

# **Divalent Sulfur mediated interactions in proteins architecture, stability and molecular recognition**

A Thesis submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

**Sanket Satish Shelke**

**20151169**



Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Thesis Supervisor

Dr. Saikrishnan Kayarat

Associate professor

Dept. of Biology

IISER Pune

Thesis Advisor

Dr. M. S. Madhusudhan

Associate professor

Dept. of Biology

IISER Pune

©Sanket Satish Shelke 2020

All rights reserved

# Certificate

This is to certify that this dissertation entitled "**Divalent Sulfur mediated interactions in proteins architecture, stability and molecular recognition.**" towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by **Sanket Satish Shelke (Reg ID: 20151169)** at Indian Institute of Science Education and Research under the supervision of **Dr. Saikrishnan Kayarat**, Associate professor, Department of Biology, IISER Pune, during the academic year 2019-2020.



Thesis Supervisor

Dr. Saikrishnan Kayarat

Associate professor

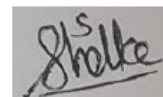
Dept. of Biology

IISER Pune

Date: 08/04/2020

# Declaration

I hereby declare that the matter embodied in the report entitled “**Divalent Sulfur mediated interactions in proteins architecture, stability and molecular recognition.**” are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of **Dr. Saikrishnan Kayarat** and the same has not been submitted elsewhere for any other degree.



Sanket Satish Shelke.

Reg ID: 20151169

Date: 08/04/2020

## Table of Contents

Abstract	06
Acknowledgments	07
Introduction	08
Materials and Methods	11
Results and Discussion	24
References	38

## List of Tables

Table No	Title of Table	Page No
1	List of primers used for cloning.	22
2	Classification of data based on electronic nature of S.	25
3	Classification of PDB data of M-S-C fragments forming H-bond.	28
4	A summary of PDB analyses performed to investigate role of H-bond and Ch-bond interaction $\alpha$ -helix capping in proteins.	29
5	Distribution of CXXXXC motif in various cases.	31
6	A summary of PDB analyses performed to investigate H-bond and Ch-bond interactions in intra- $\alpha$ -helix and $\beta$ -strand.	32

## List of Figures

Figure No	Title of Figure	Page no
1	Pictorial presentation of relative position of $\sigma$ -hole and electron lone pairs.	8
2	Definition of geometrical parameters $d$ , $\theta$ and $\delta$	12
3	Histogram showing relative preference for S...O/N and S...H-O/N interactions by divalent S in various chemical environment.	25
4	A Schematic presentation of S...H-N interaction between $C_i^{\text{th}}$ and $N_{i+n}^{\text{th}}$ residues under investigation along with definition of Ramachandran angles and geometry of the interaction.	27
5	Histogram showing frequency of appearance of S...H-N and S...O interaction at N and C-terminus $\alpha$ -helix capping.	29
6	Representative examples of examples of CXXXXC motif and Ramachandran angles distribution of the residues involved	30
7	Methionine Phosphinate bound active site of Methionyl tRNA Synthase (MetRS).	33
8	SDS PAGE Gel showing purity of protein after purification and which collected fraction has maximum protein eluted after passing through ion exchange column MonoQ.	33
9	Circular Dichroism (CD) absorption spectra of MetRS Wt and mutant Y260F.	34
10	Graph showing comparison of increase in Fluorescence intensity with respect to increasing ligand (Met) concentration in wt and mutant Y260F MetRS.	35

# Abstract

Non-covalent interactions are crucial for protein folding and stability. Traditionally, hydrogen bonding (H-bond), hydrophobic, and stacking interactions are well studied in biomolecules. Divalent Sulfur (S), which is present in small organic molecules, ligands and in proteins, also has the ability to form non-covalent interactions called chalcogen interaction (Ch-bond) and H-bond. In general, Ch-bond is made between S and nucleophiles. However, these S-mediated interactions remain unnoticed in biomolecules. In this study, we addressed the role of Ch-bond in protein structure and its effect on protein stability through extensive computational and bioinformatics analyses of high-resolution protein structures available in Protein Data Bank (PDB). This study gives unprecedented insights into the role of S present in methionine and cysteine on protein architecture. Here we showed that, H- and Ch-bond made by S can involve in capping of terminus of the  $\alpha$ -helices. Along with this, we also showed that Ch-bond can stabilize regular and non-regular secondary structural elements of proteins. In addition to the computational analyses, we also carried out biophysical and biochemical experiments to find role of Ch-bond in protein-ligand interaction, if any. For this purpose we selected methionyl-tRNA synthase (MetRS) as a model system. We found that, disruption of Ch-bond caused four-fold reduction in binding of the methionine to MetRS, demonstrating the importance of Ch-bond in ligand binding.

# Acknowledgements

First, I would like to express my heartfelt gratitude and sincere indebtedness to my mentor Dr Saikrishnan Kayarat for giving me this stupendous opportunity to carry out my thesis project with him and for his constant support. Especially hour-long morning discussions we had were constructive and insightful. I would next like to express my gratitude to my TAC member Dr M. S. Madhusudhan, for his support and thoughtful suggestions. I would like to express my appreciation to Dr Gayathri Pananghat for her advice during lab meetings.

