the number of standard deviations the anisotropy value fell above or below the mean of each microplate. Compounds that tend to lower the anisotropy of the completely bound Sox2-

DNAcomplex have negative Z-scores, while those that increase the anisotropy of the complex have positive Z-scores. As the screening experiments were carried out in duplicates, the replicate Z-scores for each compound were combined to produce a Composite Z-score. The composite Z-score was calculated as a vector projection of each Z-score in duplicate onto an imaginary <1, 1> line of perfect reproducibility(171).

The amplitude of the projection of the measurement vector of the technical replicates (Z-score screen1, Z-score screen2) onto <1, 1> is given as

$$\text{Composite Z - score} = \frac{Z - \text{score Screen1} + Z - \text{score Screen2}}{\sqrt{2}} \quad \text{--Eqn(3)}$$

Reproducibility value for each compound was calculated as the cosine of the angle between each Z-score and the imaginary<1,1> vector(171). The reproducibility values therefore range from -1 to +1.The reproducibility values for each compound were calculated based on a model that assumes “perfect reproducibility”. Reproducibility was calculated as

$$\text{Reproducibility} = \frac{(Z - \text{score Screen1} + Z - \text{score Screen2})}{\sqrt{2} * \sqrt{(Z - \text{score screen1})^2 + (Z - \text{score screen2})^2}} \quad \text{--Eqn(4)}$$

Small molecules that resulted in a change of more than three negative standard deviations from the mean (Composite Z-score < -3) with near perfect reproducibility (Reproducibility < -0.98) were considered as primary hits in the HTP screening. Compound wells that positively deviated by 3*SD from the mean total intensity of unbound fluorescent DNA was designated as auto-fluorescent, while compound wells that negatively deviated by 3*SD from the mean total

intensity of bound fluorescent DNA was designated as a quencher. The auto-fluorescent and quencher molecules were retained throughout the data analysis and were earmarked for further testing as they could be represent potential false negative or false positive molecules. Data visualization and analysis was carried out using both “Spotfire” and “R” packages.

2.3.3 IC₅₀ determination

In both fluorescence anisotropy and electrophoretic mobility shift assays, IC₅₀ values were determined by adding 45nM of Sox2-HMG to 1nM Fluorescein labelled *CCND1* (~60-70% bound) and 0.5nM of Pax6 to 1nM of its cognate *Pax6* element (~50% bound) respectively followed by addition of the inhibitor K₆[P₂Mo₁₈O₆₂]. This reaction mixture was further incubated for about an hour after which the samples were taken for reading in Spectramax M5 or used to run on a 12% native PAGE gel using Tris-glycine buffer for about 45 minutes at 100V. The IC₅₀ values were determined by fitting the data with a 4-parameter logistic model using sigma-plot. The 4-parameter logistic model is given as

$$Fraction\ bound = min + \frac{max - min}{1 + 10^{(logIC_{50} - x)Hillslope}}$$

--Eqn(5)

Here ‘Min’ and ‘Max’ refer to the maximum and minimum fraction bound protein-DNA complex and ‘x’ is the observed fraction bound in either the fluorescence anisotropy or EMSA measurements. Fraction bound is determined from the change in anisotropy, or quantification of fraction bound intensity of gel-shift bands whereas the logIC₅₀ and Hill slope terms are unknown parameters that will be fitted by non-linear regression.

2.3.4 Selectivity of Various POMs using residual DNA binding activity measurement

The effect of various polyoxometalates on the residual DNA-binding activity of a panel of transcription factors was estimated after addition of 125nM of POM to protein-DNA complexes that are approximately 70-90% bound. (all Sox-HMG proteins were used at 68nM concentration to bind to a 1nM *CCND1* DNA element, 0.5nM Pax6 on a 1nM *Pax6* element, 68nM REST on 1nM *RE-1* element, 250nM AP2 on a *HPSE* element and 60nM FoxA1 on *Foxa1* element). The residual DNA binding activities at 125nM POM is an average of three to five independent experiments and is reported as a percentage of the control without the inhibitor (0% residual binding activity would correspond to maximum inhibition, while 100% activity would correspond to no inhibition). The reaction mixture for electrophoretic mobility shift assays (EMSA) to monitor disassociation of protein-DNA complex in the presence of the inhibitor was performed under the same condition as the fluorescence anisotropy measurements. The samples were run using 1X Tris-glycine buffer on a 12% native PAGE gel for about 45 minutes at 100V.

2.4 Limited proteolysis

Proteolytic reactions were carried out using 50 μ M of Sox2-HMG in the presence and absence of 60nM of $K_6[P_2Mo_{18}O_{62}]$ in a 10mM Tris pH 8.0, 100mM NaCl solution. The final mixture contained 55 μ g of Sox2-HMG, 0.3 μ g trypsin and/or 60nM of the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$. After addition of trypsin, samples were removed at regular intervals (0, 2, 30 and 60 mins) and boiled in denaturing SDS-PAGE loading buffer to interrupt the trypsin digestion. The proteolytic fragments were visualized on a NuPage Novex 4-12% Bis-Tris gel (Invitrogen) stained with Simply Blue (Invitrogen)

2.5 Thermofluor assay

Thermofluor assay is a thermal denaturation assay that compares the change in the unfolding transition temperature (T_m) of a protein in the presence and absence of a ligand (172). Sypro-orange (Invitrogen), a dye which typically emits fluorescence upon binding to hydrophobic patches on protein surfaces was used as the fluorescent reporter to monitor the protein melting as the protein is heated. Ligands that bind to the protein can increase their thermal stability (T_m) as compared to the thermal stability of the protein alone. 14 μ M Sox2-HMG domain solubilized in a 10mM Tris pH 8.0, 100mM KCl, 50uM ZnCl₂, 10% glycerol, 3%DMSO buffer was initially dispensed in a 384 well micro-plate, followed by which increasing concentrations of the Dawson-POM K₆[Mo₁₈O₆₂P₂] (NSC 622124) was added. The protein-inhibitor complex was incubated for about an hour after which Sypro-orange was added to a final concentration of 30X to the solution. Sypro-orange was purchased commercially from Invitrogen as 5000X stock. The 384 well microplates were sealed and centrifuged before being taken for thermal unfolding. Thermofluor measurements were recorded in a Roche LC480 PCR device with excitation at 450 nm and emission followed at 568 nm.

2.6 NMR sample preparation

The solutions used for this study consisted of 0.65 mM ¹³C, ¹⁵N isotope-labeled Sox2-HMG in 90% H₂O/10% D₂O. All samples used for the two dimensional TROSY, and three dimensional HNCA, HNCACB and CBCAcoNH were adjusted to a pH of 7.0 in 50 mM K₂HPO₄/KH₂PO₄, 100mM NaCl buffer. The same buffer conditions were maintained for the Sox2-Dawson-POM (K₆[Mo₁₈O₆₂P₂]) binding studies. The Dawson-POM K₆[Mo₁₈O₆₂P₂], was prepared as a 50mM stock solution in dimethylsulfoxide-*d*₆ (DMSO-*d*₆). Addition of POM to the free-ligand solution (SOX2 only) was done stepwise to achieve final concentrations of SOX2: POM ratios of 1:0.25,

1:0.5 and 1:1. The Sox2-HMG sample without POM was used as the reference for calculation of the chemical shift perturbations ($\Delta\delta$). All experiments were conducted at 298K. The final DMSO concentration in the solution was 2% DMSO.

2.6.1 NMR spectroscopy and data processing

NMR experiments were performed on a Bruker AVANCE II 600 MHz NMR spectrometer equipped with four RF channels and a 5mm z-gradient TCI cryoprobe. The spectra were collected at a regulated temperature of 298K, sweep width for ^1H and ^{15}N were 9804Hz and 2412Hz respectively. The residual HDO resonance signal was suppressed with presaturation. A combination of experiments was used to derive the assignments of the backbone for Sox2-HMG. ^1H and ^{15}N resonances observed from the TROSY experiments were correlated with their corresponding inter- and intra-residue spin systems from 3D experiments namely HNCA, HNCACB and CBCAcoNH to sequentially correlate the amino acids. The raw data was processed using TopSpin 2.1 software, and the chemical shifts were referenced directly (^1H) to the frequency of DSS. Peak peaking and spectral analysis was done using CARAMER (173).

2.7 Docking study of the Dawson-POM [$\text{P}_2\text{Mo}_{18}\text{O}_{62}$] $^{6-}$ with Sox2-HMG

A crystallographic model of a Dawson-POM ($\text{C}_{60}\text{H}_{50}\text{Ag}_4\text{N}_{12}\text{O}_{62}\text{P}_2\text{W}_{18}$) was obtained from the crystallography open database (COD: 4304834.cif) and then modified to remove the organic groups, retaining the Dawson structure [$\text{P}_2\text{W}_{18}\text{O}_{62}$] alone using the modeling tool Avogadro (174). The partial charge of the ligand was calculated with quantum chemical semi-empirical PM6 method using MOPAC2009 software (175). The total charge of the ligand was set to -6 and the structure kept rigid throughout the calculation. Autodock Tools software was used in order to prepare Autodock4 input (pdbqt) file. H++ server was used for checking the protein structure and for adding hydrogens (176). Missing heavy atoms and hydrogens were added to the structure.

Semi-empirical assignments were performed using the PM6 method by the Mozyme function of MOPAC2009 program. The calculated partial charges were applied for further calculations. Two autodock searches were undertaken with similar settings. The first docking search was set-up so that the POM can explore the entire surface of the Sox2-HMG, while the second involved a search of the region centered around the cluster of residues M7, E66 and D69, identified by ^{15}N ^1H TROSY NMR chemical shift perturbations. Docking calculations were carried out according to the Docking Server methodology(177). Solvation parameters were added with the aid of AutoDock tools(178). Affinity (grid) maps with 0.375 Å spacing were generated to cover both the NMR chemical shift identified regions as well as the entire protein surface using the Autogrid program(178). AutoDock settings were parameterized that the orientation of the ligand molecule and its initial position was set randomly. The Lamarckian genetic algorithm and the Solis & Wets local search strategies were employed with the docking experiment derived from 100 runs set to terminate after 2500000 energy evaluations(179). The Dawson POM molecule was parameterized for a rigid docking so that the bonds do not have any free torsion during the docking calculation. Solvent accessibility analysis of pdb structures were carried out using GetArea(180).

2.8 Polyoxometalates

Sodium phosphomolybdate $\text{Na}_3[\text{PMo}_{12}\text{O}_{40}]$, Ammonium phosphomolybdate $(\text{NH}_4)_3[\text{PMo}_{12}\text{O}_{40}]$, Sodium phosphotungstate $\text{Na}_3[\text{PW}_{12}\text{O}_{40}]$ and Sodium metatungstate $3\text{Na}_2\text{W}_0.9\text{W}_0.3\text{xH}_2\text{O}$ were purchased from Sigma-Aldrich. The Dawson-phosphomolybdate $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$ and the Dawsonphosphotungstate $\text{K}_6[\text{P}_2\text{W}_{18}\text{O}_{62}]$ were kind gifts from Dr. Bernold Hasenknopf (UPMC, France). Bulk amounts of Dawson-phosphomolybdate $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$ were also obtained by custom synthesis from Asischem Inc.

CHAPTER 3

RESULTS I

Sox2-HMG: Primary high-throughput screening and secondary validation assays

3.1 Primary screening and identification of a polyoxometalate hit

One transcriptionally up regulated target of Sox2 is the *CCND1* gene encoding cyclin D1 which is linked to breast cancer progression (129). A 22bp *cis*-regulatory element of the *CCND1* gene was utilized for Sox2 binding, as the basis for fluorescence anisotropy based screening (Figure 3.1A). This binding assay is linear in the range of ~130 mA units (free FAM-*CCND1*) to 250 mA units (fully bound *CCND1*) (Figure 3.1A). The large signal window (~120mA) separating the Sox2 bound and unbound DNA facilitated the up scaling of the assay into a robust high-throughput format to identify small molecule inhibitors that can specifically disrupt Sox2-HMGDNA complex formation. The steady-state total fluorescence intensity was monitored throughout the binding reaction and no significant change was observed. A fluorescence anisotropy based DNA competition experiment was performed such that the Sox2 protein was just about saturated in binding (~85 -90% bound) to *CCND1*. Addition of unlabeled competitor DNA to a pre-incubated Sox2-HMG complex results in complete displacement of the labeled *CCND1*-Sox2-HMG complex and restores high anisotropy readings. At a concentration of ~2-3 nM unlabeled *CCND1*, the polarization began to reduce and at a concentration of 500 nM competitor DNA the fluorescently labeled *CCND1* was completely displaced (Figure 3.1B). Compound screening was carried out using the mechanistic and the challenge diversity libraries obtained from the National Cancer Institute Developmental Therapeutics Program (NCI, DTP, <http://dtp.nci.nih.gov/index.html>). The Mechanistic diversity library (825 compounds) is derived from 37,836 open compounds that have been tested on the NCI human tumor 60 cell line and represents compounds that exhibit a broad range of growth inhibition effects. The Challenge diversity set (1364 compounds) was derived from a 140,000 compound collection to create a structurally diverse library. In total, 2,189 compounds were screened in duplicates using the

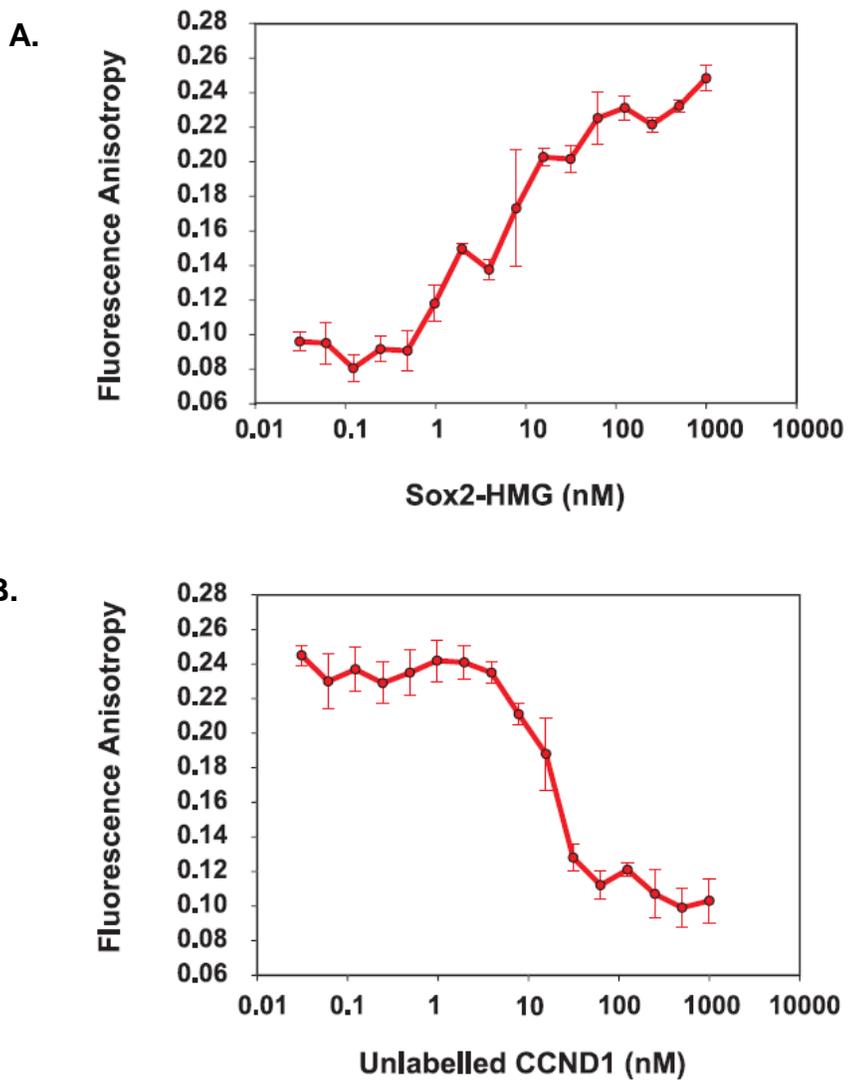


Figure 3.1

- A) Binding isotherm of Sox2-HMG with 1nM (FAM)-*CCND1*. Increasing concentrations of Sox2-HMG increases the fluorescence anisotropy indicating Sox2-HMG DNA complex formation. Data represents the average of 3 independent titrations for a given concentration of Sox2-HMG.
- B) 80 nM Sox2-HMG and 1nM (FAM)-*CCND1* complex was allowed to reach equilibrium. The saturated complex was displaced using competing unlabeled *CCND1* DNA.

SciClone ALH-3000 workstation in 384 well microplates and a SpectraMaxM5 plate reader. Measurements obtained from the positive and negative control samples in each microplate were used to calculate the Z' factor for that plate. To minimize edge-related bias, the positive (4 μ l of 1 μ M unlabeled DNA) and negative controls (4 μ l of 100% DMSO) for the assay were placed in the 2nd and 23rd columns of the microplate such that they appear equally on each of the 16 rows and each of the two columns (Figure 3.2A). For all the individual microplates of the mechanistic and challenge diversity screens, the Z' factor was found to be > 0.60 (Figure 3.2B). Comparison of the replicate screens of the challenge and mechanistic diversity libraries revealed a high degree of correlation indicating that the screening results were reproducible (Figure 3.2C). Compounds with Composite Z-score < -3 and Reproducibility < -0.98 were identified as primary hits and were ranked based on their inhibitory activity (Figure 3.2D). In effect, the ranking system identified 51 compounds as potential primary hits, 33 compounds from the mechanistic diversity library and 18 compounds from the challenge diversity library (see Appendix E for the primary hit list). Many primary hits included planar and aromatic compounds with possible DNA intercalator activity and were therefore excluded from validation experiments. One of the interesting candidate molecules that emerged from the primary screening was a Dawson-type polyoxometalate NSC 622124 $K_6[P_2Mo_{18}O_{62}]$, also referred to as the Dawson-phosphomolybdate. Polyoxometalates (POMs) are usually polyanionic, nanometer sized compounds with well documented biological activity (138,139). Studies from a number of *invitro* experiments on polyoxometalates suggests that these class of compounds bind to proteins like HIV-1 reverse transcriptase, *E.coli* Klenow DNA pol I, and eukaryotic DNA pol β inhibiting nucleic acid binding activity. Therefore, the identification of a polyoxometalate scaffold as a hit molecule in the primary screening was very encouraging. NSC 622124 in itself has been

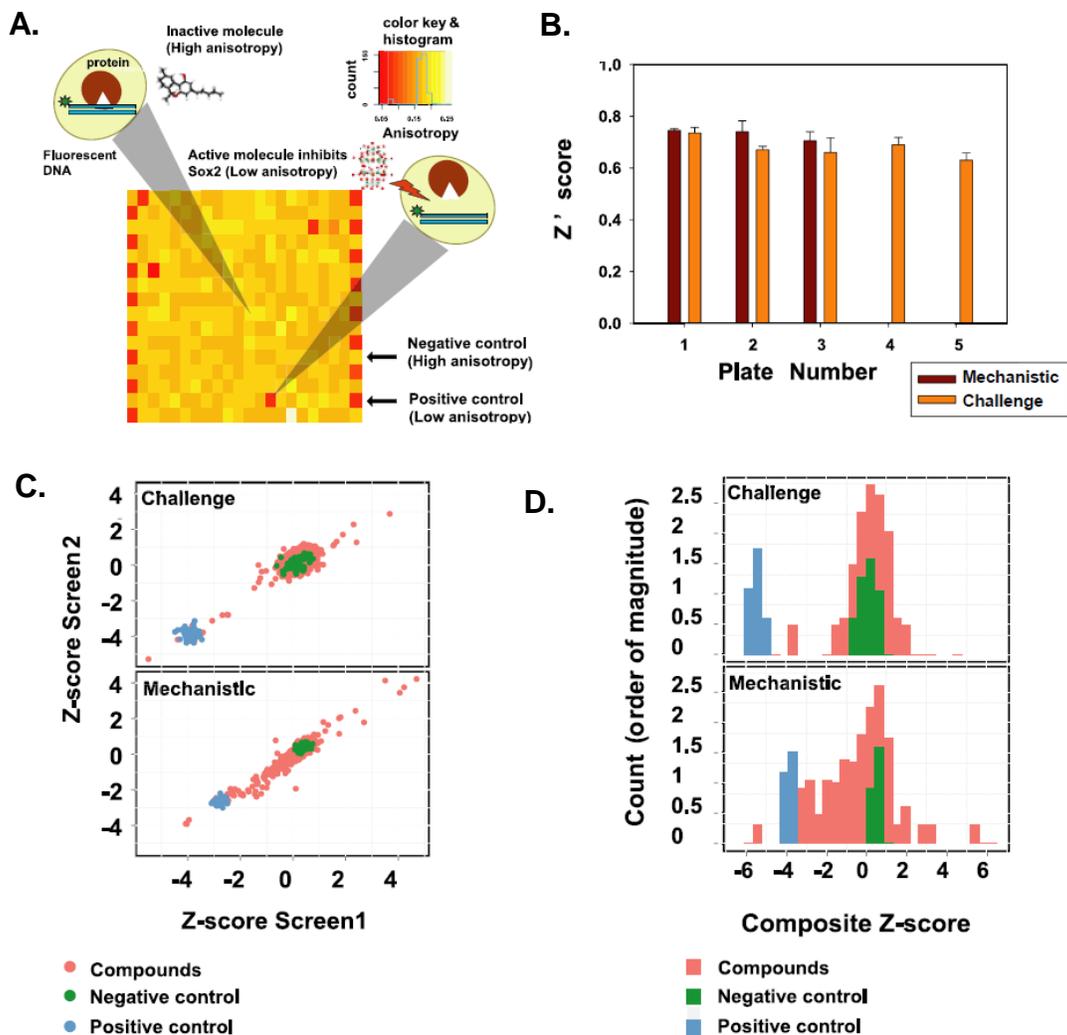


Figure 3.2

- A) Assays are carried out in a 384-well microplate depicted schematically as a heatmap (color coded by anisotropy values. Red color indicates higher inhibition potential while yellow color indicates relatively lesser inhibition). Compounds are added to each well, while the positive and negative controls were alternately added to the peripheral columns (column 2 and 23)(4).
- B) Duplicate screens (Screen1 and Screen2) of the challenge and mechanistic diversity libraries revealed a Z' factor larger than 0.6 indicating a sufficiently large signal window for robust hit identification.
- C) Z-scores from duplicate screens correlate well highlighting the reproducibility of the assay.
- D) Screening results are shown as histograms of composite Z-scores. Primary hits were defined as having a composite Z-score ≤ 3 and reproducibility < -0.98 .

recently reported to selectively bind and inhibit the protein kinase CK2 and certain members of the kinesin family of proteins from high-throughput screening studies (159,181). Given the potential of POMs to selectively inhibit proteins it was decided to further study the inhibitory features of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ along with a series of other POM structures.

3.2 Active Dawson phosphomolybdate $K_6[P_2Mo_{18}O_{62}]$ species responsible for inhibition of Sox2-HMG DNA binding activity

The Dawson-phosphomolybdate $K_6[P_2Mo_{18}O_{62}]$ is known to undergo multiple condensation-hydrolysis equilibria in solution depending on the pH and temperature (140). It is therefore possible that the Sox2-HMG domain is inhibited by a hydrolysis product of the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$ under equilibrium conditions. For example, phospho-molybdic Dawson-type POMs like $(NH_4)_6P_2Mo_{18}O_{62} \cdot 12H_2O$ decompose into the lacunar Keggin-type anion $H_xPMo_{11}O_{39}^{(7-x)-}$, pentamolybdodiphosphate $H_xMo_5P_2O_{23}^{(6-x)-}$, phosphate and oxomolybdate regardless of the acidity-basicity of the solution (140). However, the degradation products molybdate ($[MoO_4]^{3-}$), the phosphate ($[HPO_4]^{2-}$), and the equivalent Keggin phosphomolybdate ($[PMo_{12}O_{40}]^{3-}$) by themselves were found to be insufficient to inhibit Sox2-HMG activity (Figure 3.3). It can thus be concluded that a structurally intact Dawson-structure interacts with the Sox2-HMG.

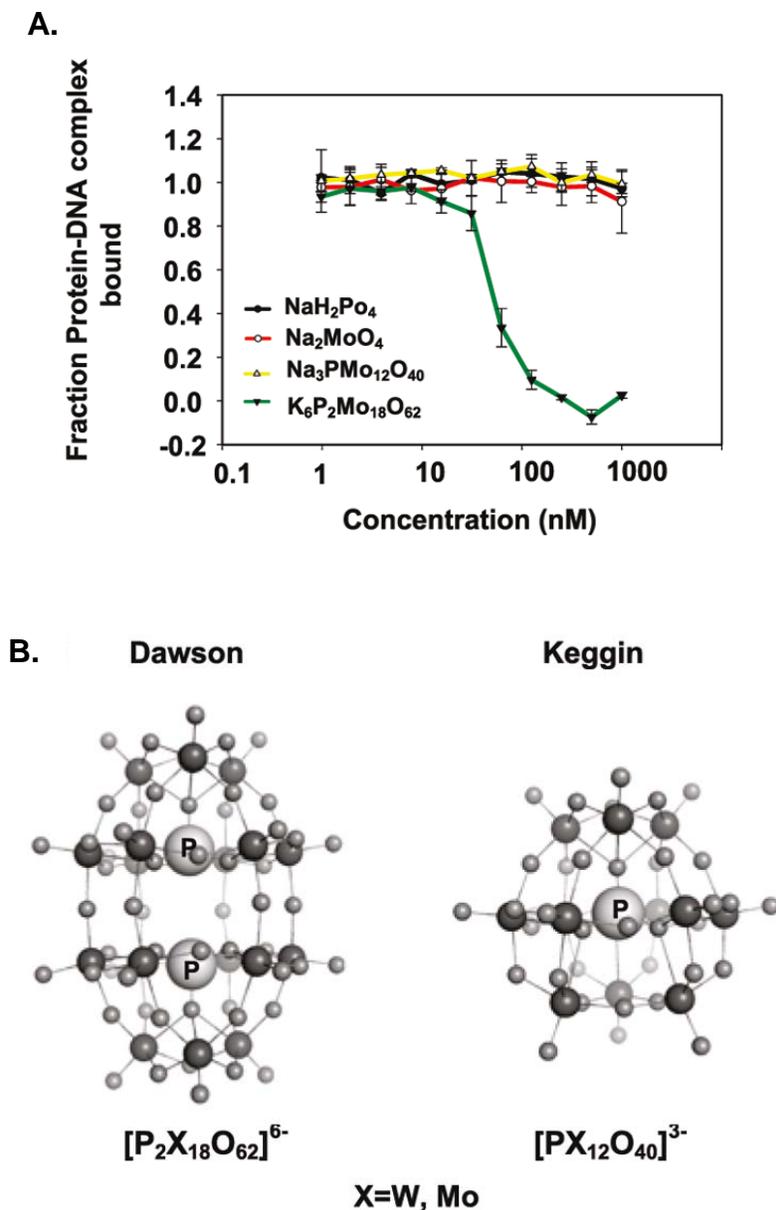


Figure 3.3

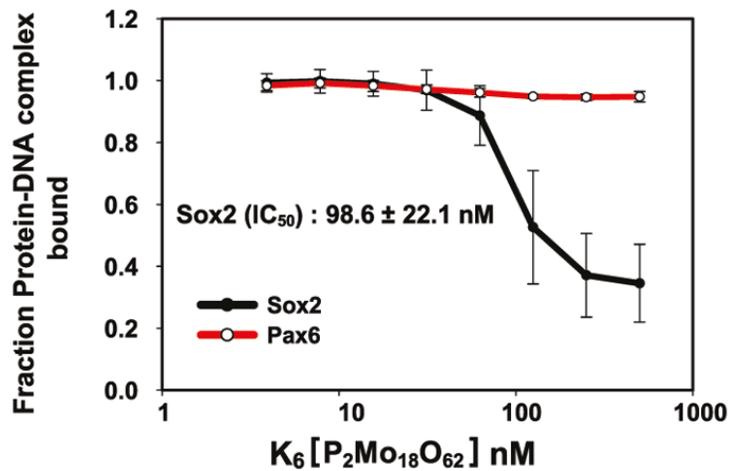
A) Fluorescence anisotropy assay showing that degradation products of $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$, namely the phosphate ($[\text{HPO}_4]^{2-}$), the molybdate ($[\text{MoO}_4]^{2-}$), and the Keggin phosphomolybdate ($[\text{PMo}_{12}\text{O}_{40}]^{3-}$), do not disrupt a half-saturated Sox2-HMG DNA complex. Ball and stick representation of the Dawson and Keggin POMs.

B) Small light gray spheres are oxygen atoms and the bigger dark spheres are transition metals like Mo and W. The central phosphate atoms are labeled.

3.3 Preliminary selectivity studies on the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$

The selectivity of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ for Sox2-HMG was assessed in a preliminary experiment against an unrelated transcription factors Pax6 using electrophoretic mobility shift assays (EMSAs). EMSA experiments were performed such that the Pax6 and Sox2-HMG proteins are incubated with their respective cognate fluorescently labeled DNA elements at concentrations resulting in a fractional binding of 50-70% prior to the addition of the Dawson-POM. It was found that $K_6[Mo_{18}O_{62}P_2]$ inhibits Sox2-HMG DNA binding with an IC_{50} value of 98.6 ± 22.1 nM, while Pax6 did not show any inhibition at all (Figure 3.4). Addition of unlabeled Pax6 consensus DNA sequence (100nM DNA element) serves as the positive control for complete inhibition of Pax6-fluorescein labeled DNA complex (Appendix F). This preliminary selectivity study demonstrated that $K_6[Mo_{18}O_{62}P_2]$ exhibits varying inhibition profiles against unrelated TF structural classes suggesting that $K_6[Mo_{18}O_{62}P_2]$ discriminates between different transcription factor-DNA complexes.

A.



B.

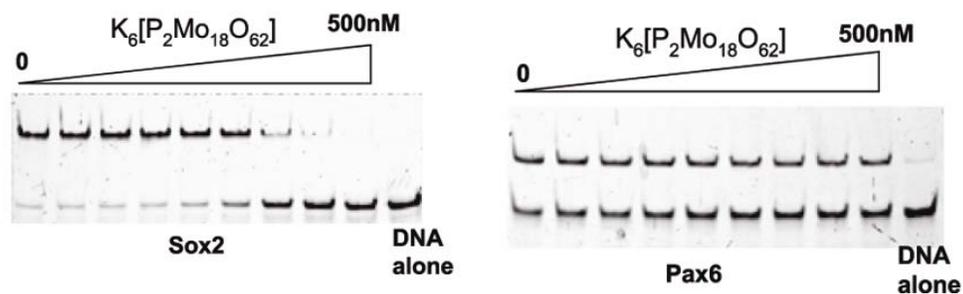


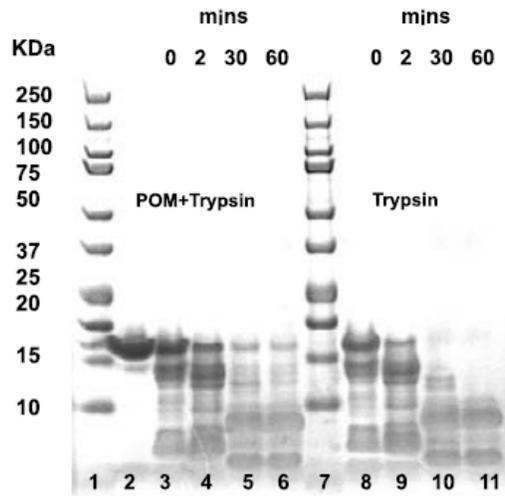
Figure 3.4

- A) Gel-shift assays using varying POM concentrations shows that $K_6[P_2Mo_{18}O_{62}]$ selectively inhibits Sox2-HMG with an IC_{50} value of 98.6 ± 22.1 nM.
- B) Representative EMSA experiment showing dose-dependent titrations of $K_6[P_2Mo_{18}O_{62}]$ with 40 nM Sox2-HMG and 1 nM CCND1 (~50-70% fraction bound) and 0.5 nM Pax6 and 1 nM Pax6 DNA element (~50% fraction bound) reveals selective inhibition of the HMG domain.

3.4 The Dawson-POM $K_6[Mo_{18}O_{62}P_2]$ physically interacts with the Sox2-HMG domain

A series of secondary assays were performed to verify the interaction of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ with the Sox2-HMG domain. Firstly, limited trypsin digestion of Sox2-HMG in the presence and absence of the Dawson-POM revealed that the POM confers sustained resistance to proteolytic digestion, whereas the Sox2-HMG alone is degraded more rapidly (Figure 3.5A). Secondly, a comparison of the thermal unfolding of the Sox2-HMG domain in the absence and presence of the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$, revealed differences that can be attributed to a stabilization effect of the POMs on the melting profile of Sox2-HMG. The addition of increasing concentrations of the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$ transformed the melting profile and increased the unfolding transition suggesting that the Dawson-phosphomolybdate directly binds and stabilizes the Sox2-HMG domain (Figure 3.5B). The thermal denaturation profiles of Sox2-HMG alone and the Sox2-HMG inhibitor complex at 1:1 ratio could not be modeled reliably with Van't Hoff models and hence parameters like enthalpy and melting temperature could not be retrieved (182). However, the nature of the melting profile provided sufficient qualitative information to suggest a physical interaction between the Dawson-POM $K_6[Mo_{18}O_{62}P_2]$ and Sox2-HMG. It must also be mentioned at this context that attempts to measure the binding affinity of the POM to Sox2 by ITC resulted in a failure to obtain titrations with good curvature. This loss of binding curvature in ITC experiments is well documented and is known to occur for ligands with a very high binding affinity. With very tight binding events in the lower nanomolar range (1-100nM) there can be a loss of curvature and accurate estimates of binding constants cannot be obtained as the range of binding constants that can be directly measured by ITC is approximately between 1000 μ M and 10nM(183). It is

A.



B.

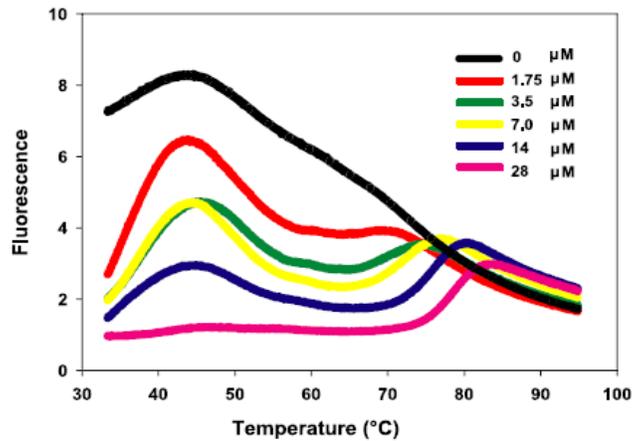


Figure 3.5

A) Limited proteolysis reveals that interaction of Sox2 with $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$ confers resistance to proteolytic digestion by trypsin. Sox2-HMG was incubated with trypsin in the presence (lanes 3-6) and absence (lanes 8-11) of the Dawson-POM $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$. Reactions were stopped after different time points and analyzed by 4-12% SDS-PAGE. Molecular weight markers are added in lanes 1 and 7. Lane 2 contains the Sox2-HMG incubated for 60 min but not subjected to trypsin digestion.

B) Thermal melting profiles of Sox2-HMG monitored in the presence of Sypro-orange with and without increasing concentrations of the Dawson-POM $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$.

presumable from the IC₅₀ estimation (98.6±22.1nM) that the binding affinity could potentially be on the lower nanomolar range and hence on the tighter binding end for reliable ITC estimation.

CHAPTER 4

RESULTS II

**Sox2-HMG $K_6[P_2Mo_{18}O_{62}]$ interaction: structure-
function relationship**

4.1 Studies of Sox2 Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ interaction by NMR

To understand the inhibitory mechanism and to facilitate the follow-up chemistry aimed at optimizing the selectivity and potency of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$, multidimensional NMR experiments were carried out. To assign backbone-resonances ^{13}C and ^{15}N double-labeled Sox2-HMG domain was recombinantly purified and HNCA, HNCACB, and CBCAcoNH spectra were recorded for this sample. Approximately 80% of the 114 residues of the Sox2-HMG construct could be unambiguously assigned, with the unassigned residues predominantly belonging to the largely flexible N- and C-termini. Importantly, residues comprising the helical core of the Sox2-HMG domain and DNA contacting residues could be reliably assigned. To study the effect of POM binding to Sox2-HMG, ^{15}N 1H TROSY measurements were recorded in the absence and presence of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$. When the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ was added additively to 0.65mM protein, the backbone resonances were observed to selectively shift distinct peaks, further corroborating a direct interaction with the Sox2-HMG and suggesting a structurally defined binding region (Figure 4.1). Chemical shift perturbations from these experiments were quantified by using an overall weighted plot ($\Delta\delta = [\Delta\delta^2H_N + (0.2\Delta\delta N)^2]^{1/2}$) (184,185). The combined chemical shift is displayed on a Sox2 model which was derived from a Sox2-Oct1/DNA (PDB:1GT0) crystallographic structure (Figure 4.2). Based on the distribution of the weighted chemical shift values, residues were classified into (i) those that undergo significant chemical shift changes ($\Delta\delta \geq 0.065$ ppm), (ii) those that undergo moderate chemical shift changes ($0.04\text{ppm} \geq \Delta\delta < 0.065$ ppm) and (iii) those with low chemical shift changes ($\Delta\delta < 0.04$ ppm) (Figure 4.3). The residues Met7, His29, Glu66 and Asp69 were shifted most significantly (numbering based on Sox2-model in the crystal structure 1GT0). Among these, Met7 is directly involved in DNA binding whereas Glu66 and Asp69 are not in direct

A.

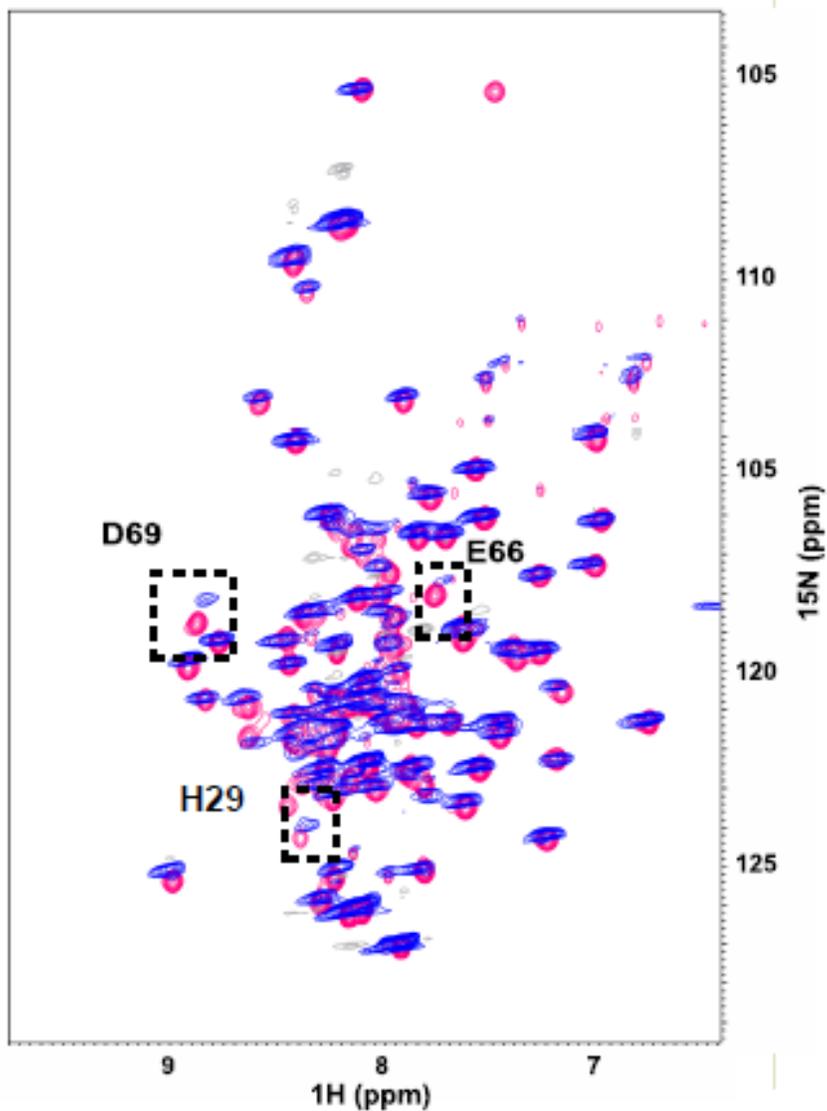


Figure 4.1

Superimposed spectra of two-dimensional TROSY spectra of free Sox2 (pink) and Sox2 bound to POM (blue). Each cross-peak represents a bonded N-H pair. The axes correspond to the chemical shifts of N and H atoms in ppm (parts per million). The peaks that undergo significant shifts upon complex formation namely Glu66, Asp69 and His42 are highlighted.

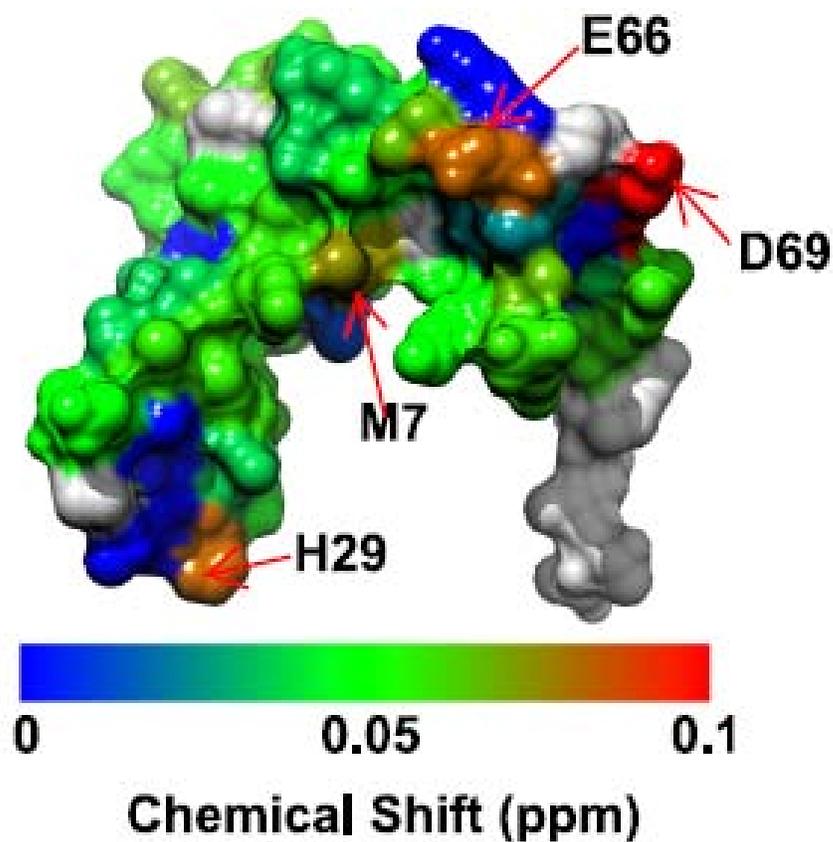


Figure 4.2

The weighted change in chemical shift perturbations ($\Delta\delta = [\Delta\delta^2\text{H}_N + (0.2\Delta\delta\text{N})^2]^{1/2}$) obtained from the ^{15}N ^1H TROSY experiments are mapped on the entire Sox2-HMG surface. Residues which are significantly shifted are depicted by an arrow. The colored spectrum bar displays the extent of NMR chemical shift perturbations in ppm. Unassigned residues are colored in gray