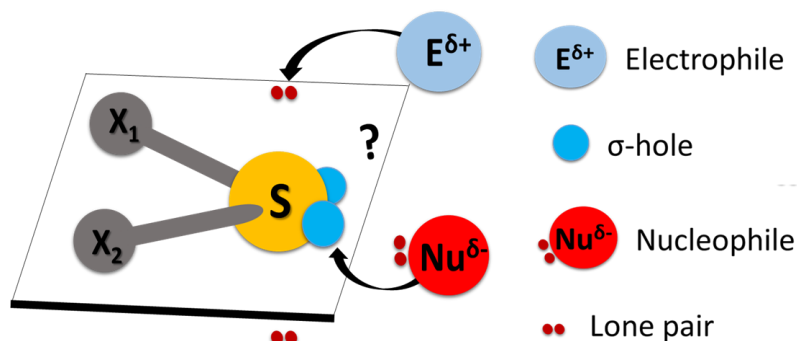
I would like to extend my gratitude to my friend Vishal Adhav. Vishal and I were working on the same project. He taught me the experiments and techniques when I joined the lab. He has guided me and helped me throughout the project. Next, I would like to thank Mahesh, Om, Sujata, Shekhar, Mrunmayee, Manil and other SK-G3 lab members for making life at the lab enjoyable.

Friends are the greatest assets I have gained during my time at IISER. Without them, life at IISER is something I cannot imagine. I would like to thank Yuvraaj, Digvijay and other friends at IISER in the memory of all the great time we had. Most importantly, I would like to express my gratitude to my family for their love, support and sacrifices they made.

Whenever my mind was clouded with doubts about my capability, all these people reassured and directed me on the right path. Part of the journey is the end. When I drift off, I will always remember these great people and beautiful memories.

# Introduction

Non-covalent interactions have always been a matter of attraction in the field of biology. It is not only because of the key role they play in maintaining structure but also they regulate the activity of biomolecules. There can be attractive or repulsive forces between molecules or atom and in general attractive forces increases the overall stability of molecules. Commonly known non-covalent interactions in biomolecules are van der Waals interactions, hydrogen bonding, stacking interactions and hydrophobic interactions (Dill et al., 2012; Nick Pace et al., 2014). These interactions are crucial for maintaining the overall architecture of the biomolecules along with the molecular recognition phenomena. For example, hydrogen bond is important for formation different protein secondary structures such as  $\alpha$ -helix and  $\beta$ -sheet. In addition, stacking interaction are important for double helical structure of DNA. In general, various elements such as C, N, O and H are abundantly present the biomolecules. These atoms are also involved in various polar and non-polar interactions.



**Figure 1.** Pictorial presentation of relative position of  $\sigma$ -hole and electron lone pairs on S.

Apart from these atoms, divalent S can also be found in proteins. However, its frequency is only significant when compared with other atoms. In proteins, divalent S is present in side chain of amino acids such as methionine and cysteine. This S can also be part of various non-covalent interactions and graced with the unique ability to interact with both an electrophile as well as a nucleophile. Divalent S has two lone-pair regions that are negatively charged and can potentially interact with



electrophiles. Interestingly, in addition to these negatively charged regions, S also has a pair of positively charged regions that are present along each covalent bond (Rosenfield et al., 1977). These two positively charged regions are called  $\sigma$ -holes (Refer Figure 1) (Kilgore and Raines, 2018; Murray et al., 2012; Politzer et al., 2017). These regions are mainly generated because of the large polarizability of S. These  $\sigma$ -holes allows S to interact with nucleophiles.

The relative orientation of lone-pairs and  $\sigma$ -holes on S are well studied in literature and summarized in Figure 1. A lone-pair region on S are located above and below of the plane made by C, S and C. However, two  $\sigma$ -holes lie on the plane (Refer Figure 1). As the lone-pairs and  $\sigma$ -holes regions have directional behaviour hence, interaction made by these regions are also directional. Lone-pairs of S can interact with H to form H-bond. Similarly,  $\sigma$ -holes can also interact with nucleophiles such as O or N. These interactions are in general called as '**Chalcogen interaction**' or '**Chalcogen bond**' (Ch-bond). The Ch-bonds, however, are relatively new advancement in the field and its impact on biomolecules is not clear.

### **Aim and Scope of thesis**

Using statistical investigations, existence of S mediated Ch-bond has been reported previously in proteins (Iwaoka et al., 2002; Taylor et al., 2012). However, the nature of S mediated Ch-bond and its precise role in proteins is yet to be characterized. As discussed in previous section, S can interact with electrophiles as well as nucleophiles. This thesis aims to address the origin selectivity between H-bond (S-electrophile interaction) and Ch-bond (S-nucleophile interaction) in proteins. This analysis was carried by probing directional behaviour of these interactions on high-resolution protein structures from the Protein Data Bank (PDB). Apart from this, we also aim to understand the role of the Ch-bond in protein architecture, if any. This was performed analysing various secondary structural elements of protein, which has local Ch-bonds.