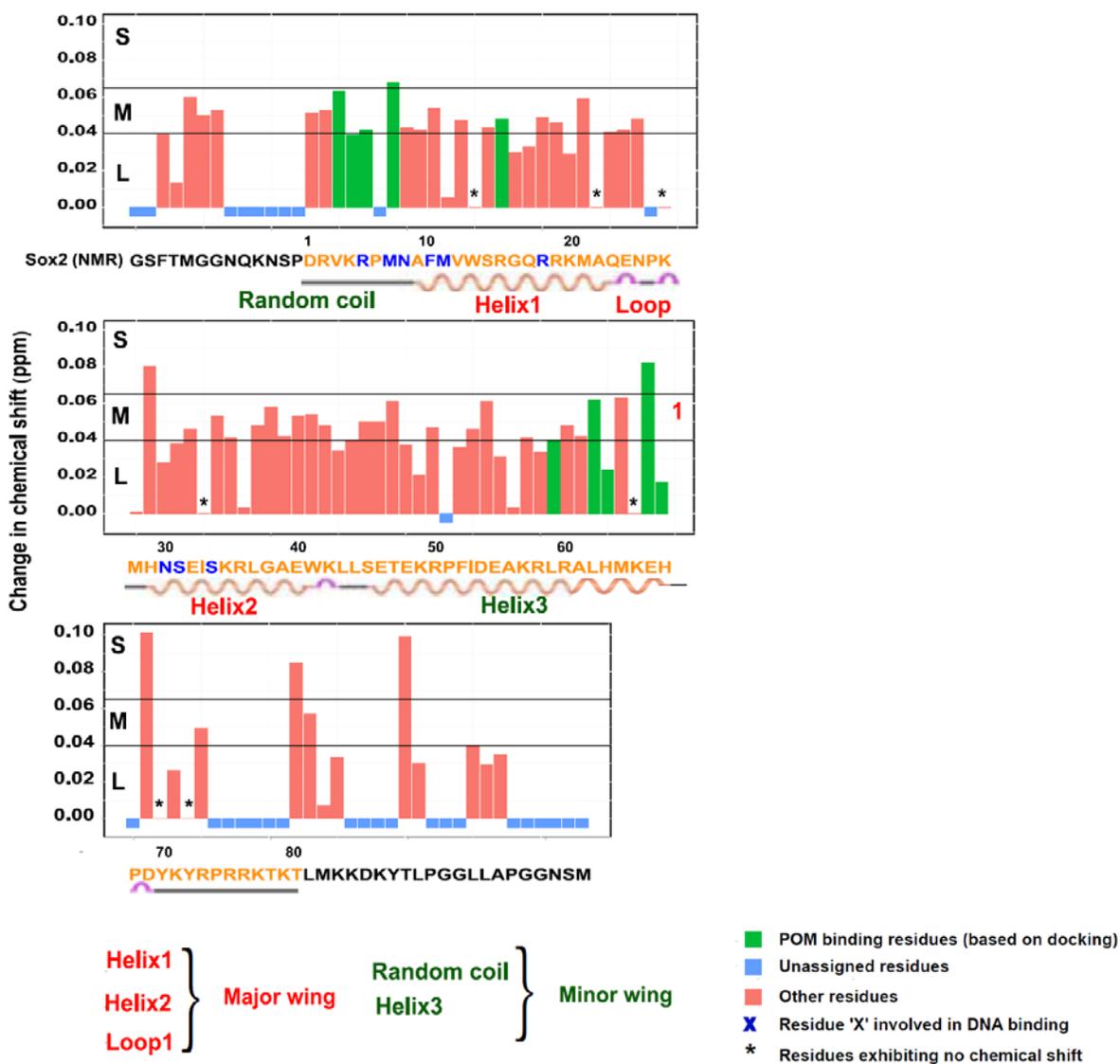


Figure 4.3

Changes in chemical shift ($\Delta\delta = [\Delta\delta^2\text{H}_N + (0.2\Delta\delta\text{N})^2]^{1/2}$) upon POM binding is plotted against the Sox2 amino acid sequence (numbered according to Sox2: 1GT0). Threshold windows indicating Significant (S), Moderate (M) and Low (L) chemical shift perturbations are depicted as straight lines. Green colored bars indicate residues which have been implicated in direct binding to POM based on docking studies. Residues that are unchanged in TROSY are indicated with an asterisk (*). Unassigned residues are colored in light blue and given arbitrary negative chemical shift values solely for data visualization purposes. Sox2-HMG residue sequences involved in DNA binding are colored in blue. Secondary structural elements like alpha helices are named and colored to distinguish whether they belong to the major or minor wing of the HMG domain.

proximity to the DNA binding region (Figure 4.2). Among the significantly shifted residues Met7, Glu66 and Asp69 are in spatial proximity whereas His29 is located distantly in loop 1 of the HMG's major wing (Figure 4.2). The structural role of these significantly perturbed residues in POM binding was investigated further by docking studies (described in the section below) and was aided by comparison with solution structures of Sox17 (PDB:2YUL), Sox5 (PDB:1I11) and the DNA bound structure of Sox2 (PDB:1GT0). Buried core residues like Trp13, Ala22, Lys27 and Ile33 which cause the stacking of helices 1 and 2 of the HMG major wing, were found to be least perturbed upon POM binding, suggesting that the structural integrity of the HMG domain is retained upon inhibitor binding. A majority of the Sox2-HMG residues, however, exhibited moderate chemical shift changes suggesting a POM induced reorganization of the backbone (Figure 4.3). It is well documented that residues showing ^{15}N ^1H TROSY shifts do not necessarily constitute the POM binding site but that such perturbations may also occur at a distance from the actual binding site via indirect and allosteric effects (186,187). To further investigate Sox2-HMGPOM interactions, the chemical shift perturbations were employed to constrain ligand binding sites on the Sox2-HMG surface by docking studies.

4.2 Preferential binding site of the Dawson-POM $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$ on the Sox2-HMG surface

In order to identify the site of interaction of the Dawson-POM $\text{K}_6[\text{P}_2\text{Mo}_{18}\text{O}_{62}]$ with the Sox2-HMG, two autodock searches were undertaken. First, a blind docking search was set-up such that the POM can explore the entire surface of a Sox2-HMG model derived from Sox2/Oct1/DNA crystal structure (70). The second search was restrained by drawing a grid in the spatial zone defined by the cluster of the residues Met7, Glu66 and Asp69 which were identified as being significantly shifted from ^{15}N ^1H TROSY NMR. Both, the blind and NMR chemical shift based

docking calculations identified a common docking site showing that this binding site is highly preferred by POM. The binding pocket is formed by the C-terminal helix-3 and the N-terminal region of the minor wing of the L-shaped Sox2-HMG structure (Figure 4.4A). A comparison of the Sox2-POM docked structure with the X-ray structure of ternary Oct1.Sox2.DNA complex (PDB code: 1GT0) reveals that the binding of POM to this site would interfere with the negative phosphate backbone of DNA by competing with the neighboring positively charged residues in Sox2 (Figure 4.4B).

As the ligand surface is negatively charged, the calculated docking energy was found to be very low (-10.6 kcal/mol) corresponding to a binding affinity of ~50 nM. Such a tight binding affinity presumably also explains the difficulty in obtaining reliable ITC curves as discussed earlier. The negatively charged surface of POM can form many favorable electrostatic interactions within the binding cavity of Sox2 in the docked complex geometry (Figure 4.5A). The binding cavity is formed mostly by basic residues: Lys4, Arg5, Arg15, His63 and His67. The positively charged side chains of these residues can form hydrogen bonds and/or electrostatic interactions with the negatively charged oxygens of POM. The backbone amide of Lys4 forms an additional hydrogen bond with POM (Figure 4.4A). An alternative expanded snap-shot of the docked configuration is provided in Appendix G to provide adequate visualization of the range of electrostatic and hydrogen bond interactions. In accordance with the docking studies, the positively charged residues Lys 4, Arg5 and Arg15 exhibit moderate chemical shift changes while His63 and His67 exhibit insignificant perturbation. Solvent accessibility analysis of DNA free Sox2-HMG (1GT0) reveals that His63 (only 13% solvent exposed) could be classified as a buried residue, while His67 is (30% solvent exposed) a partially exposed residue (Figure 4.5B). The lack of significant chemical perturbation for His63 and His67 is possibly a reflection of the inflexible nature of the

A.

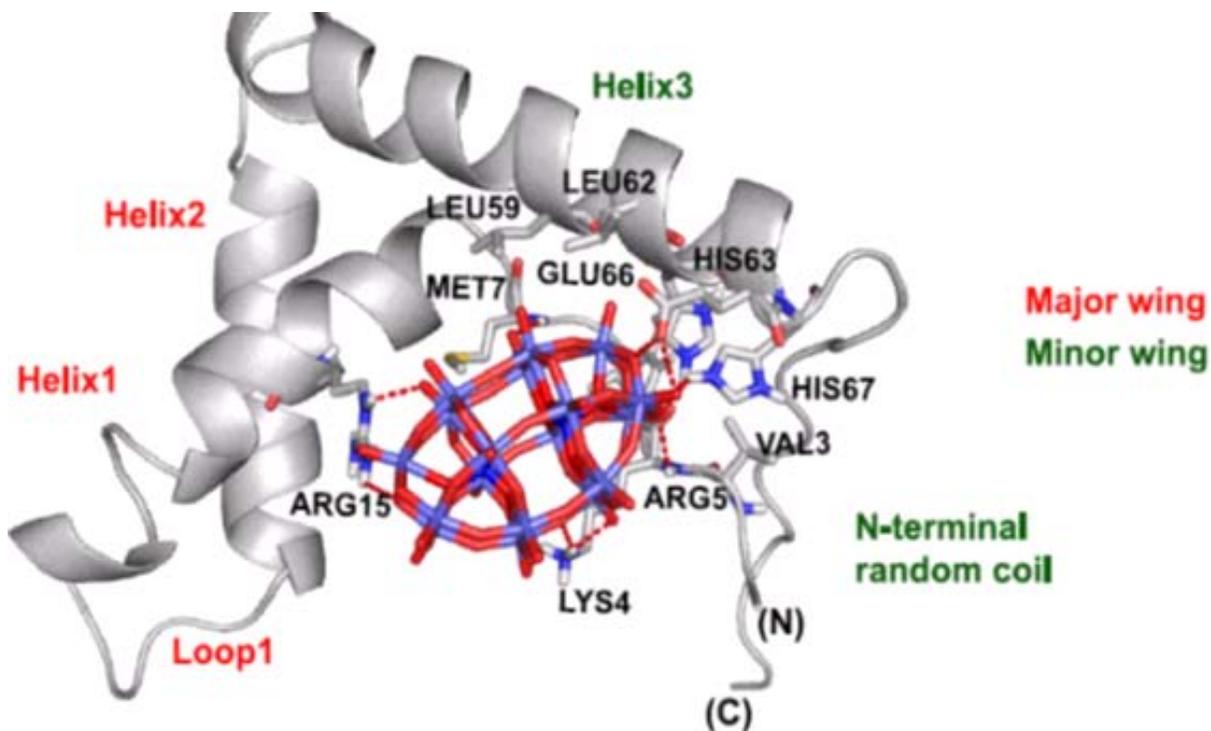


Figure 4.4

A) The lowest energy Sox2-HMG-POM complex structure from the Autodock searches shows the POM positioned within a pocket in the minor wing of the Sox2-HMG structure. Lys4, Arg5, Arg15, His63 and His67 are potentially involved in electrostatic or hydrogen bond interactions. Glu66 can donate hydrogen bond in a protonated form. Leu59, Leu62, Met7 and Val3 contribute to shaping the hydrophobic cavity. The Sox2 structure is shown as cartoon and the interacting amino-acids are shown as sticks. Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ is also shown in stick representation. Hydrogen bonds and electrostatic interactions less than 3.5\AA are shown in red dots. Residue numbering is based on the PDB structure 1GT0.

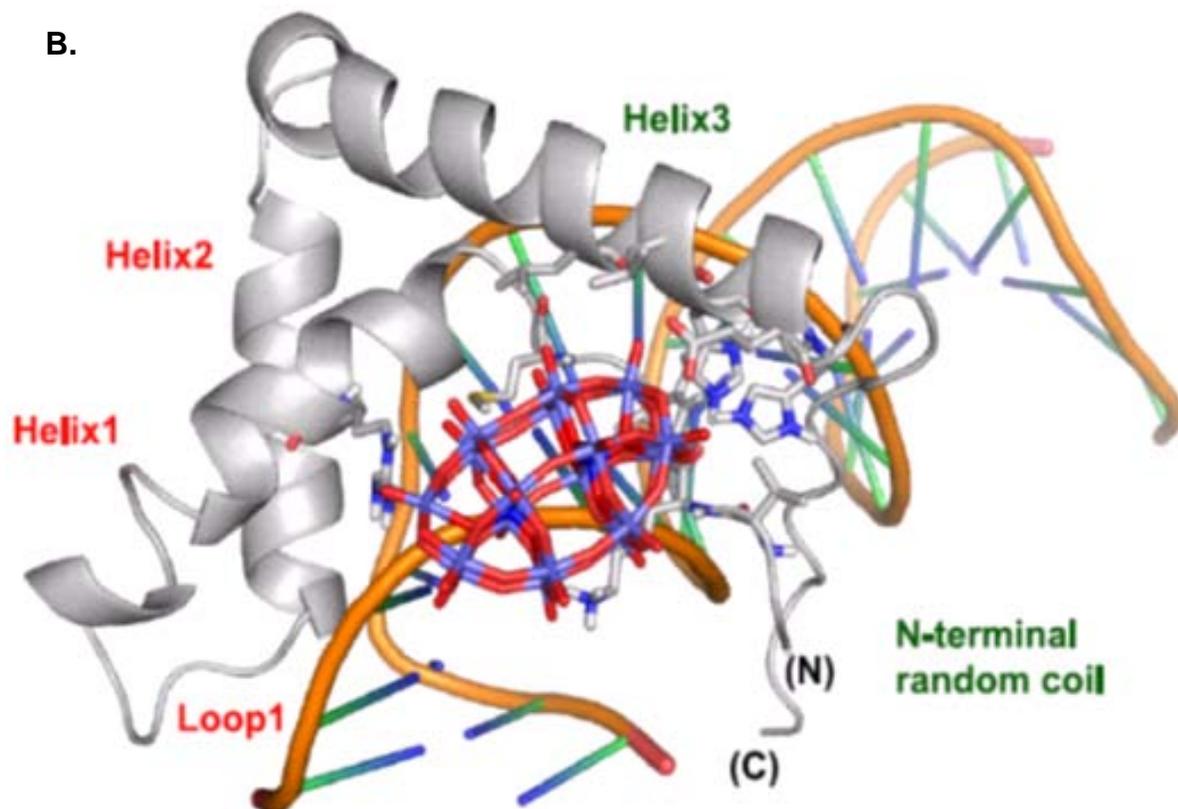


Figure 4.4

B) Comparison of the docked model from Figure 4.4A with the Sox2-DNA complex X-ray structure (1GT0) reveals that binding of POM to this site would directly interfere with DNA binding due to charge repulsion. The Sox2-DNA complex structure is shown as cartoon and the interacting amino-acids of Sox2 is depicted as sticks. The Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ is shown in stick representation.

chemical environment around solvent inaccessible residues. The single acidic residue in the POM binding cavity Glu66, was found to be most strongly shifted following POM addition in the TROSY measurements (Figure 4.2 & Figure 4.3). Even though a protonated state of Glu66 cannot be ruled out due to a slightly altered pH microenvironment, such a state at neutral pH is unlikely. It is more conceivable that POM binding repels Glu66 due to unfavorable negative charges causing it to structurally re-orient with respect to the unbound Sox2-HMG. Asp69 is a residue strongly perturbed in TROSY but is not present in the POM docking site. Asp69 is located in the C-terminal of helix 3 which is a region subject to inherent conformational changes due to protein flexibility (Figure 4.4A). This is evident from NMR studies of Sox4 and Sox5 which reveal that the C-terminus comprises the most dynamic region of the protein (66,188). Following DNA binding however, the C-terminus is markedly rearranged to participate in the DNA interaction(71). Thus, the pronounced perturbation detected for Asp69 likely illustrated the dynamics of this region and its structural adjustments accompanying molecular recognition events. The strongly perturbed His29 is also a residue that is not located in the POM docking site but is positioned in loop 1, connecting helices 1 and 2 (Figure 4.3). This loop was found to exhibit the most pronounced C α RMSD (Root mean square deviation) deviation within the helical core of the HMG domain when the DNA-bound Sox17 was compared with Sox5 and Sox17 NMR structures (71). By inspecting the different conformers of the solution structure of Sox17 (PDB: 2YUL), it is evident that the N-terminal region of the minor wing is very dynamic and includes conformations where the unstructured minor wing N-terminus approaches loop1 indicating a potential for cross-talk between these regions. The Sox2-HMG construct employed in this study contains an N-terminus extended by 13 amino acids as compared to the homologous Sox17 structure (PDB:2YUL). Thus, it is conceivable that a longer dynamic Sox2 N-terminus

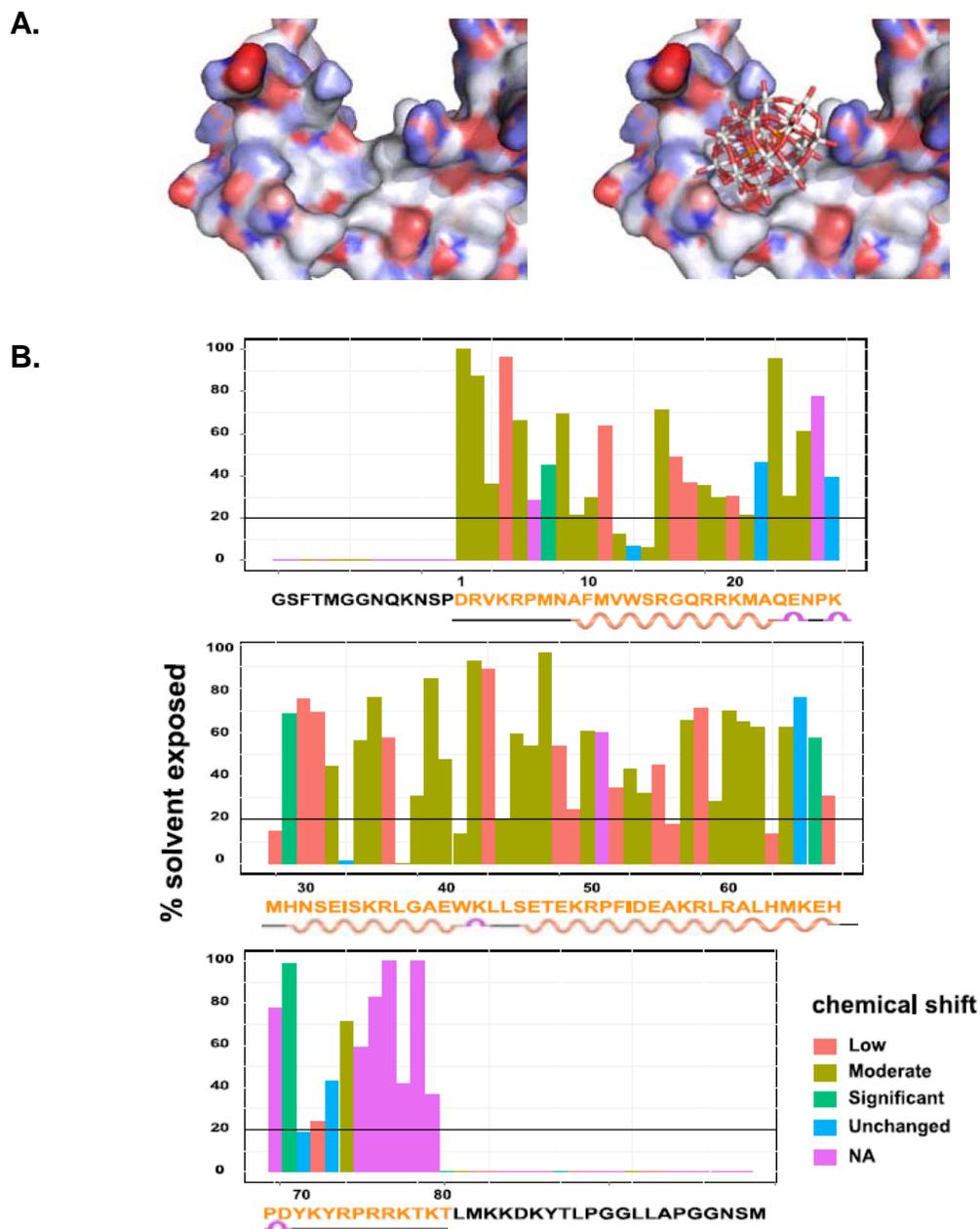


Figure 4.5

- A) Interaction surface of Sox2-HMG colored by PM6 partial charges without and with docked ligand (blue – positive, red – negative).
- B) Solvent accessibility per residue of DNA free Sox2-HMG structure (PDB: 1GT0) is plotted against the Sox2-HMG sequence and bars colored to depict their corresponding chemical shift perturbation category (NA-backbone unassigned residue).

would come into contact with His29 in loop1 due to conformational reorganization induced by POM binding. In addition to the charged amino acids, several hydrophobic residues like Val3, Met7, Leu59 and Leu62 shape the cavity of the POM docking site (Figure 4.4A). Consistently, Met7 exhibited a significant chemical shift in the TROSY experiment while Val3, Leu59 and Leu62 are moderately affected. Together, NMR and docking studies suggest a model illustrating how the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ brings about the inhibition of the Sox2-HMGDNA interaction

CHAPTER 5

RESULTS III

Selectivity of polyoxometalates

5.1 Selectivity studies on the inhibition of Sox-HMG family of TFs by polyoxometalates

To assess the selectivity of a panel of POMs, extensive experiments were carried out on different members of the Sox-HMG family as well as with transcription factors of unrelated structural classes namely Pax6, REST, FoxA1 and AP2 (Full length protein sequences were used for FoxA1 and AP2). Pax6 contains a bi-partite paired DNA binding domain, the REST has eight Cys₂/His₂ Zinc-finger DNA binding domains, Ap2 has a beta-sheet fold and FoxA1 has a forkhead/winged helix DNA binding domain (167,189-191). In total, the residual DNA binding activities of 15 different TFs was estimated against a panel of 15 different polyoxometalates. The panel of polyoxometalates that were tested could be broadly classified into the “Keggin” POMs the “Dawson” POMs, and other simpler POMs like sodium metatungstate and decavanadate. The details of the type of POMs employed in the study, including the type of organic modification is provided in a table along with structures of organic Dawson POMs (Figure 5.1 & Table 5.1). Keggin POMs exert lowered inhibition on the Sox-HMG members as compared to the Dawson POM leading to the observation that the size of inhibitor polyoxometalate is an important consideration in the inhibition of the Sox-HMG family (Figure 5.2). To determine selectivity, fluorescence anisotropy (all Sox-HMGs, REST, AP2 and FoxA1) and EMSA experiments (Pax6) were performed such that the proteins were incubated with their respective cognate fluorescently labeled DNA elements at concentrations resulting in a fractional binding of 70-90%, prior to the addition of various POMs. The residual DNA binding activity is estimated from three to five independent experiments and is reported as a percentage of the control without the inhibitor (0% residual binding activity would correspond to maximum inhibition, while 100% activity would correspond to no inhibition).

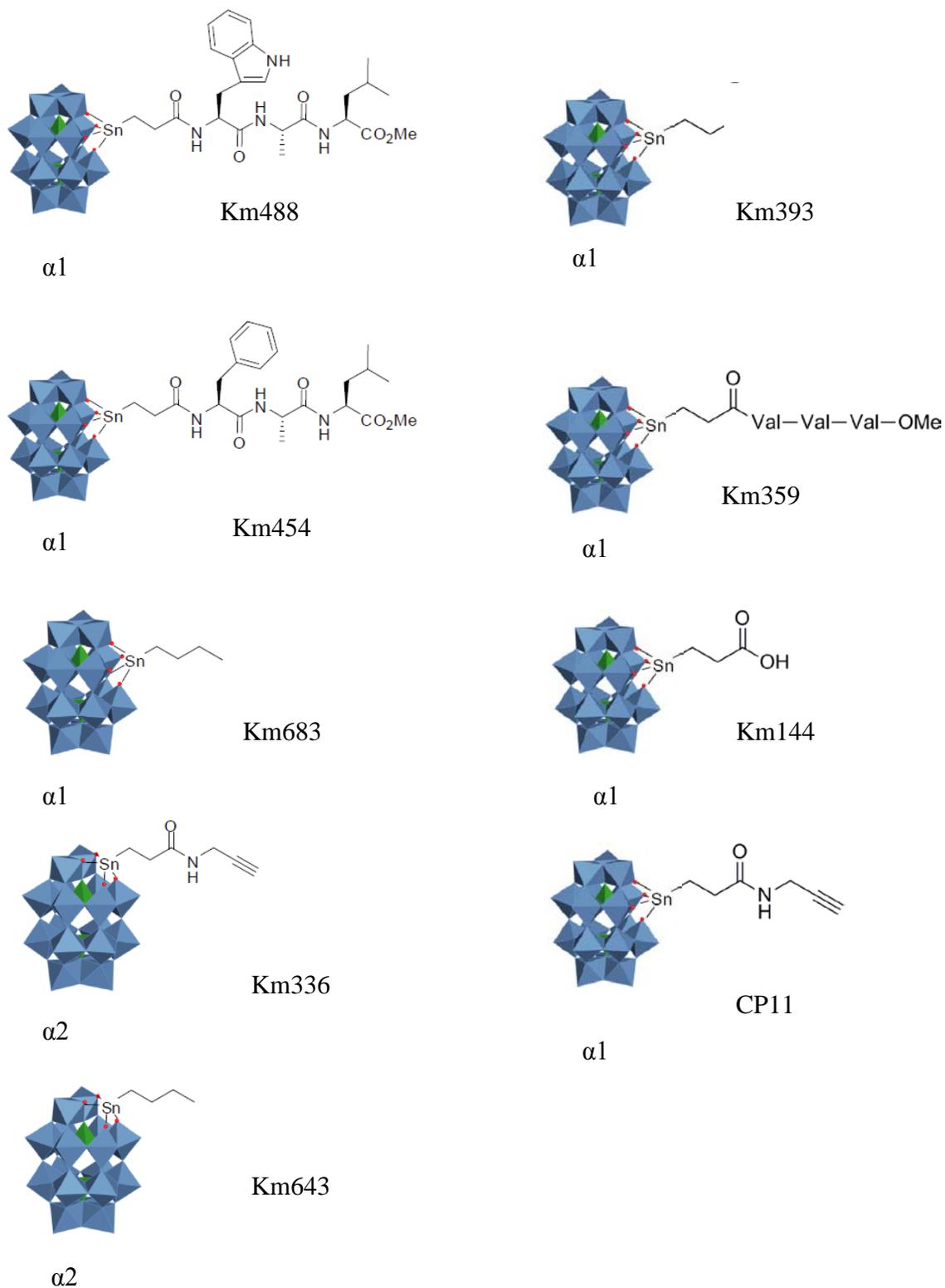


Figure 5.1

The regioselective (α_1/α_2) organic side chains of tin substituted Dawson POMs used in the study.

Table 5.1 Panel of 15 POMs screened for inhibition of DNA binding activity of 15 transcription factors

POMs	Chemical formula	Type	Functional modification	Mol.wt g/mol
km488	$\alpha 1 P_2W_{17}O_{66}SnC_{24}H_{34}N_4(NC_{16}H_{36})_7$	Dawson	Trp-Ala-Leu	6438
km454	$\alpha 1 P_2W_{17}O_{66}SnC_{22}H_{32}(NC_{16}H_{36})_7$	Dawson	Phe-Ala-Leu	6356
km683	$\alpha 1 P_2W_{17}O_{61}SnC_4H_{10}(NC_{16}H_{36})_6$	Dawson	Aliphatic	5794
km643	$\alpha 2 P_2W_{17}O_{61}SnC_4H_{10}(NC_{16}H_{36})_6$	Dawson	Aliphatic	5794
km336	$\alpha 2 P_2W_{17}O_{62}SnC_6H_8N_4(NC_{16}H_{36})_7$	Dawson	Aliphatic	6089
CP11	$\alpha 1 P_2W_{17}O_{62}SnC_6H_8N_4(NC_{16}H_{36})_7$	Dawson	Aliphatic	6089
km393	$\alpha 2 P_2W_{17}O_{62}SnC_6H_{11}N_4O(NC_{16}H_{36})_7$	Dawson	Aliphatic	6135
km359	$\alpha 2 P_2W_{17}O_{62}SnC_{16}H_{23}N_3O_2(NC_{16}H_{36})_7$	Dawson	Val-Val-Val	6357
km633	$H_3V_{10}O_{28}(NC_{16}H_{36})_3$	Decavandate	Vanadium	1688
km144	$\alpha 2 P_2W_{17}O_{62}SnC_3H_4O_2(NC_{16}H_{36})_6$	Dawson	Cyclic aliphatic	5810
DpomMo	$K_6[Mo_{18}O_{62}P_2]$	Dawson	None	3016
DpomW18	$K_6[W_{18}O_{62}P_2]$	Dawson	None	4597
KeggMo	$[PMo_{12}O_{40}]^{3-}$	Keggin	None	1892
KeggW	$[PW_{12}O_{40}]^{3-}$	Keggin	None	2946
NaMeta	$3Na_2WO_4 \cdot 9WO_3 \cdot xH_2O$	Sodium metatungstate	None	2969

5.1.1 Unmodified Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ and decavandate $H_3V_{10}O_{28}$ are relatively selective TF inhibitors

The average residual DNA binding activity of 15 different TFs against a panel of POMs from three-five independent experiments was used in a clustering analysis using an “R heatmap.2” package. From the clustering analysis it could be clearly observed that the Keggin and Dawson class of polyoxometalates exhibit a marked dichotomy in their selectivity and inhibition potential of the Sox-HMG family of transcription factors (Figure 5.2). The Dawson POMs were found to be highly potent in their inhibition profiles of not only the Sox-HMG members but also other TF families like FoxA1, REST and AP2 (Table 5.2). Modified organic Dawson POMs were tested under the assumption that their functionalized organic side-chains would be capable of exploiting subtle differences in the Sox-HMG domain sequences to exhibit Sox family specific inhibition. Contrary to expected behavior, the modified Dawson polyoxometalates did not bring about a commensurate change in selective discrimination of the HMG family members. The modified Dawson polyoxometalates invariably amplified the inhibitory potency of the pristine “Dawson” scaffold against various Sox-HMG members (Figure 5.3). The potency of the organically modified Dawson-POMs in inhibiting the Sox-HMG family merits investigation through crystallographic/NMR methods for a better understanding of the physico-chemical nature of these interactions. The organically modified Dawson-POMs while being moderately effective against AP2 was also found to be very effective against FoxA1. Among the transcription factors tested it was observed that Pax6 was the most inert to POM treatment, with a majority of the POMs not having any effect on the Pax6 paired domain activity tested using EMSA (Appendix H). The only POMs that were relatively effective in inhibiting Pax6 activity were KM393, KM359 and the Keggin tungstate (Figure 5.2 & Table 5.2). Overall, it could be observed that the

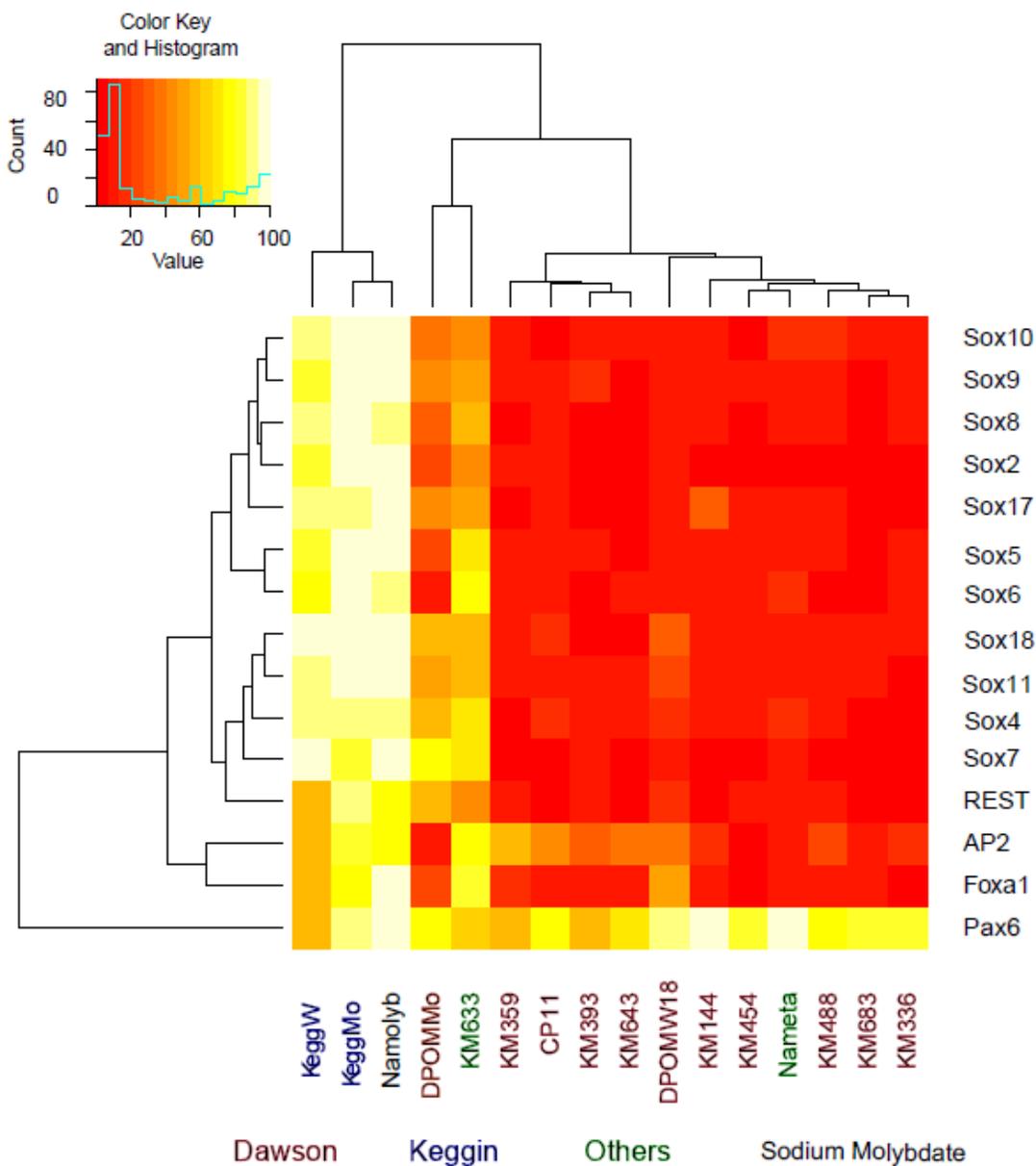


Figure 5.2

A heatmap of the average residual DNA binding activity ('value' in %) of 15 different TFs against a panel of 15 POMs, clustered by their inhibition profiles (Red color indicates higher inhibition, while yellow color indicates relatively lesser inhibition). Keggin POMs exert lowered inhibition on the Sox-HMG members leading to the observation that the size of polyoxometalate is an important consideration in the inhibition of the Sox-HMG family. Inhibitor compounds are color coded according to their polyoxometalate class as indicated in Table 5.1 (Dawson, Keggin or other simpler POMs like decavandate and Sodium metatungstate).

Table 5.2 Residual DNA binding activity (in %) of 15 TFs against 15 different POMs from three-five independent experiments expressed as mean \pm standard deviation

	Dawson							
	KM488	KM454	KM683	KM336	KM643	CP11	KM393	KM359
Pax6	75 \pm 1	86 \pm 1	82 \pm 2	81 \pm 2	70 \pm 2	76 \pm 4	59 \pm 6	56 \pm 9
AP2	21 \pm 10	6 \pm 4	15 \pm 10	17 \pm 15	39 \pm 17	45 \pm 20	30 \pm 15	55 \pm 22
REST	10 \pm 5	9 \pm 2	6 \pm 6	4 \pm 3	6 \pm 4	4 \pm 6	9 \pm 3	10 \pm 5
Foxa1	12 \pm 16	4 \pm 5	10 \pm 9	4 \pm 5	9 \pm 4	10 \pm 7	8 \pm 7	18 \pm 13
Sox2	8 \pm 3	6 \pm 4	4 \pm 5	7 \pm 6	6 \pm 4	12 \pm 6	6 \pm 5	10 \pm 2
Sox4	14 \pm 10	12 \pm 6	4 \pm 3	6 \pm 6	10 \pm 2	18 \pm 3	12 \pm 7	4 \pm 4
Sox5	11 \pm 7	10 \pm 8	8 \pm 6	9 \pm 8	6 \pm 6	12 \pm 5	10 \pm 5	10 \pm 5
Sox6	7 \pm 8	13 \pm 3	2 \pm 3	9 \pm 5	11 \pm 4	10 \pm 3	7 \pm 4	11 \pm 4
Sox7	8 \pm 2	7 \pm 6	3 \pm 4	2 \pm 1	8 \pm 6	7 \pm 3	9 \pm 2	5 \pm 4
Sox8	12 \pm 6	8 \pm 4	7 \pm 5	12 \pm 7	6 \pm 4	11 \pm 7	6 \pm 7	7 \pm 4
Sox9	10 \pm 6	10 \pm 8	6 \pm 6	10 \pm 5	7 \pm 5	10 \pm 8	15 \pm 5	11 \pm 8
Sox10	15 \pm 3	7 \pm 5	8 \pm 6	9 \pm 7	10 \pm 6	5 \pm 5	10 \pm 5	11 \pm 8
Sox11	12 \pm 4	10 \pm 5	9 \pm 6	5 \pm 10	9 \pm 7	11 \pm 3	8 \pm 5	10 \pm 5
Sox17	9 \pm 8	8 \pm 7	5 \pm 5	6 \pm 5	7 \pm 3	10 \pm 4	7 \pm 5	7 \pm 3
Sox18	14 \pm 5	14 \pm 8	10 \pm 5	9 \pm 7	2 \pm 3	17 \pm 8	3 \pm 3	10 \pm 4

	Dawson			Other		Keggin		Salt
	KM144	DpomW18	DpomMo	MetaW	KM633	KeggMo	KeggW	Namolyb
Pax6	100 \pm 0	93 \pm 5	77 \pm 2	99 \pm 1	64 \pm 5	93 \pm 2	56 \pm 4	100 \pm 0
AP2	19 \pm 6	35 \pm 16	13 \pm 2	10 \pm 11	74 \pm 21	83 \pm 14	60 \pm 17	76 \pm 19
REST	6 \pm 5	15 \pm 3	57 \pm 26	11 \pm 4	41 \pm 6	88 \pm 7	59 \pm 26	76 \pm 31
Foxa1	9 \pm 5	48 \pm 15	26 \pm 9	11 \pm 2	81 \pm 13	77 \pm 19	59 \pm 9	96 \pm 6
Sox2	7 \pm 5	12 \pm 10	25 \pm 21	8 \pm 3	41 \pm 16	95 \pm 5	81 \pm 9	95 \pm 4
Sox4	9 \pm 10	17 \pm 4	54 \pm 19	15 \pm 9	70 \pm 12	91 \pm 17	89 \pm 8	91 \pm 9
Sox5	13 \pm 8	12 \pm 4	24 \pm 15	12 \pm 8	73 \pm 24	99 \pm 3	84 \pm 5	95 \pm 4
Sox6	10 \pm 5	11 \pm 4	14 \pm 6	17 \pm 3	77 \pm 9	95 \pm 6	80 \pm 8	90 \pm 9
Sox7	6 \pm 5	13 \pm 5	74 \pm 18	8 \pm 2	69 \pm 19	81 \pm 27	95 \pm 5	98 \pm 3
Sox8	9 \pm 5	14 \pm 2	29 \pm 16	11 \pm 4	56 \pm 10	96 \pm 2	89 \pm 5	92 \pm 2
Sox9	10 \pm 8	11 \pm 9	43 \pm 8	10 \pm 5	48 \pm 11	94 \pm 5	85 \pm 7	96 \pm 5
Sox10	11 \pm 2	14 \pm 6	37 \pm 14	16 \pm 8	46 \pm 22	96 \pm 2	87 \pm 8	94 \pm 6
Sox11	12 \pm 6	25 \pm 12	53 \pm 33	10 \pm 4	59 \pm 27	97 \pm 3	92 \pm 6	96 \pm 4
Sox17	28 \pm 39	12 \pm 8	43 \pm 39	9 \pm 7	48 \pm 25	92 \pm 3	93 \pm 6	94 \pm 5
Sox18	14 \pm 8	30 \pm 14	61 \pm 14	8 \pm 2	55 \pm 18	97 \pm 5	94 \pm 4	95 \pm 5

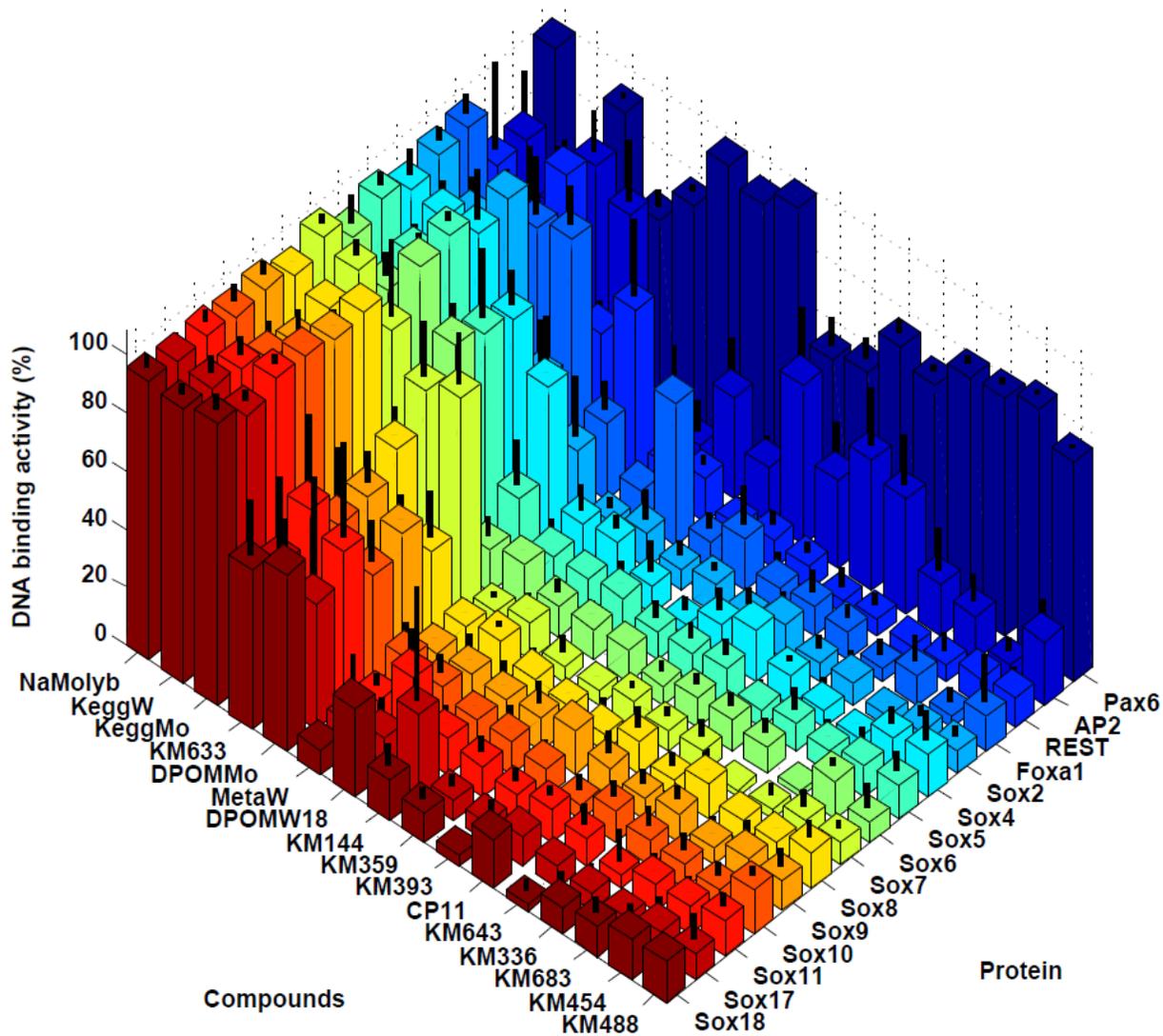


Figure 5.3

A 3D bar plot depiction of the selectivity study of 15 TFs against a panel of 15 POMs from three-five independent experiments

Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ and the decavanadate, KM633 ($H_3V_{10}O_{28}$) were the only two POMs that exhibited a rather diverse and therefore relatively selective inhibitory pattern against the panel of 15 TFs (Figure 5.4). The diversity between the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ and the decavanadate KM633 ($H_3V_{10}O_{28}$) is most evident from their inhibition potential of the same Sox-HMG member. For example, the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ inhibits Sox6 so that it retains only $14 \pm 6\%$ of its DNA binding activity, whereas decavanadate is least effective against Sox6 retaining $77 \pm 9\%$ of its DNA binding activity highlighting the diversity in the inhibitory potential between the Dawson-POM and the decavanadate against a Sox-HMG member (Table 5.2). It must be mentioned at this juncture that decavanadates are known to have established properties of interaction with redox and hydrolytic enzymes and is also known to affect calcium pumps and the function of actin and myosin(138). However, one of the limitations of the decavanadate, is that it has a limited chemistry with respect to being modified by organic groups when compared to the Dawson-POMs(139). The selectivity potential of polyoxometalates is most evident from the inhibition potential of the same polyoxometalate against different Sox-HMG members. For example, it can be observed that the group F members Sox7 and Sox18 were the most inert of all Sox-HMG proteins tested against $K_6[P_2Mo_{18}O_{62}]$, retaining $74 \pm 18\%$ and $61 \pm 14\%$ of their DNA binding activity respectively, highlighting the differential interaction between different Sox-HMG members against the same inhibitory polyoxometalate $K_6[P_2Mo_{18}O_{62}]$ (Table 5.2 and Figure 5.4). The inhibition potential of $K_6[P_2Mo_{18}O_{62}]$ on Group F members Sox7 and Sox18 differs significantly from inhibition of the group B member Sox2, as evaluated by a Student's t-test (Appendix I). This relatively selective interaction behavior of $K_6[P_2Mo_{18}O_{62}]$ could not however be extended to the other Group F member Sox17 because of the high experimental noise in Sox17 DNA binding activity estimation. Sox17 was found to retain $43 \pm 39\%$ of its

DNA binding activity with $K_6[P_2Mo_{18}O_{62}]$ treatment. This experimental noise could be because of potential intrinsic instability of Sox17 over the course of experiment (eg; gradual loss of DNA binding activity due to mis-folding). Knowledge of the putative binding site of $K_6[P_2Mo_{18}O_{62}]$ on Sox2 reveals potential reasons for the observed differential interaction of this POM with the other Group F members Sox7 and Sox18. Firstly, multiple sequence alignment reveals consistent differences between the Group F members (Sox7 and Sox18) and Sox2 in at least six out of ten amino-acid positions that has been proposed to bind to the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ (Figure 5.5). A distinct relationship between the substitution of an amino-acid and its corresponding effect on $K_6[P_2Mo_{18}O_{62}]$ binding to different Sox-HMG family members can be discerned by observing that the amino-acid position corresponding to Val3 and Leu62 (PDB: 1GT0 numbering) in Sox2 is replaced by Ile3 and Gln62 in Sox7 and Sox18. Val3 along with Met7, Leu59 and Leu62 are involved in shaping the hydrophobic cavity of Sox2 for the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ to bind (Figure 4.4A). It can be envisaged that substitution of Val3 and Leu62 in Sox2 by the homologous amino-acids (Ile3 and Gln62) in Sox7 and Sox18 leads to an altered shape in the hydrophobic Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ binding cavity. Similarly, the substitution of Lys4 and Arg15 in Sox2 by Arg4 and Lys15 in Sox7 and Sox18 could be envisaged to provide an altered electrostatic and hydrogen bonding environment for the $K_6[P_2Mo_{18}O_{62}]$ binding. Thus it can be expected the pristine Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ is sensitive to binding shape cavity and electrostatics of the Sox-HMG member (Figure 5.5 and 4.4A).

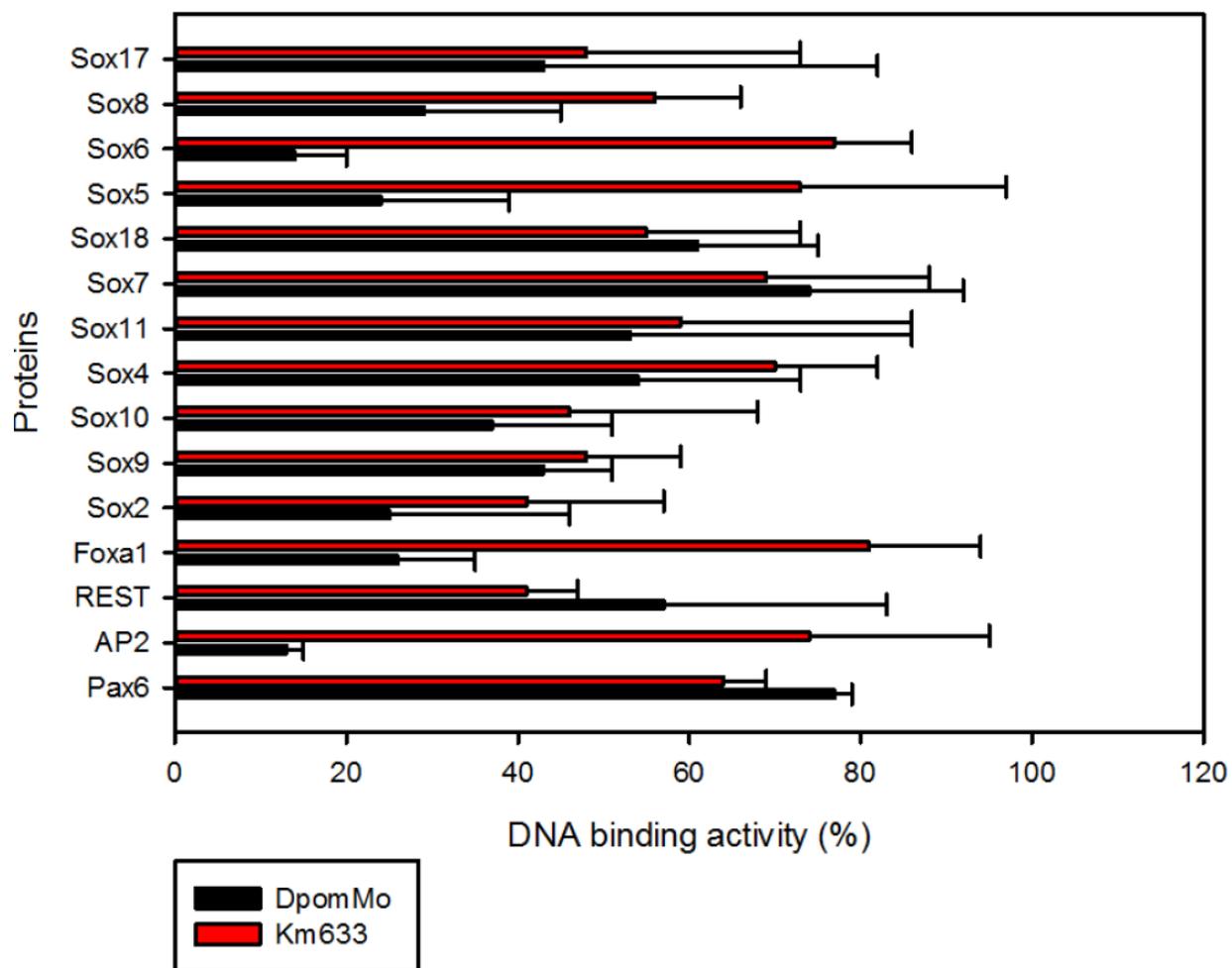


Figure 5.4

A 2D bar plot extract of figure 5.3 depicting the diverse and relatively selective inhibition effect of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ and KM633 ($H_3V_{10}O_{28}$) in inhibiting a panel of 15 TFs from three-five independent experiments

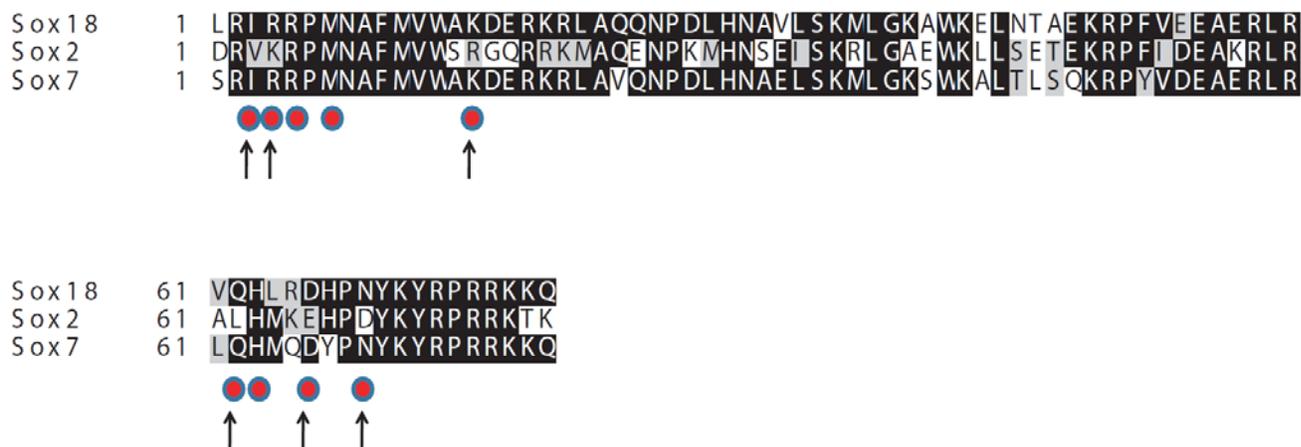


Figure 5.5

Multiple sequence alignment of the core HMG-domain reveals differences between the Group F members (Sox7 and Sox18) and Sox2 in 6 out of 10 amino-acid positions proposed to be involved in $K_6[P_2Mo_{18}O_{62}]$ binding with Sox2. Residues potentially involved in $K_6[P_2Mo_{18}O_{62}]$ binding based on docking studies with Sox2 are indicated by red dots. Homologous Sox-HMG residue positions involved in POM binding that exhibit consistent differences between the Group F members (Sox7 and Sox18) and Sox2 across the sequence alignment are indicated by an arrow. Numbering is based on Sox2 structure from PDB: 1GT0.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

6.1 Summary of Results

1. A high throughput fluorescence anisotropy screen of a library of 2,189 structurally and mechanistically diverse compounds for inhibitors of Sox2-HMG DNA binding activity resulted in the identification of a nanomolar wells-Dawson-Polyoxometalate $K_6[P_2Mo_{18}O_{62}]$ inhibitor.
2. The Dawson-Polyoxometalate $K_6[P_2Mo_{18}O_{62}]$ physically interacts with the Sox2-HMG domain and inhibits DNA binding activity with an IC_{50} of 98.6 ± 22.1 nM.
3. Preliminary selectivity studies of $K_6[P_2Mo_{18}O_{62}]$ revealed that this Dawson-POM can discriminate between Sox2-HMG, Pax6-paired domain and the REST C2H2 DNA binding domain surfaces in its inhibitory potential.
4. NMR and docking studies suggest that $K_6[P_2Mo_{18}O_{62}]$ binds to a binding pocket formed by the C-terminal helix-3 and the N-terminal region of the minor wing of the L-shaped Sox2-HMG structure.
5. A comprehensive selectivity panel consisting of 15 different TFs against a panel of 15 different polyoxometalates revealed a marked dichotomy in the selectivity and inhibition potential of the Dawson and Keggin polyoxometalates.
6. Dawson polyoxometalates modified with organic moieties were found to invariably amplify the inhibitory potency of the pristine “Dawson” scaffold against all the Sox-HMG members exhibiting a rather low selectivity in its inhibitory potential.

6.2 Mechanism of Sox-HMG inhibition by the Dawson-POM

$K_6[P_2Mo_{18}O_{62}]$

The DNA binding domains of transcription factors have so far been considered too impervious to be tackled as drug targets although upregulated transcription factors are a major cause of cancer and other diseases (21,24,192). A Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ has been identified as an unconventional but potent scaffold to inhibit the DNA binding activity of Sox2 from this study. Based on NMR and autodock experiments, a structural mechanism of inhibition of Sox2-HMG by the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ has been proposed. Firstly, a comparison of the Sox2-POM docked structure with the X-ray structure of ternary Oct1.Sox2.DNA complex (PDB code: 1GT0) reveals that the binding of POM to this site would interfere with the negative phosphate backbone of DNA by competing with the neighboring positively charged residues in Sox2 (Figure 4.4B and Figure 6.1). Secondly, the binding of POM could induce structural rearrangements of the N-terminal Sox2-HMG minor wing favoring a closed conformation of the major and minor wings. Taken together, the POM induced repulsion of the negatively charged DNA backbone followed by a preferential closure of the wings of the HMG domain would cause structural rearrangements that would result in exclusion of the DNA from the HMG scaffold. The mode of interaction of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ involves predominantly electrostatic interactions at the pocket just outside of the DNA binding region but still adequately positioned to interfere with the DNA binding residues to effectively compete for the negatively charged DNA backbone (Figure 6.1). While obtaining an atomic structure of a Sox2-HMG/POM complex would provide valuable insights to test the proposed structural mechanism of POM inhibition, it must be mentioned that one of the challenges that the current study faced was obtaining crystallographic complexes of Sox2 with either one of the wells-Dawson

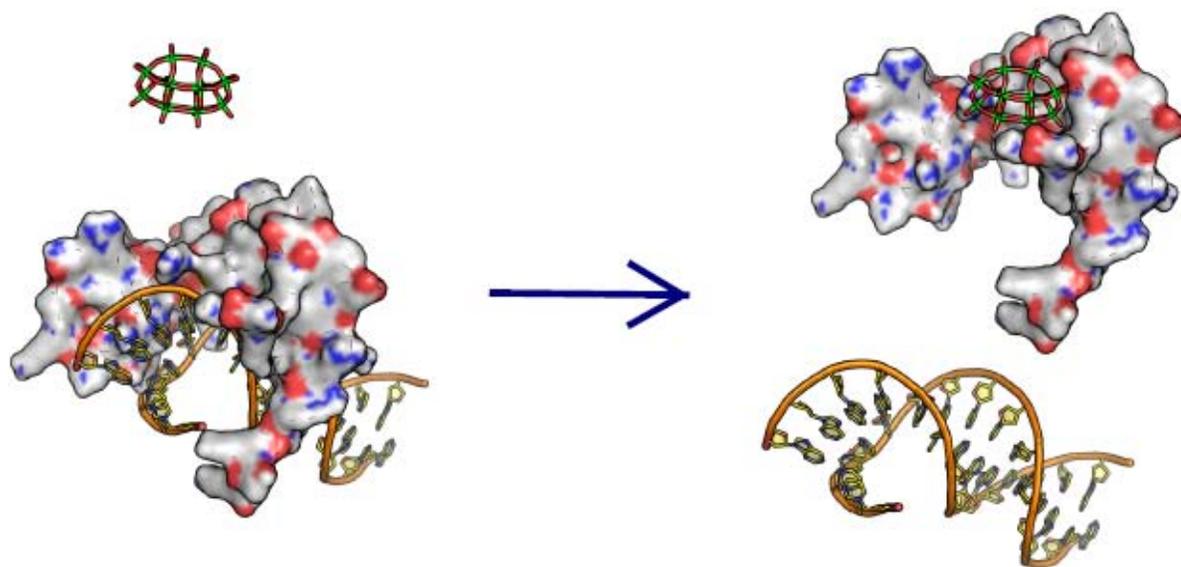


Figure 6.1

A schematic representation of Sox2-HMG bound DNA complex inhibited by the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$

polyoxometalates $K_6[P_2Mo_{18}O_{62}]$ or $K_6[P_2W_{18}O_{62}]$. Crystallization of Sox2 with $K_6[P_2Mo_{18}O_{62}]$ had the problem that the Dawson phosphomolybdate $K_6[P_2Mo_{18}O_{62}]$ is intrinsically unstable for longer time periods in aqueous buffers and breaks down into fragmentary POMs, while the relatively hydrolysis stable Dawson phosphotungstate $K_6[P_2W_{18}O_{62}]$ had the shortcoming of being insoluble in aqueous buffers at higher concentrations that would be necessary for crystallization.

6.3 Future experiments to test the mechanism of Sox2-HMG inhibition by the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$

The biophysical mechanism that has been proposed to explain the inhibition of Sox2-HMG by $K_6[P_2Mo_{18}O_{62}]$ can be tested by a series of precise biochemical mutation experiments. Firstly, critical residues in Sox2-HMG that are potentially responsible for making hydrogen bond /electrostatic contacts with $K_6[P_2Mo_{18}O_{62}]$ like Lys4, Arg5, Met7, Arg15, His63 and His67 can be mutated into alanine by either point or combinatorial mutations. These mutations can be envisaged to have only a minimal effect on the DNA binding activity of Sox2-HMG, but could have a significant impact on the inhibitory potential (higher IC_{50} values) of $K_6[P_2Mo_{18}O_{62}]$ and can be verified from fluorescence anisotropy experiments. Conversely, mutations of His29 and Asp69 into alanine should not have any direct effect on the inhibitory potential of $K_6[P_2Mo_{18}O_{62}]$ as these residues are expected to be perturbed only due to allosteric/secondary binding effects. Summarily, the above biochemical mutation studies can be expected to validate or falsify the biophysical mechanism proposed in this study. Secondly, stopped flow kinetics experiments or surface plasmon resonance based experiments could be carried out to estimate parameters like K_{on} and K_{off} in the presence and absence of the inhibitor to assess whether the inhibitory mechanism is competitive, non-competitive or un-competitive type inhibition.

6.4 Assessment of the selectivity of Dawson-polyoxometalates

Firstly, a detailed investigation on a panel of different transcription factors against an expanded set of various polyoxometalates reveal that the Keggin and Dawson class of polyoxometalates exhibit a marked dichotomy in their selectivity. This leads to the observation that the size of inhibitor polyoxometalate is an important consideration in the inhibition of the Sox-HMG family. Secondly, the functionally modified Dawson POMs do not show any selective inhibition pattern when tested against a panel of transcription factors. The amplified potency of the modified Dawson POMs clearly seem to be due to physical interactions of the organic side chains with the transcription factors when compared to the potential of the unmodified Dawson polyoxometalate alone. While the idea of achieving selectivity using derivatized POMs to various classes of DNA binding is in principle a promising approach, the outcome of testing a limited sub-set of modified Dawson POMs indicate substantial methodological challenges involved in achieving selectivity against different TF DNA binding domains. Another important methodological issue that should be considered in these studies is the effect of the buffer composition. A recent study on HIV-1 protease inhibition by organic Dawson POMs has shown that inhibitory profiles are to a large extent affected by the buffer composition, namely the concentration of DMSO and the type of counter ion (160). Future studies could be performed to take this into account when studying TF inhibition by different Dawson-POMs. The inhibitory chemistry of the organically modified Dawson-POMs towards the Sox-HMG family though challenging, merits further investigation through crystallographic/NMR methods for a better understanding of the physico-chemical nature of these interactions. Such a detailed structural study may help in exploiting the utility of the Dawson polyoxometalate scaffold for inhibition of a wide variety of transcription factors. Thirdly, the Dawson scaffold $K_6[P_2Mo_{18}O_{62}]$ seems to

exhibit a subtle difference in inhibition of Sox7 and Sox18 over other Sox-HMG members like the Group B Sox2 (Figure 5.4 and Table 5.2). Based on previous NMR and auto-docking studies it was observed that there are consistent differences between Sox7 and Sox18 with Sox2 in six out of ten amino-acid positions known to bind to the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ leading to a subtle change in the surface interaction chemistry potentially accounting for the observed difference in the inhibition of Sox2 as compared to Sox7 and Sox18 (Figure 5.5).

6.5 Potential Strategies that can be adapted from polyoxometalate based inhibition chemistry to target the DBDs of Transcription factors

Traditional drug discovery approaches from high throughput chemical screening of compound libraries on a target by target basis followed by development of primary hits into leads by organic synthesis have only been modestly successful in the evolution of general strategies for targeting different structural classes of transcription factors. Even though targeting the DNA binding domain of TFs is challenging, evolving a core inhibitor chemistry that can be functionally customized to specifically target the DNA binding domains of different of TFs will have a significant impact in the pharmacology of targeting TFs. The development of click chemistry to incorporate a range of organic substrates in tin substituted Dawson POMs by reacting with monovacant $\alpha 1/\alpha 2-[P_2W_{17}O_{61}]^{10-}$ lacunary POMs has expanded the horizon of organic modifications that could be achieved in polyoxometalate chemistry(139,144,145,178). Future polyoxometalate chemistry can be expected to progress towards even more complex conjugation steps wherein natural biological molecules like carbohydrates, steroids and peptides can be incorporated . It can be therefore envisaged that subsets of transcription factor interaction domains or whole peptides can be conjugated with synthetic Dawson polyoxometalates to target regulatory or DNA binding domains in a truly specific way. For example it can be conceived that

designing specific recognition peptides capable of interacting with the Sox2-HMG surface could be linked with the inorganic framework of Dawson-POMs to potentially achieve a very targeted inhibition of Sox2-HMG.

The only comparable effort to target systematically the DNA binding domains of transcription factors with a core inhibition chemistry is the usage of nucleic acid decoy-like aptamers (65). Just as with POMs, however aptamers suffer from the disadvantage that introducing large charged molecules into cells can be a variable process (65). Bulk production of aptamers could also be expensive (100-500 USD/g) as compared to synthesis of polyoxometalates which generally require only cheap salts like sodium molybdate/tungstate, phosphoric acid and an organic substrate of interest to link to the inorganic framework (193).

Overall, the inhibitory mechanism of the Dawson-POM demonstrated here could eventually spawn the development of novel drugs that could incorporate strategies from the POM based inhibition chemistry of TFs, to specifically combat aberrant gene expression. The study has demonstrated that challenging molecular architectures like the DNA binding domains of transcription factors are tractable drug targets and that Dawson polyoxometalates can be proposed to act as molecular scaffolds for transcription factor inhibition.

REFERENCES

1. Kiefer, J.C. (2007) Back to basics: Sox genes. *Dev Dyn*, **236**, 2356-2366.
2. Huson, D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68-73.
3. Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, **30**, 3059-3066.
4. Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J. and Nadon, R. (2006) Statistical practice in high-throughput screening data analysis. *Nat Biotechnol*, **24**, 167-175.
5. Pope, A., Thor, M., (1994) *Polyoxometalates: From Platonic Solids to Anti-retroviral Activity*. 3 ed. Springer.
6. Philips, R., Kondev, J., Theriot, R., J. (2009) *Physical Biology of the cell*. Garland Science, Taylor & Francis group, New York.
7. Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., Matsudaira, P., (2007) *Molecular Cell Biology*. W.H. Freeman.
8. Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D.R., Afzal, V. *et al.* (2008) Human-specific gain of function in a developmental enhancer. *Science*, **321**, 1346-1350.
9. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, **467**, 430-435.
10. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem*, **79**, 233-269.
11. Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X. *et al.* (2009) Diversity and Complexity in DNA Recognition by Transcription Factors. *Science*.
12. Morris, Q., Bulyk, M.L. and Hughes, T.R. (2011) Jury remains out on simple models of transcription factor specificity. *Nat Biotechnol*, **29**, 483-484.
13. Jauch, R., Ng, C.K., Narasimhan, K. and Kolatkar, P.R. (2011) Crystal structure of the Sox4 HMG/DNA complex suggests a mechanism for the positional interdependence in DNA recognition. *Biochem J*.
14. Nutiu, R., Friedman, R.C., Luo, S., Khrebtukova, I., Silva, D., Li, R., Zhang, L., Schroth, G.P. and Burge, C.B. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat Biotechnol*, **29**, 659-664.
15. Berger, M.F. and Bulyk, M.L. (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc*, **4**, 393-411.
16. Zhao, Y., Granas, D. and Stormo, G.D. (2009) Inferring binding energies from selected binding sites. *PLoS Comput Biol*, **5**, e1000590.
17. Maerkl, S.J. and Quake, S.R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, **315**, 233-237.

18. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.
19. Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G. and Chothia, C. (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res*, **28**, 257-259.
20. Weirauch, M.T. and Hughes, T.R. (2011) A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell Biochem*, **52**, 25-73.
21. Koehler, A.N. (2010) A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol*, **14**, 331-340.
22. Darnell, J.E., Jr. (2002) Transcription factors as targets for cancer therapy. *Nat Rev Cancer*, **2**, 740-749.
23. Cai, W., Hu, L. and Foulkes, J.G. (1996) Transcription-modulating drugs: mechanism and selectivity. *Curr Opin Biotechnol*, **7**, 608-615.
24. Berg, T. (2008) Inhibition of transcription factors with small organic molecules. *Curr Opin Chem Biol*, **12**, 464-471.
25. Jung, D., Choi, Y. and Uesugi, M. (2006) Small organic molecules that modulate gene transcription. *Drug Discov Today*, **11**, 452-457.
26. Koehler, A.N., Shamji, A.F. and Schreiber, S.L. (2003) Discovery of an inhibitor of a transcription factor using small molecule microarrays and diversity-oriented synthesis. *J Am Chem Soc*, **125**, 8420-8421.
27. Arndt, H.D. (2006) Small molecule modulators of transcription. *Angew Chem Int Ed Engl*, **45**, 4552-4560.
28. Zhang, Z., Burch, P.E., Cooney, A.J., Lanz, R.B., Pereira, F.A., Wu, J., Gibbs, R.A., Weinstock, G. and Wheeler, D.A. (2004) Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome. *Genome Res*, **14**, 580-590.
29. Riggs, B.L. and Hartmann, L.C. (2003) Selective estrogen-receptor modulators -- mechanisms of action and application to clinical practice. *N Engl J Med*, **348**, 618-629.
30. Parent, A.A., Gunther, J.R. and Katzenellenbogen, J.A. (2008) Blocking estrogen signaling after the hormone: pyrimidine-core inhibitors of estrogen receptor-coactivator binding. *J Med Chem*, **51**, 6512-6530.
31. Gunther, J.R., Moore, T.W., Collins, M.L. and Katzenellenbogen, J.A. (2008) Amphipathic benzenes are designed inhibitors of the estrogen receptor alpha/steroid receptor coactivator interaction. *ACS Chem Biol*, **3**, 282-286.
32. LaFrate, A.L., Gunther, J.R., Carlson, K.E. and Katzenellenbogen, J.A. (2008) Synthesis and biological evaluation of guanylhydrazone coactivator binding inhibitors for the estrogen receptor. *Bioorg Med Chem*, **16**, 10075-10084.
33. Arnold, L.A., Estebanez-Perpina, E., Togashi, M., Jouravel, N., Shelat, A., McReynolds, A.C., Mar, E., Nguyen, P., Baxter, J.D., Fletterick, R.J. *et al.* (2005) Discovery of small molecule inhibitors of the interaction of the thyroid hormone receptor with transcriptional coregulators. *J Biol Chem*, **280**, 43048-43055.
34. Hwang, J.Y., Arnold, L.A., Zhu, F., Kosinski, A., Mangano, T.J., Setola, V., Roth, B.L. and Guy, R.K. (2009) Improvement of pharmacological properties of irreversible thyroid receptor coactivator binding inhibitors. *J Med Chem*, **52**, 3892-3901.

35. Yuan, X., Ta, T.C., Lin, M., Evans, J.R., Dong, Y., Bolotin, E., Sherman, M.A., Forman, B.M. and Sladek, F.M. (2009) Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS One*, **4**, e5609.
36. Le Guevel, R., Oger, F., Lecorgne, A., Dudasova, Z., Chevance, S., Bondon, A., Barath, P., Simonneaux, G. and Salbert, G. (2009) Identification of small molecule regulators of the nuclear receptor HNF4alpha based on naphthofuran scaffolds. *Bioorg Med Chem*, **17**, 7021-7030.
37. Hardcastle, I.R., Ahmed, S.U., Atkins, H., Farnie, G., Golding, B.T., Griffin, R.J., Guyenne, S., Hutton, C., Kallblad, P., Kemp, S.J. *et al.* (2006) Small-molecule inhibitors of the MDM2-p53 protein-protein interaction based on an isoindolinone scaffold. *J Med Chem*, **49**, 6209-6221.
38. Shangary, S. and Wang, S. (2009) Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. *Annu Rev Pharmacol Toxicol*, **49**, 223-241.
39. Schust, J., Sperl, B., Hollis, A., Mayer, T.U. and Berg, T. (2006) Stattic: a small-molecule inhibitor of STAT3 activation and dimerization. *Chem Biol*, **13**, 1235-1242.
40. Xu, Y., Shi, J., Yamamoto, N., Moss, J.A., Vogt, P.K. and Janda, K.D. (2006) A credit-card library approach for disrupting protein-protein interactions. *Bioorg Med Chem*, **14**, 2660-2673.
41. Yin, X., Giap, C., Lazo, J.S. and Prochownik, E.V. (2003) Low molecular weight inhibitors of Myc-Max interaction and function. *Oncogene*, **22**, 6151-6159.
42. Takeda, K. and Akira, S. (2000) STAT family of transcription factors in cytokine-mediated biological responses. *Cytokine Growth Factor Rev*, **11**, 199-207.
43. McBride, O.W., Merry, D. and Givol, D. (1986) The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). *Proc Natl Acad Sci U S A*, **83**, 130-134.
44. Kern, S.E., Kinzler, K.W., Bruskin, A., Jarosz, D., Friedman, P., Prives, C. and Vogelstein, B. (1991) Identification of p53 as a sequence-specific DNA-binding protein. *Science*, **252**, 1708-1711.
45. Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat Rev Drug Discov*, **1**, 727-730.
46. Nickols, N.G., Jacobs, C.S., Farkas, M.E. and Dervan, P.B. (2007) Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chem Biol*, **2**, 561-571.
47. Mao, C., Patterson, N., Cherian, M., Aninye, I., Zhang, C., Montoya, J., Cheng, J., Putt, K., Hergenrother, P., Wilson, E. *et al.* (2008) A new small molecule inhibitor of estrogen receptor alpha binding to estrogen response elements blocks estrogen-dependent growth of cancer cells. *J Biol Chem*, **283**, 12819-12830.
48. Rishi, V., Potter, T., Laudeman, J., Reinhart, R., Silvers, T., Selby, M., Stevenson, T., Krosky, P., Stephen, A., Acharya, A. *et al.* (2005) A high-throughput fluorescence-anisotropy screen that identifies small molecule inhibitors of the DNA binding of B-ZIP transcription factors. *Anal Biochem*, **340**, 259-271.
49. Chan, L., Pineda, M., Heeres, J., Hergenrother, P. and Cunningham, B. (2008) A general method for discovering inhibitors of protein-DNA interactions using photonic crystal biosensors. *ACS Chem Biol*, **3**, 437-448.

50. Ng, P., Tang, Y., Knosp, W., Stadler, H. and Shaw, J. (2007) Synthesis of diverse lactam carboxamides leading to the discovery of a new transcription-factor inhibitor. *Angew Chem Int Ed Engl*, **46**, 5352-5355.
51. Rungeler, P., Castro, V., Mora, G., Goren, N., Vichnewski, W., Pahl, H.L., Merfort, I. and Schmidt, T.J. (1999) Inhibition of transcription factor NF-kappaB by sesquiterpene lactones: a proposed molecular mechanism of action. *Bioorg Med Chem*, **7**, 2343-2352.
52. Anzellotti, A., Liu, Q., Bloemink, M., Scarsdale, J. and Farrell, N. (2006) Targeting retroviral Zn finger-DNA interactions: a small-molecule approach using the electrophilic nature of trans-platinum-nucleobase compounds. *Chem Biol*, **13**, 539-548.
53. Wang, L., Yang, X., Zhang, X., Mihalic, K., Fan, Y., Xiao, W., Howard, O., Appella, E., Maynard, A. and Farrar, W. (2004) Suppression of breast cancer by chemical modulation of vulnerable zinc fingers in estrogen receptor. *Nat Med*, **10**, 40-47.
54. Ng, P.Y., Tang, Y., Knosp, W.M., Stadler, H.S. and Shaw, J.T. (2007) Synthesis of diverse lactam carboxamides leading to the discovery of a new transcription-factor inhibitor. *Angew Chem Int Ed Engl*, **46**, 5352-5355.
55. Hellsten, R., Johansson, M., Dahlman, A., Dizeyi, N., Sterner, O. and Bjartell, A. (2008) Galiellalactone is a novel therapeutic candidate against hormone-refractory prostate cancer expressing activated Stat3. *Prostate*, **68**, 269-280.
56. Turkson, J., Zhang, S., Mora, L.B., Burns, A., Sebti, S. and Jove, R. (2005) A novel platinum compound inhibits constitutive Stat3 signaling and induces cell cycle arrest and apoptosis of malignant cells. *J Biol Chem*, **280**, 32979-32988.
57. Mackenzie, G.G., Delfino, J.M., Keen, C.L., Fraga, C.G. and Oteiza, P.I. (2009) Dimeric procyanidins are inhibitors of NF-kappaB-DNA binding. *Biochem Pharmacol*, **78**, 1252-1262.
58. Kong, D., Park, E.J., Stephen, A.G., Calvani, M., Cardellina, J.H., Monks, A., Fisher, R.J., Shoemaker, R.H. and Melillo, G. (2005) Echinomycin, a small-molecule inhibitor of hypoxia-inducible factor-1 DNA-binding activity. *Cancer Res*, **65**, 9047-9055.
59. Brennan, P., Donev, R. and Hewamana, S. (2008) Targeting transcription factors for therapeutic benefit. *Mol Biosyst*, **4**, 909-919.
60. Bielinska, A., Shivdasani, R.A., Zhang, L.Q. and Nabel, G.J. (1990) Regulation of gene expression with double-stranded phosphorothioate oligonucleotides. *Science*, **250**, 997-1000.
61. Taniguchi, H., Fujiwara, Y., Doki, Y., Sugita, Y., Sohma, I., Miyata, H., Takiguchi, S., Yasuda, T., Tomita, N., Morishita, R. *et al.* (2007) Gene therapy using ets-1 transcription factor decoy for peritoneal dissemination of gastric cancer. *International journal of cancer. Journal international du cancer*, **121**, 1609-1617.
62. Hoel, A.W. and Conte, M.S. (2007) Edifoligide: a transcription factor decoy to modulate smooth muscle cell proliferation in vein bypass. *Cardiovascular drug reviews*, **25**, 221-234.
63. Leong, P.L., Andrews, G.A., Johnson, D.E., Dyer, K.F., Xi, S., Mai, J.C., Robbins, P.D., Gadiparthi, S., Burke, N.A., Watkins, S.F. *et al.* (2003) Targeted inhibition of Stat3 with a decoy oligonucleotide abrogates head and neck cancer cell growth. *Proc Natl Acad Sci U S A*, **100**, 4138-4143.
64. El-Andaloussi, S., Johansson, H., Magnusdottir, A., Jarver, P., Lundberg, P. and Langel, U. (2005) TP10, a delivery vector for decoy oligonucleotides targeting the Myc protein.

- Journal of controlled release : official journal of the Controlled Release Society*, **110**, 189-201.
65. Thiel, K.W. and Giangrande, P.H. (2009) Therapeutic applications of DNA and RNA aptamers. *Oligonucleotides*, **19**, 209-222.
 66. van Houte, L.P., Chuprina, V.P., van der Wetering, M., Boelens, R., Kaptein, R. and Clevers, H. (1995) Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *J Biol Chem*, **270**, 30516-30524.
 67. Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947-956.
 68. Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705-714.
 69. Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791-795.
 70. Remenyi, A., Lins, K., Nissen, L.J., Reinbold, R., Scholer, H.R. and Wilmanns, M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*, **17**, 2048-2059.
 71. Palasingam, P., Jauch, R., Ng, C.K. and Kolatkar, P.R. (2009) The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J Mol Biol*, **388**, 619-630.
 72. Murphy, E.C., Zhurkin, V.B., Louis, J.M., Cornilescu, G. and Clore, G.M. (2001) Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J Mol Biol*, **312**, 481-499.
 73. Scaffidi, P. and Bianchi, M.E. (2001) Spatially precise DNA bending is an essential activity of the sox2 transcription factor. *J Biol Chem*, **276**, 47296-47302.
 74. Weiss, M.A. (2001) Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition. *Mol Endocrinol*, **15**, 353-362.
 75. Kuhlbrodt, K., Herbarth, B., Sock, E., Enderich, J., Hermans-Borgmeyer, I. and Wegner, M. (1998) Cooperative function of POU proteins and SOX proteins in glial cells. *J Biol Chem*, **273**, 16050-16057.
 76. Tsukamoto, T., Inada, K., Tanaka, H., Mizoshita, T., Mihara, M., Ushijima, T., Yamamura, Y., Nakamura, S. and Tatematsu, M. (2004) Down-regulation of a gastric transcription factor, Sox2, and ectopic expression of intestinal homeobox genes, Cdx1 and Cdx2: inverse correlation during progression from gastric/intestinal-mixed to complete intestinal metaplasia. *J Cancer Res Clin Oncol*, **130**, 135-145.
 77. Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. and Kondoh, H. (2001) Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*, **15**, 1272-1286.
 78. Kondoh, H. and Kamachi, Y. (2010) SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int J Biochem Cell Biol*, **42**, 391-399.
 79. Wilson, M. and Koopman, P. (2002) Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr Opin Genet Dev*, **12**, 441-446.

80. Stefanovic, S., Abboud, N., Desilets, S., Nury, D., Cowan, C. and Puceat, M. (2009) Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *J Cell Biol*, **186**, 665-673.
81. Maruyama, M., Ichisaka, T., Nakagawa, M. and Yamanaka, S. (2005) Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *J Biol Chem*, **280**, 24371-24379.
82. Stros, M., Launholt, D. and Grasser, K.D. (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci*, **64**, 2590-2606.
83. Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev*, **17**, 126-140.
84. Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663-676.
85. Fong, H., Hohenstein, K.A. and Donovan, P.J. (2008) Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. *Stem Cells*, **26**, 1931-1938.
86. Graham, V., Khudyakov, J., Ellis, P. and Pevny, L. (2003) SOX2 functions to maintain neural progenitor identity. *Neuron*, **39**, 749-765.
87. Yuan, H., Corbi, N., Basilico, C. and Dailey, L. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev*, **9**, 2635-2645.
88. Nishimoto, M., Fukushima, A., Okuda, A. and Muramatsu, M. (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol*, **19**, 5453-5465.
89. Jauch, R., Aksoy, I., Hutchins, A.P., Ng, C.K., Tian, X.F., Chen, J., Palasingam, P., Robson, P., Stanton, L.W. and Kolatkar, P.R. (2011) Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells*, **29**, 940-951.
90. Koopman, P. (2005) Sex determination: a tale of two Sox genes. *Trends Genet*, **21**, 367-370.
91. Berta, P., Hawkins, J.R., Sinclair, A.H., Taylor, A., Griffiths, B.L., Goodfellow, P.N. and Fellous, M. (1990) Genetic evidence equating SRY and the testis-determining factor. *Nature*, **348**, 448-450.
92. Koopman, P. (1999) Sry and Sox9: mammalian testis-determining genes. *Cell Mol Life Sci*, **55**, 839-856.
93. Pevny, L.H. and Lovell-Badge, R. (1997) Sox genes find their feet. *Curr Opin Genet Dev*, **7**, 338-344.
94. Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R. and Kondoh, H. (1995) Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO J*, **14**, 3510-3519.
95. Kamachi, Y., Uchikawa, M., Collignon, J., Lovell-Badge, R. and Kondoh, H. (1998) Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. *Development*, **125**, 2521-2532.
96. Fantes, J., Ragge, N.K., Lynch, S.A., McGill, N.I., Collin, J.R., Howard-Peebles, P.N., Hayward, C., Vivian, A.J., Williamson, K., van Heyningen, V. *et al.* (2003) Mutations in SOX2 cause anophthalmia. *Nat Genet*, **33**, 461-463.

97. Ferri, A.L., Cavallaro, M., Braidà, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S. *et al.* (2004) Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development*, **131**, 3805-3819.
98. Wilmore, H.P., Smith, M.J., Wilcox, S.A., Bell, K.M. and Sinclair, A.H. (2000) SOX14 is a candidate gene for limb defects associated with BPES and Mobius syndrome. *Hum Genet*, **106**, 269-276.
99. Kiso, M., Tanaka, S., Saba, R., Matsuda, S., Shimizu, A., Ohyama, M., Okano, H.J., Shiroishi, T., Okano, H. and Saga, Y. (2009) The disruption of Sox21-mediated hair shaft cuticle differentiation causes cyclic alopecia in mice. *Proc Natl Acad Sci U S A*, **106**, 9292-9297.
100. Futaki, S., Hayashi, Y., Emoto, T., Weber, C.N. and Sekiguchi, K. (2004) Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression. *Mol Cell Biol*, **24**, 10492-10503.
101. Francois, M., Caprini, A., Hosking, B., Orsenigo, F., Wilhelm, D., Browne, C., Paavonen, K., Karnezis, T., Shayan, R., Downes, M. *et al.* (2008) Sox18 induces development of the lymphatic vasculature in mice. *Nature*, **456**, 643-647.
102. Irrthum, A., Devriendt, K., Chitayat, D., Matthijs, G., Glade, C., Steijlen, P.M., Fryns, J.P., Van Steensel, M.A. and Vikkula, M. (2003) Mutations in the transcription factor gene SOX18 underlie recessive and dominant forms of hypotrichosis-lymphedema-telangiectasia. *Am J Hum Genet*, **72**, 1470-1478.
103. Lee, H.J., Goring, W., Ochs, M., Muhlfeld, C., Steding, G., Paprotta, I., Engel, W. and Adham, I.M. (2004) Sox15 is required for skeletal muscle regeneration. *Mol Cell Biol*, **24**, 8428-8436.
104. Ikeda, T., Kamekura, S., Mabuchi, A., Kou, I., Seki, S., Takato, T., Nakamura, K., Kawaguchi, H., Ikegawa, S. and Chung, U.I. (2004) The combination of SOX5, SOX6, and SOX9 (the SOX trio) provides signals sufficient for induction of permanent cartilage. *Arthritis Rheum*, **50**, 3561-3573.
105. Laumonier, F., Ronce, N., Hamel, B.C., Thomas, P., Lespinasse, J., Raynaud, M., Paringaux, C., Van Bokhoven, H., Kalscheuer, V., Fryns, J.P. *et al.* (2002) Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am J Hum Genet*, **71**, 1450-1455.
106. Tanaka, S., Kamachi, Y., Tanouchi, A., Hamada, H., Jing, N. and Kondoh, H. (2004) Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. *Mol Cell Biol*, **24**, 8834-8846.
107. Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R. and Episkopou, V. (1998) Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes Dev*, **12**, 776-781.
108. Ekonomou, A., Kazanis, I., Malas, S., Wood, H., Alifragis, P., Denaxa, M., Karagozeos, D., Constanti, A., Lovell-Badge, R. and Episkopou, V. (2005) Neuronal migration and ventral subtype identity in the telencephalon depend on SOX1. *PLoS Biol*, **3**, e186.
109. Uchikawa, M., Kamachi, Y. and Kondoh, H. (1999) Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech Dev*, **84**, 103-120.
110. Sandberg, M., Kallstrom, M. and Muhr, J. (2005) Sox21 promotes the progression of vertebrate neurogenesis. *Nat Neurosci*, **8**, 995-1001.

111. Kamachi, Y., Uchikawa, M. and Kondoh, H. (2000) Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet*, **16**, 182-187.
112. Bergsland, M., Werme, M., Malewicz, M., Perlmann, T. and Muhr, J. (2006) The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev*, **20**, 3475-3486.
113. Stolt, C.C., Schlierf, A., Lommes, P., Hillgartner, S., Werner, T., Kosian, T., Sock, E., Kessar, N., Richardson, W.D., Lefebvre, V. *et al.* (2006) SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. *Dev Cell*, **11**, 697-709.
114. Lefebvre, V., Li, P. and de Crombrughe, B. (1998) A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J*, **17**, 5718-5733.
115. Stolt, C.C., Lommes, P., Hillgartner, S. and Wegner, M. (2008) The transcription factor Sox5 modulates Sox10 function during melanocyte development. *Nucleic Acids Res*, **36**, 5427-5440.
116. Stolt, C.C., Lommes, P., Friedrich, R.P. and Wegner, M. (2004) Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. *Development*, **131**, 2349-2358.
117. Pingault, V., Bondurand, N., Kuhlbrodt, K., Goerich, D.E., Prehu, M.O., Puliti, A., Herbarth, B., Hermans-Borgmeyer, I., Legius, E., Matthijs, G. *et al.* (1998) SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet*, **18**, 171-173.
118. Kim, J., Lo, L., Dormand, E. and Anderson, D.J. (2003) SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. *Neuron*, **38**, 17-31.
119. Dong, C., Wilhelm, D. and Koopman, P. (2004) Sox genes and cancer. *Cytogenet Genome Res*, **105**, 442-447.
120. Rodriguez-Pinilla, S.M., Sarrío, D., Moreno-Bueno, G., Rodriguez-Gil, Y., Martínez, M.A., Hernandez, L., Hardisson, D., Reis-Filho, J.S. and Palacios, J. (2007) Sox2: a possible driver of the basal-like phenotype in sporadic breast cancer. *Mod Pathol*, **20**, 474-481.
121. Chen, Y., Shi, L., Zhang, L., Li, R., Liang, J., Yu, W., Sun, L., Yang, X., Wang, Y., Zhang, Y. *et al.* (2008) The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J Biol Chem*, **283**, 17969-17978.
122. Spisek, R., Kukreja, A., Chen, L.C., Matthews, P., Mazumder, A., Vesole, D., Jagannath, S., Zebroski, H.A., Simpson, A.J., Ritter, G. *et al.* (2007) Frequent and specific immunity to the embryonal stem cell-associated antigen SOX2 in patients with monoclonal gammopathy. *J Exp Med*, **204**, 831-840.
123. Schmitz, M., Temme, A., Senner, V., Ebner, R., Schwind, S., Stevanovic, S., Wehner, R., Schackert, G., Schackert, H.K., Fussel, M. *et al.* (2007) Identification of SOX2 as a novel glioma-associated antigen and potential target for T cell-based immunotherapy. *Br J Cancer*, **96**, 1293-1301.
124. Bass, A.J., Watanabe, H., Mermel, C.H., Yu, S., Perner, S., Verhaak, R.G., Kim, S.Y., Wardwell, L., Tamayo, P., Gat-Viks, I. *et al.* (2009) SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet*, **41**, 1238-1242.

125. Hide, T., Takezaki, T., Nakatani, Y., Nakamura, H., Kuratsu, J. and Kondo, T. (2009) Sox11 prevents tumorigenesis of glioma-initiating cells by inducing neuronal differentiation. *Cancer Res*, **69**, 7953-7959.
126. Frierson, H.F., Jr., El-Naggar, A.K., Welsh, J.B., Sapinoso, L.M., Su, A.I., Cheng, J., Saku, T., Moskaluk, C.A. and Hampton, G.M. (2002) Large scale molecular analysis identifies genes with altered expression in salivary adenoid cystic carcinoma. *Am J Pathol*, **161**, 1315-1323.
127. Reya, T., Morrison, S.J., Clarke, M.F. and Weissman, I.L. (2001) Stem cells, cancer, and cancer stem cells. *Nature*, **414**, 105-111.
128. Nicolis, S.K. (2007) Cancer stem cells and "stemness" genes in neuro-oncology. *Neurobiol Dis*, **25**, 217-229.
129. Chen, Y., Shi, L., Zhang, L., Li, R., Liang, J., Yu, W., Sun, L., Yang, X., Wang, Y., Zhang, Y. *et al.* (2008) The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J Biol Chem*, **283**, 17969-17978.
130. Chen, S., Hilcove, S. and Ding, S. (2006) Exploring stem cell biology with small molecules. *Mol Biosyst*, **2**, 18-24.
131. Chen, S., Do, J.T., Zhang, Q., Yao, S., Yan, F., Peters, E.C., Scholer, H.R., Schultz, P.G. and Ding, S. (2006) Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A*, **103**, 17266-17271.
132. Borowiak, M., Maehr, R., Chen, S., Chen, A.E., Tang, W., Fox, J.L., Schreiber, S.L. and Melton, D.A. (2009) Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell*, **4**, 348-358.
133. Martis, E., Radhakrishnan, R. and Badve, R. (2011) High-Throughput Screening: The Hits and Leads of Drug Discovery- An Overview. *JAPS*, **01**, 02-10.
134. Monge, A., Arrault, A., Marot, C. and Morin-Allory, L. (2006) Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. *Molecular diversity*, **10**, 389-403.
135. (2007) The academic pursuit of screening. *Nature chemical biology*, **3**, 433.
136. Kaiser, J. (2008) Molecular biology. Industrial-style screening meets academic biology. *Science*, **321**, 764-766.
137. Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J. and Bryant, S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*, **37**, W623-633.
138. Rhule, J.T., Hill, C.L., Judd, D.A. and Schinazi, R.F. (1998) Polyoxometalates in Medicine. *Chem Rev*, **98**, 327-358.
139. Hasenknopf, B. (2005) Polyoxometalates: introduction to a class of inorganic compounds and their biomedical applications. *Front Biosci*, **10**, 275-287.
140. Briand, L.E., Valle, G.M. and Thomas, H.J. (2002) Stability of the phospho-molybdic Dawson-type ion P2Mo18O626- in aqueous media. *Journal of Materials Chemistry*, **12**, 299-304.
141. Prudent, R., Moucadel, V., Laudet, B., Barette, C., Lafanechère, L., Hasenknopf, B., Li, J., Bareyt, S., Lacôte, E., Thorimbert, S. *et al.* (2008) Identification of polyoxometalates as nanomolar noncompetitive inhibitors of protein kinase CK2. *Chem Biol*, **15**, 683-692.
142. Hill, C.L., Weeks, M.S. and Schinazi, R.F. (1990) Anti-HIV-1 activity, toxicity, and stability studies of representative structural families of polyoxometalates. *J Med Chem*, **33**, 2767-2772.

143. Bareyt, S., Piligkos, S., Hasenknopf, B., Gouzerh, P., Lacote, E., Thorimbert, S. and Malacria, M. (2003) Highly efficient peptide bond formation to functionalized Wells-Dawson-type polyoxotungstates. *Angew Chem Int Ed Engl*, **42**, 3404-3406.
144. Bareyt, S., Piligkos, S., Hasenknopf, B., Gouzerh, P., Lacote, E., Thorimbert, S. and Malacria, M. (2005) Efficient preparation of functionalized hybrid organic/inorganic Wells-Dawson-type polyoxotungstates. *J Am Chem Soc*, **127**, 6788-6794.
145. Micoine, K., Hasenknopf, B., Thorimbert, S., Lacote, E. and Malacria, M. (2007) A general strategy for ligation of organic and biological molecules to Dawson and Keggin polyoxotungstates. *Organic letters*, **9**, 3981-3984.
146. Raynaud, M., Chermann, J.C., Plata, F., Jasmin, C., Mathe, G. and Sinoussi, F. (1972) Properties of new inhibitors of murine leukaemia and sarcoma viruses. *Prog Immunobiol Stand*, **5**, 285-288.
147. Raynaud, M., Chermann, J.C., Plata, F., Jasmin, C. and Mathe, G. (1971) [Inhibitors of the murine leukemia sarcoma group viruses. Silicotungstates]. *C R Acad Sci Hebd Seances Acad Sci D*, **272**, 347-348.
148. Fischer, J., Ricard, L. and Weiss, R. (1976) The structure of the heteropolytungstate (NH₄)₁₇Na NaW₂₁Sb₉O₈₆-14H₂O. An inorganic cryptate. *J Am Chem Soc*, **98**, 3050-3052.
149. Moskovitz, B.L. (1988) Clinical trial of tolerance of HPA-23 in patients with acquired immune deficiency syndrome. *Antimicrob Agents Chemother*, **32**, 1300-1303.
150. Mukherjee, H.N. (1965) Treatment of Cancer of the Intestinal Tract with a Complex Compound of Phosphotungstic Phosphomolybdic Acids and Caffeine. *J Indian Med Assoc*, **44**, 477-479.
151. Wang, X., Liu, J., Li, J., Yang, Y., Li, B. and Pope, M.T. (2003) Synthesis and antitumor activity of cyclopentadienyltitanium substituted polyoxotungstate [CoW₁₁O₃₉(CpTi)]⁷⁻ (Cp=η⁵-C₅H₅). *J Inorg Biochem*, **94**, 279-284.
152. Wang, X., Liu, J. and Pope, M.T. (2003) New polyoxometalate/starch nanomaterial: synthesis, characterization and antitumoral activity. *Dalton Transactions*, 957-960.
153. Sarafianos, S., Kortz, U., Pope, M. and Modak, M. (1996) Mechanism of polyoxometalate-mediated inactivation of DNA polymerases: an analysis with HIV-1 reverse transcriptase indicates specificity for the DNA-binding cleft. *Biochem J*, **319** (Pt 2), 619-626.
154. Judd, D., Nettles, J., Nevins, N., Snyder, J., Liotta, D., Tang, J., Ermolieff, J., Schinazi, R. and Hill, C. (2001) Polyoxometalate HIV-1 protease inhibitors. A new mode of protease inhibition. *J Am Chem Soc*, **123**, 886-897.
155. Li, Y., He, Y. and Luo, Y. (2009) Crystal structure of an archaeal Rad51 homologue in complex with a metatungstate inhibitor. *Biochemistry*, **48**, 6805-6810.
156. Laramas, M., Pasquier, D., Filhol, O., Ringeisen, F., Descotes, J.L. and Cochet, C. (2007) Nuclear localization of protein kinase CK2 catalytic subunit (CK2α) is associated with poor prognostic factors in human prostate cancer. *Eur J Cancer*, **43**, 928-934.
157. Kim, J.S., Eom, J.I., Cheong, J.W., Choi, A.J., Lee, J.K., Yang, W.I. and Min, Y.H. (2007) Protein kinase CK2α as an unfavorable prognostic marker and novel therapeutic target in acute myeloid leukemia. *Clin Cancer Res*, **13**, 1019-1028.
158. Vale, R.D., Reese, T.S. and Sheetz, M.P. (1985) Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell*, **42**, 39-50.

159. Learman, S., Kim, C., Stevens, N., Kim, S., Wojcik, E. and Walker, R. (2009) NSC 622124 Inhibits Human Eg5 and Other Kinesins via Interaction with the Conserved Microtubule-Binding Site (dagger). *Biochemistry*, **48**, 1754-1762.
160. Flutsch, A., Schroeder, T., Grutter, M.G. and Patzke, G.R. (2011) HIV-1 protease inhibition potential of functionalized polyoxometalates. *Bioorg Med Chem Lett*, **21**, 1162-1166.
161. Shigeta, S., Mori, S., Yamase, T. and Yamamoto, N. (2006) Anti-RNA virus activity of polyoxometalates. *Biomed Pharmacother*, **60**, 211-219.
162. WANG JP, H.D.S.Z. (2010) Molecular simulation study of the binding mechanism of [α -PTi₂W₁₀O₄₀]⁷⁻ for its promising broad-spectrum inhibitory activity to FluV-A neuraminidase. *Chinese Science Bulletin*, **55(23)**, 2497-2504.
163. Dong, Z., Tan, R., Cao, J., Yang, Y., Kong, C., Du, J., Zhu, S., Zhang, Y., Lu, J., Huang, B. *et al.* (2011) Discovery of polyoxometalate-based HDAC inhibitors with profound anticancer activity in vitro and in vivo. *Eur J Med Chem*, **46**, 2477-2484.
164. Cibert, C. and Jasmin, C. (1982) Determination of the intracellular localization of a polyoxotungstate (HPA-23) by Raman Laser and X fluorescence spectroscopies. *Biochem Biophys Res Commun*, **108**, 1424-1433.
165. Wang, X., Li, F., Liu, S. and Pope, M.T. (2005) New liposome-encapsulated-polyoxometalates: synthesis and antitumoral activity. *J Inorg Biochem*, **99**, 452-457.
166. Ng, C.K., Palasingam, P., Venkatachalam, R., Baburajendran, N., Cheng, J., Jauch, R. and Kolatkar, P.R. (2008) Purification, crystallization and preliminary X-ray diffraction analysis of the HMG domain of Sox17 in complex with DNA. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **64**, 1184-1187.
167. Johnson, R., Samuel, J., Ng, C.K., Jauch, R., Stanton, L.W. and Wood, I.C. (2009) Evolution of the vertebrate gene regulatory network controlled by the transcriptional repressor REST. *Mol Biol Evol*, **26**, 1491-1507.
168. Ng, C.K.L., Noel, X., Li, Chee, S., Prabhakar, S., Kolatkar, P.R. and Jauch, R. (2012) Deciphering the Sox-Oct Partner Code by Quantitative Cooperativity Measurements. *Nucleic Acids Res*.
169. Lundblad, J.R., Laurance, M. and Goodman, R.H. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol Endocrinol*, **10**, 607-612.
170. Zhang, J.H., Chung, T.D. and Oldenburg, K.R. (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen*, **4**, 67-73.
171. Seiler, K.P., George, G.A., Happ, M.P., Bodycombe, N.E., Carrinski, H.A., Norton, S., Brudz, S., Sullivan, J.P., Muhlich, J., Serrano, M. *et al.* (2008) ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res*, **36**, D351-359.
172. Cummings, M.D., Farnum, M.A. and Nelen, M.I. (2006) Universal screening methods and applications of ThermoFluor. *J Biomol Screen*, **11**, 854-863.
173. keller, R. (2004), Swiss Federal Institute of Technology (ETH), Zurich, Zurich.
174. Sha, J., Peng, J., Lan, Y., Su, Z., Pang, H., Tian, A., Zhang, P. and Zhu, M. (2008) pH-dependent assembly of hybrids based on Wells-Dawson POM/Ag chemistry. *Inorg Chem*, **47**, 5145-5153.

175. Stewart, J.J. (2007) Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model*, **13**, 1173-1213.
176. Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S. and Onufriev, A. (2005) H⁺⁺: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Res*, **33**, W368-371.
177. Bikadi, Z. and Hazai, E. (2009) Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J Cheminform*, **1**, 15.
178. Morris, G.M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J computational chemistry*, **19**, 1639-1662.
179. Solis, F.J.W., R.J-B. (1981) Minimization by random search techniques. *Mathematics of Operation Research*, **6**, 19-30.
180. Fraczekiewicz, R.a.B., W. (1998) Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *journal of computational chemistry*, **19**, 319-333.
181. Prudent, R., Moucadet, V., Laudet, B., Barette, C., Lafanechere, L., Hasenknopf, B., Li, J., Bareyt, S., Lacote, E., Thorimbert, S. *et al.* (2008) Identification of polyoxometalates as nanomolar noncompetitive inhibitors of protein kinase CK2. *Chem Biol*, **15**, 683-692.
182. John, D.M. and Weeks, K.M. (2000) van't Hoff enthalpies without baselines. *Protein Sci*, **9**, 1416-1419.
183. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-137.
184. Schumann, F.H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.P. and Kalbitzer, H.R. (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *Journal of biomolecular NMR*, **39**, 275-289.
185. Clarkson, J. and Campbell, I.D. (2003) Studies of protein-ligand interactions by NMR. *Biochemical Society transactions*, **31**, 1006-1009.
186. Gao, G., Williams, J.G. and Campbell, S.L. (2004) Protein-protein interaction analysis by nuclear magnetic resonance spectroscopy. *Methods Mol Biol*, **261**, 79-92.
187. Akabayov, S.R., Biron, Z., Lamken, P., Piehler, J. and Anglister, J. (2010) NMR mapping of the IFNAR1-EC binding site on IFNalpha2 reveals allosteric changes in the IFNAR2-EC binding site. *Biochemistry*, **49**, 687-695.
188. Cary, P.D., Read, C.M., Davis, B., Driscoll, P.C. and Crane-Robinson, C. (2001) Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5. *Protein Sci*, **10**, 83-98.
189. Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, **13**, 1263-1275.
190. Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K. and Cheung, E. (2011) AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *EMBO J*, **30**, 2569-2581.
191. Nakshatri, H. and Badve, S. (2007) FOXA1 as a therapeutic target for breast cancer. *Expert Opin Ther Targets*, **11**, 507-514.

192. Rigby, A.C. (2009) Exploring novel chemical space through the use of computational and structural biology. *Comb Chem High Throughput Screen*, **12**, 927-928.
193. Pendergrast, P.S., Marsh, H.N., Grate, D., Healy, J.M. and Stanton, M. (2005) Nucleic acid aptamers for target validation and therapeutic applications. *Journal of biomolecular techniques : JBT*, **16**, 224-234.

List of publications:

1. Narasimhan, K., Pillay, S., Bin Ahmad, N.R., Bikadi, Z., Hazai, E., Yan, L., Kolatkar, P.R., Pervushin, K. and Jauch, R. (2011) Identification of a polyoxometalate inhibitor of the DNA binding activity of Sox2. *ACS Chem Biol*, **6**, 573-581.
2. Baburajendran, N., Palasingam, P., Narasimhan, K., Sun, W., Prabhakar, S., Jauch, R. and Kolatkar, P.R. (2010) Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic Acids Res*, **38**, 3477-3488.
3. Baburajendran, N., Jauch, R., Tan, C.Y., Narasimhan, K. and Kolatkar, P.R. (2011) Structural basis for the cooperative DNA recognition by Smad4 MH1 dimers. *Nucleic Acids Res*, **39**, 8213-8222.
4. Jauch, R., Ng, C.K., Narasimhan, K. and Kolatkar, P.R. (2011) Crystal structure of the Sox4 HMG/DNA complex suggests a mechanism for the positional interdependence in DNA recognition. *Biochem J*.
5. Kamesh, N., Aradhyam, G.K. and Manoj, N. (2008) The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. *BMC Evol Biol*, **8**, 129

Conference paper:

1. Jauch, R., Hutchins, A., Baburajendran, N., Kamesh, N., Ng, C.K., and Kolatkar, P.R. Decoding the structural basis for the *cis*-regulatory logic of early developmental switches *Acta Cryst A*. 2009 A65, s21

Conference and Poster presentations:

1. Transcription factors involved in stem cell pluripotency: Key targets for small molecule modulation. *Narasimhan K, Melamed P, Kolatkar P.R., and Jauch R. 13th Biological Sciences Graduate Congress*, December 2009, Chulalongkorn University, Thailand. (Best oral presentation award).
2. Identification of a lead inhibitor of Sox2-DNA interaction using a high-throughput fluorescence anisotropy screen. *Narasimhan K, Kolatkar P.R., and Jauch R. A*STAR Scientific conference*, October 2009, Biopolis, Singapore.

APPENDIX A

> FoxA1FL

MEGHETSDWNSYYADTQEAYSSVPVSNMNSGLGSMNSMNTYMTMNTMTTSGNMT PASFNMSYAN
PGLGAGLSPGAVAGMPGGSAGAMNSMTAAGVTAMGTALSPSGMGAMGAQQAASMNGLGPYAAAM
NPCMSPMAYAPSNLGRSRAGGGGDAKTFKRSYPHAKPPYSYISLITMAIQQAPSKMLTLSEIYQ
WIMDLFPYYRQNRWQNSIRHSLSFNDCFVKVARSPDKPGKGSYWTLHPDSGNMFENGCYLRR
QKRFKCEKQPGAGGGGGSGSGGSGAKGGPE SRKDP SGASNPSADSPLHRGVHGKTGQLEGAPAP
GPAASPQTLDHSGATATGGASELKT PASSTAPP I SSGPGALASVPASHPAHGLAPHESQLHLKG
DPHYSFNHPFSINNLMSSSEQQHKLDFKAYEQALQYSPYGSTLPASLPLGSASVTTRSPIEPSA
LEPAYYQGVYSRPVLNTS

> Sox2HMG

GSFTMGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRP
FIDEAKRLRALHMKEHPDYKYRPRRKT TLMKKDKYTLP GGLLAPGGNSM

> Sox7

SRIRRP MNAFMVWAKDERKRLAVQNPDLHNAELSKMLGKSWKALTLSQKRPYVDEAERLRLQHM
QDYPNYKYRPRRKKQ

> Sox10

PHVKRPMNAFMVWAQAARRK LADQYPHLHNAELSKTLGKLRLLNESDKRPFIEEAERLRMQHK
KDHPDYKYQPRRRKN

> Sox5

PHIKRPMNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKAMTNLEKQPYYEEQARLSKQHL
EKYPDYKYKPRPKRT

> Sox17

SRIRRP MNAFMVWAKDERKRLAQQNPDLHNAELSKMLGKSWKALT LAEKRPFFVEEAERLRVQHM
QDHPNYKYRPRRKKQ

> Sox15

EKVKRPMNAFMVWSSVQRRQMAQQNP KMHNSEISKRLGAQWKLLGDEEKRPFFVEEAERLRARHL
RDYPDYKYRPRRKS

> Sox8

PHVKRPMNAFMVWAQAARRK LADQYPHLHNAELSKTLGKLRLLSESEKRPFFVEEAERLRVQHK
KDHPDYKYQPRRKS

> Sox11

GHIKRP MNAFMVWSKIERRKIMEQSPDMHNAEISKRLGKRWKMLKDSEKIPFIREAERLR LKHM
ADYPDYKYRPRKKPK

>Sox9

PHVKRPMNAFMVWAQAARRKLDQYPHLHNAELSKTLGKLRLLNESEKRPFVEEAERLRVQHK
KDHPDYKYQPRRRKS

>Sox4

GHIKRPMPNAFMVWSQIERRKIMEQSPDMHNAEISKRLGKRWKLKDSKIPFIQEAERLRLLKHM
ADYDPDYKYRPRKKVK

>Sox6

PHIKRPMPNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNQEKQPYEEQARLSKIHL
EKYPNYKYKPRPKRT

>Sox18

LRIRRPMPNAFMVWAKDERKRLAQONPDLHNAVLSKMLGKAWKELNTAEKRPFVEEAERLRVQHL
RDHPNYKYRPRRKKQ

>Ap2 MBP

MKIHSHHHHHEEGKLVIIWINGDKGYNGLAEVGGKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGP
DIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDL
LPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENKDYDIKDVGVND
AGAKAGLTFVLVDLIKKNHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPT
FKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYYEELAK
DPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSITSLYKKAG
MLWKITDNVKEYEEDCEDRHDGSSNGNPRVPHLSSAGQHLYSPAPPLSHTGVAEYQPPPYFPPPY
QQLAYSQSADPYSHLGEAYAAAINPLHQAPTGSQQQAWPGRQSQEGAGLPSHHGRPAGLLPHL
SGLEAGAVSARRDAYRRSDLLLPHAHALDAAGLAENLGLHDMPHQMDEVQNVDDQHLLLDQTV
IRKGPISMTKNPLNLPQCQKELVGAVMNPTEVFCSVPGRLSLLSSTSKYKVTVAEVQRRLSPPEC
LNASLLGGVLRRAKSKNGRSLREKLDKIGLNLPAARRKAAHVTLTSLVEGEAVHLARDFAYV
CEAEFPSKPVAEYLTRPHLGGRNEMAARKNMLLAAQQLCKEFTTELLSQDRTPHGTSRLAPVLET
NIQNCLSHFSLITHGFGSQAICAAVSALQNYIKEALIVIDKSYMNPQDQSPADSNKTLEKMEKH
RK

>Pax6

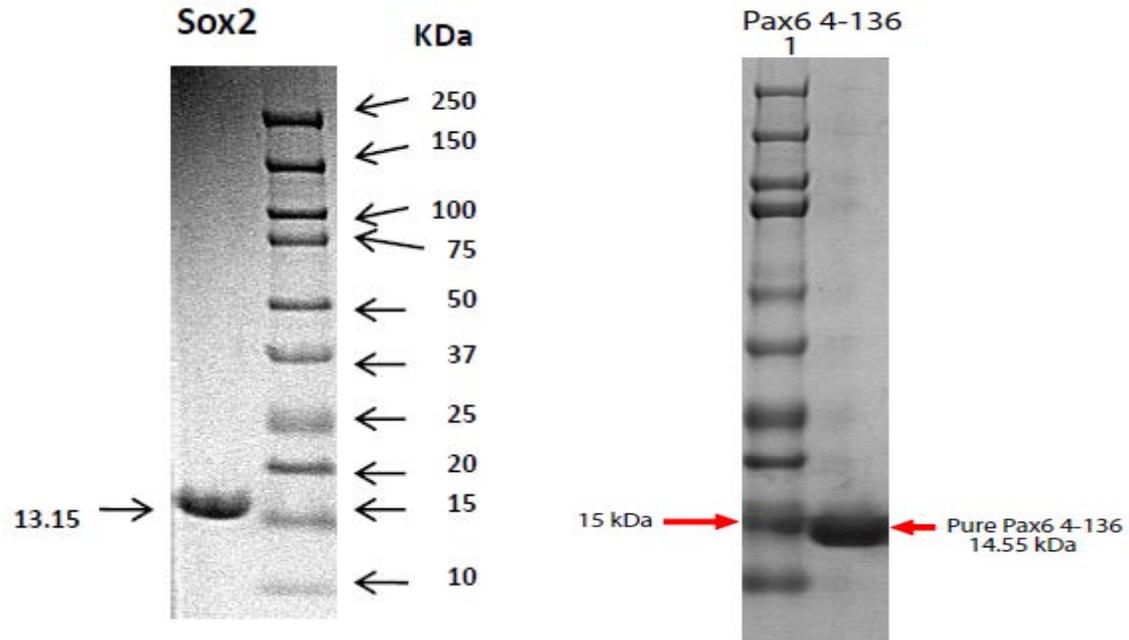
SHSGVNQLGGVFNVRPLPDSTRQKIVELAHSGARPCDISRILQVSNCGVSKILGRYYETGSIR
PRAIGGSKPRVATPEVVSKIAQYKRECPISFAWEIRDRLLESEGVCTNDNIPSVSSINRVLRNLA
SEKQQ

>RESTHIS₆MBP

GSSTAEEGDFSKGPIRCDRCGYNTNRYDHYTAHLKHHTRAGDNERVYKCIICTYTTVSEYHWRK
HLRNHFPRKVYTCGKCNYSRDRKNNYVQHVRTHGTGERPYKCELCOPYSSSQKTHLTRHMRTHSGE
KPFKCDQCSYVASNQHEVTRHARQVHNGPKPLNCPHCDYKTADRSNFKKHVELHVNPRQFNCPV
CDYAASKKCNLQYHFKSKHPTCPNKTMDVSKVKLKKTKKREADLPDN

The protein sequences used in TF binding experiments like fluorescence anisotropy and EMSA

APPENDIX B



Representative 12% SDS gel images of the purified Sox2-HMG and Pax6 protein used in the study

APPENDIX C

***CCND1* (22bp)**

5' (FAM) -CTGCCGGGCTTTGATCTTTGCT-3'

***Lama-1* (21bp)**

5' (Cy5) - ATCCAGGACAATAGAGACTGT-3'

***FGF-4* (18bp)**

5' (FAM) -AAAACCTCTTTGTTTGGAT-3'

***RE-1* (22bp)**

5' (FAM) -CTTCAGCACCTCGGACAGCTCC-3'

***FoxA1 element* (19bp)**

5' (FAM) -TGCCAAGTAAATAGTGCAG-3'

***AP2-HPSE* (19bp)**

5' (FAM) -AAAGTGCCCAGAGCCCATG-3'

The DNA duplexes used in TF binding experiments like fluorescence anisotropy and EMSA

Appendix D

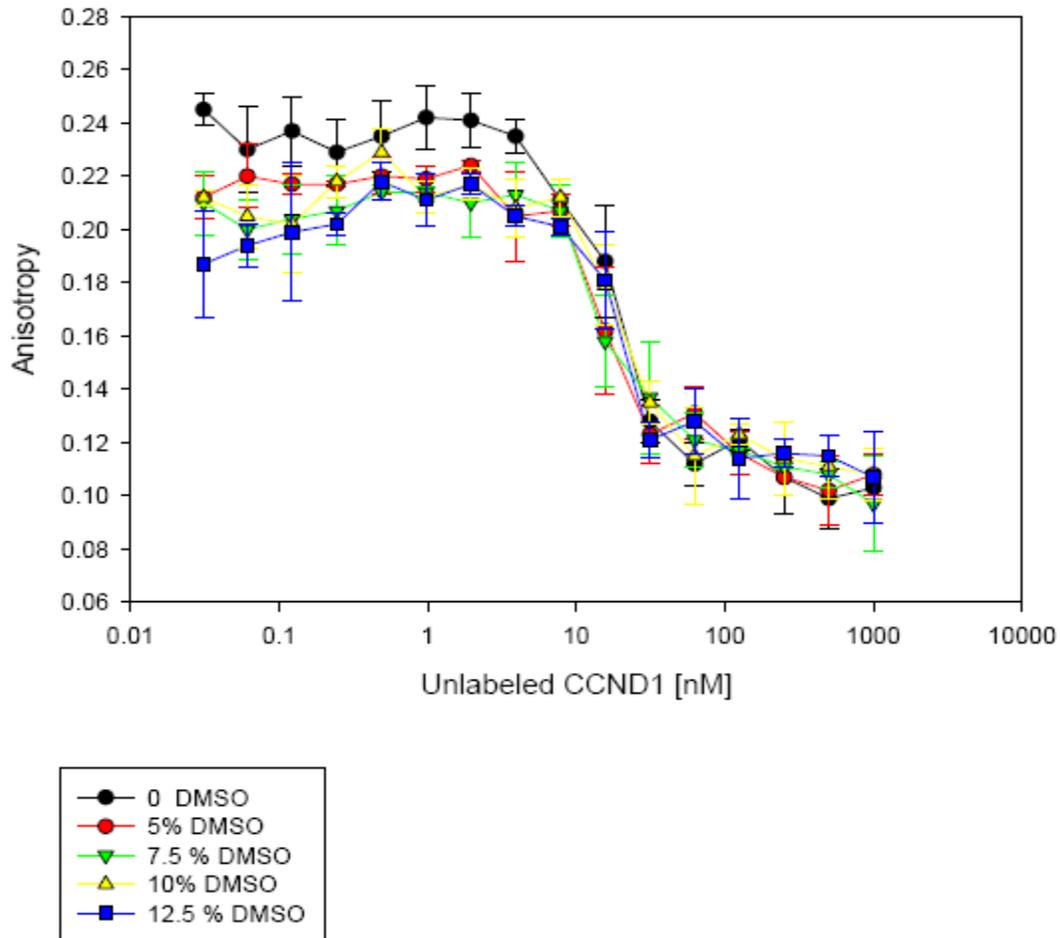


Figure. A saturated complex of 1nM DNA and 50 nM Sox2 was competed by addition of unlabeled CCND1 in the presence of varying concentrations of DMSO. The assay shows tolerance even at high DMSO concentrations (>10% DMSO).

APPENDIX E

Appendix table 1. Primary hits identified from the screening

NSC	Molecular formula	Mol.wt (g mol ⁻¹)	Composite Z-score	Reproducibility
3064	C ₁₆ H ₁₆ N ₄ O ₄	328	-7.619112165	-0.999797067
25678	C ₁₈ H ₁₅ NO ₂	277	-6.060548322	-0.999716549
45536	C ₁₆ H ₁₁ FN ₂ O	266	-5.779801696	-0.999029388
7572	C ₁₃ H ₁₄ N ₂ O ₃ S	278	-5.64302035	-0.998102966
13973	C ₁₅ H ₁₅ N ₃ .ClH	274	-5.638405917	-0.999735078
605756	C ₁₃ H ₁₁ N ₃ .8/5ClH	268	-5.624649541	-0.999886179
403374	C ₈ H ₉ ClN ₄ S	229	-5.606006746	-0.999815322
12644	C ₉ H ₁₂ AsNO ₅	289	-5.603219665	-0.999990452
117197	C ₁₄ H ₁₃ N ₅ O ₃	299	-5.600512132	-0.998941935
6731	C ₁₇ H ₁₈ O ₂	254	-5.591618633	-0.999922619
373535	C ₁₂ H ₁₂ N ₂ O ₃	232	-5.577978008	-0.999994223
337851	C ₄₀ H ₄₅ N ₃ O ₆ S	696	-5.400999762	-0.999360347
402083	C ₁₃ H ₁₀ N ₂ O ₂	226	-5.350655033	-0.999934197
660301	C ₁₀ H ₉ ClN ₂ O ₂	225	-5.238020222	-0.998603973
503425	C ₇ H ₄ ClNS ₂	202	-5.20617902	-0.999999266
107022	C ₁₂ H ₆ O ₈	278	-5.13495199	-0.998360358
667251	C ₁₇ H ₁₆ N ₂	248	-5.090616878	-0.998804377
637308	C ₁₂ H ₁₀ N ₂ O ₄	246	-4.380567029	-0.999965859
622124	K ₆ Mo ₁₈ O ₆₂ P ₂	3015	-4.06659908	-0.999486466
99027	C ₁₀ H ₇ BrO ₂	239	-4.065993889	-0.999940475
663996	C ₁₄ H ₁₈ N ₂ O ₂	246	-4.065768749	-0.998114385
82025	C ₁₄ H ₁₁ N ₃ O ₆	317	-4.032896914	-0.9978595
638352	C ₃₂ H ₂₂ N ₆ O ₁₄ S _{4.4} Na	935	-3.999039966	-0.999799055

375392	C ₁₄ H ₁₄ N ₂	210	-3.882518672	-0.999740793
34931	C ₃₄ H ₂₄ N ₄ O ₁₄ S ₄ .4Na	933	-3.86807345	-0.999933053
261726	C ₆ H ₆ N ₄ O	150	-3.863854659	-0.999291024
191389	C ₁₇ H ₁₇ NO	251	-3.795361323	-0.996068431
616232	C ₆ H ₁₂ Br ₂ O ₄	308	-3.75795373	-0.999874457
659437	C ₁₀ H ₇ N ₃ O ₄	233	-3.735222308	-0.99862276
636132	C ₁₈ H ₁₆ N ₂ O ₂	292	-3.733185925	-0.999997171
63878	C ₉ H ₁₃ N ₃ O ₅ .ClH	280	-3.730940183	-0.999679245
638080	C ₁₀ H ₈ N ₂ O ₄	220	-3.723164705	-0.997994358
688795	C ₁₂ H ₁₂ O ₄	220	-3.698398749	-0.99837388
635441	C ₁₁ H ₁₂ N ₂ O ₃	220	-3.664808403	-0.999853742
80396	C ₁₂ H ₇ F ₃ O ₂ S	272	-3.658615322	-0.99998664
7833	C ₁₄ H ₁₀ N ₂ O ₂	238	-3.605805426	-0.999583929
5200	C ₁₀ H ₁₀ N ₆ O	230	-3.571323176	-0.999620574
295156	C ₇ H ₈ O ₂	124	-3.565955577	-0.999960147
285166	C ₉ H ₇ NO	145	-3.533904375	-0.999128027
283162	C ₉ H ₁₈ N ₆ O ₃	258	-3.519693925	-0.999035146
606532	C ₇ H ₇ N ₃ O ₅	213	-3.477039911	-0.999241966
31702	C ₁₈ H ₁₃ N ₃ O	287	-3.345152209	-0.998497235
157004	C ₁₈ H ₁₆ N ₂ O ₅	340	-3.229268921	-0.999830257
28002	C ₁₈ H ₁₄ N.Cl	280	-3.228235201	-0.999982325
4170	C ₁₁ H ₈ O ₂	172	-3.223886845	-0.999923013
659501	C ₁₃ H ₁₀ N ₂ O ₂ S	258	-3.12676136	-0.999529284
637993	C ₂₂ H ₂₆ N ₄ O ₂ .2ClH	451	-3.062857238	-0.995853966
1771	C ₆ H ₁₂ N ₂ S ₄	240	-3.058732054	-0.9949141
672425	C ₁₉ H ₁₄ N ₂ O ₄ .HNO ₃	397	-3.046819292	-0.998450148
32946	C ₅ H ₁₂ N ₈ .2ClH	257	-3.014661188	-0.997342004
625748	C ₈ H ₁₂ N ₆	192	-3.009402527	-0.997944332

Appendix F

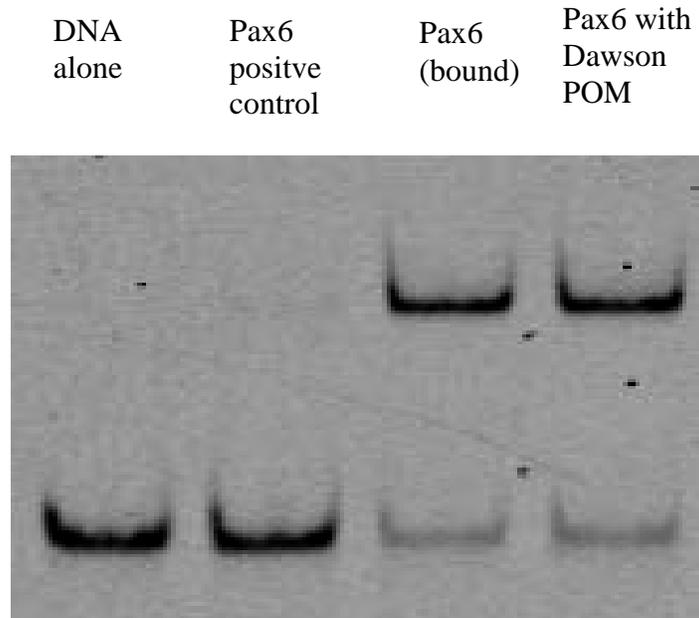


Figure. 8nM of Pax6 was added to 1nM of a fluorescently labeled Pax6 consensus DNA sequence. Addition of unlabeled Pax6 consensus DNA sequence (100nM DNA element) serves as the positive control for complete inhibition of the Pax6-fluorescein labeled DNA complex. Addition of 200nM of Dawson POM to the previously bound Pax6-DNA complex has no effect on Pax6-DNA binding

Appendix G

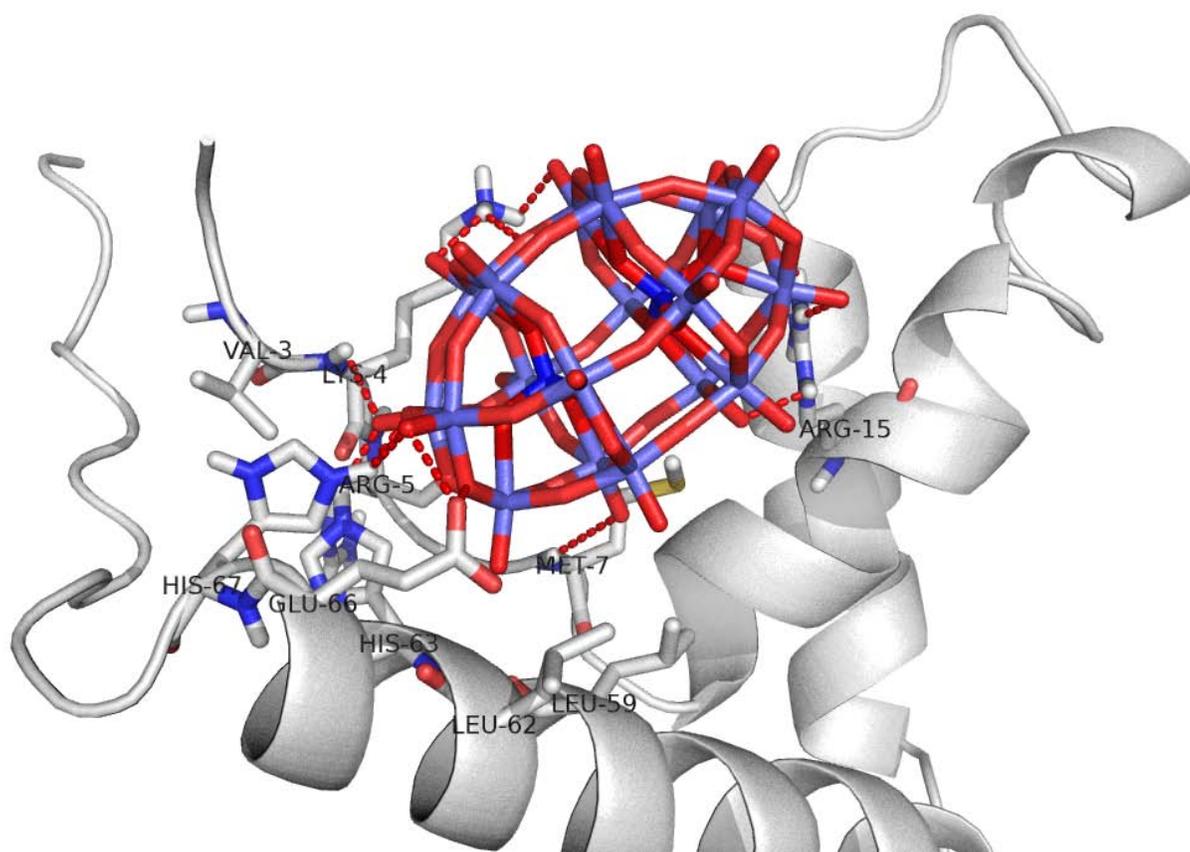
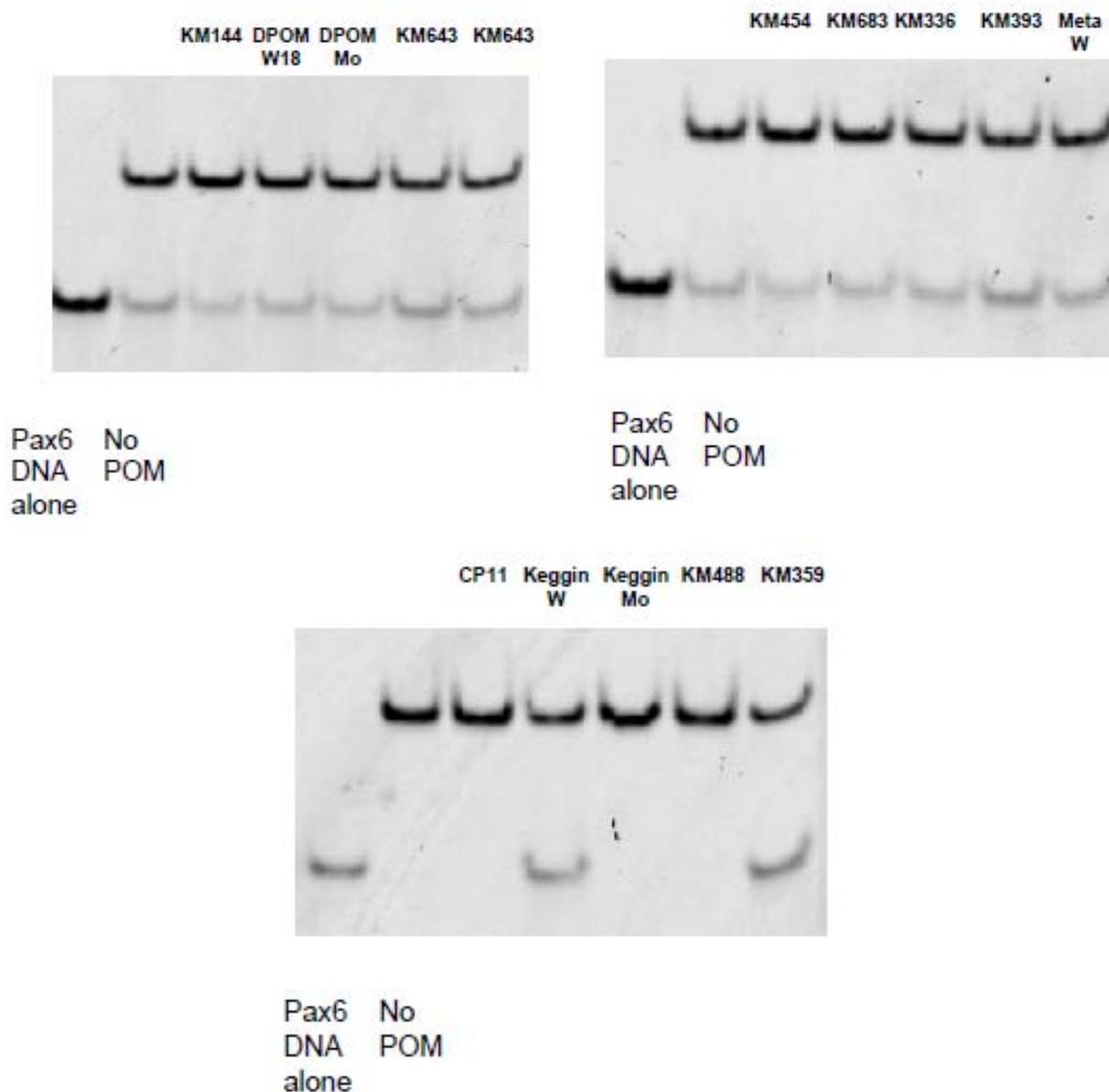


Figure. An expanded snapshot of the docked configuration of the Dawson-POM $K_6[P_2Mo_{18}O_{62}]$ with Sox2. Hydrogen bonds and electrostatic interactions less than 3.5\AA are shown in red dots. Residue numbering is based on the PDB structure 1GT0.

Appendix H



The inhibition of Pax6 by 15 different polyoxometalates studied using EMSA. DNA binding activity was estimated from maximally bound Pax6-DNA (no POM) and free DNA gel-shift intensities (Pax6 DNA alone).

Appendix I

Appendix Table 2. P-value of two-tailed, unpaired T-test (assuming equal variance) on residual DNA binding activities of 15 TFs upon $K_6[P_2Mo_{18}O_{62}]$ treatment. P-values less than 0.05 were taken as being statistically significant.

	Sox2	Sox4	Sox5	Sox6	Sox7	Sox8	Sox9	Sox10	Sox11	Sox17	Sox18	AP2	REST	Foxa1
Sox2	1	0.0808	0.9486	0.3716	0.011	0.7415	0.1388	0.3531	0.1933	0.4389	0.0292	0.3049	0.0967	0.9152
Sox4		1	0.0473	0.0074	0.1832	0.0881	0.337	0.1934	0.938	0.6116	0.6237	0.0053	0.8675	0.0377
Sox5			1	0.2873	0.005	0.6408	0.0597	0.2383	0.1557	0.3944	0.0125	0.2081	0.0669	0.8194
Sox6				1	0.0007	0.1248	0.001	0.0215	0.0576	0.1934	0.001	0.7134	0.0175	0.0805
Sox7					1	0.0087	0.0191	0.0158	0.2947	0.1918	0.2815	0.0005	0.3214	0.0029
Sox8						1	0.1483	0.4683	0.2354	0.5354	0.0249	0.0851	0.1115	0.7349
Sox9							1	0.4403	0.5964	0.972	0.0807	0.0002	0.3488	0.0266
Sox10								1	0.4049	0.7898	0.0546	0.0123	0.2167	0.2212
Sox11									1	0.7045	0.6773	0.0493	0.8412	0.1613
Sox17										1	0.4212	0.1732	0.5587	0.4273
Sox18											1	0.0006	0.8242	0.0067
AP2												1	0.014	0.0347
REST													1	0.0623
Foxa1														1