In addition to the statistical analysis of high-resolution protein structures, it was very tempting to study the functional relevance of Ch-interaction in proteins. As no attempt were made in literature to study the role of  $\sigma$ -hole mediated Ch-bond in biochemical processes. Towards this goal, we also planned for biophysical and

biochemical experiments to understand enzyme substrate recognition. For this purpose we selected methionyl-tRNA synthase also called as MetRS as a model system. This enzyme catalyses acylation of methionine (Met) to its cognate tRNA. Acylation happens through a two-step mechanism beginning with activation of the methionine by ATP to yield aminoacyl adenylate followed by transfer onto the 3'-end of the tRNA molecule. This enzyme has been studied thoroughly because of its involvement in translating genetic code into proteins (Deniziak and Barciszewski, 2001). Earlier studies have concluded that binding of methionine to the enzyme occurs through induced-fit mechanism. Aromatic residues at the core undergo large rearrangement that enables the enzyme to create a hydrophobic pocket around the methionine side chain. Residues crucial for binding of methionine include main-chain NH of L13, side-chain of Y260 and H301 assists the recognition of the S from methionine (Mechulam et al., 1999; Serre et al., 2001). Here, we speculate that –OH group of Y260 forming Ch-bond with S and perhaps, recognises S from methionine.

The overall aim to characterize divalent S mediated interaction in protein which subsequently would have an importance in fields like proteins designing, developing models or computer programmes that predict the structure of proteins. In addition, experimental validation of role played by Ch-bond in molecular recognition would be important in field of drug designing.

# Materials and Methods

The study of divalent S mediated interactions in proteins was divided into two parts: **a.** Statistical analyses of the high-resolution structures available in Protein Data Bank (PDB) and **b.** Experimentally studying the role of S mediated Ch-bond in molecular recognition of ligand for the chosen protein system.

## Materials and methods of computational section:

Structural data of the proteins determined using X-ray crystallography was downloaded in January 2018. The essential information in this data is 3D coordinates of the atoms constituting biological macromolecule. In addition to the atom coordinates, this data also includes the name of molecules, primary and secondary structural information, sequence database references, details about data collection and structure solution. This downloaded data was divided into two sets according to following criterion:

### Set 1:

Resolution  $\leq 2.0$  Å, pairwise sequence identity  $\leq 90\%$ ,  $R_{\text{factor}} \leq 25\%$  and file size  $\leq 1$  MB.

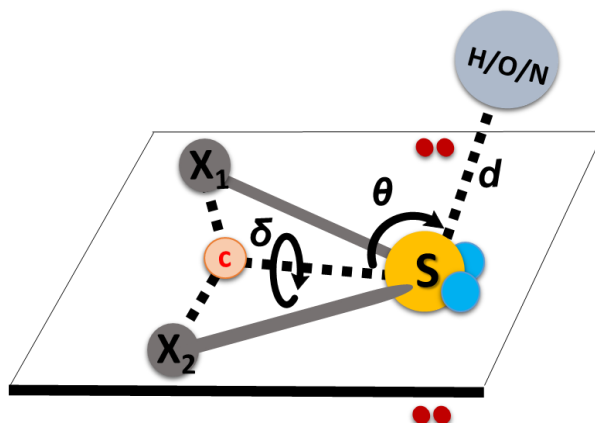
### Set 2:

Resolution  $\leq 2.5$  Å, pairwise sequence identity  $\leq 90\%$ ,  $R_{\text{factor}} \leq 30\%$  and file size  $\leq 1$  MB.

Set 1 resulted in 16851 PDB files whereas Set 2 had 25423 PDB files. For filtering PDB files based on the above mentioned criteria, PISCES (Wang and Dunbrack, 2003) programme was used. In these datasets, we searched for S having distance  $\leq 3.32$  Å with O (van der Waals radius of S + van der Waals radius of O =  $1.80$  Å +  $1.52$  Å =  $3.32$  Å (Craik, 2008)). Similarly, S having distance  $\leq 3.35$  Å with N (van der Waals radius of S + van der Waals radius of N =  $1.80$  Å +  $1.55$  Å =  $3.35$  Å) (Craik, 2008).

Along with these distance criteria, we also applied angular parameters  $\theta$  and  $\delta$  (Figure 2). This was mainly to study directional behaviour of S mediated interactions. Pictorial representation of all these geometric parameters are shown in

Figure 2. For all analyses, we used S from either methionine, disulfide bonded cysteine, metal chelated cysteines or S of ligands.  $X_1$  and  $X_2$  are atoms covalently bonded with the same S. Angle between vectors CS and SO was called  $\theta$  that ranges from  $0^\circ$  to  $180^\circ$ . Whereas, dihedral angle  $\delta$  was angle between two planes. First plane considered for calculation of  $\delta$  comprises of points  $X_1/X_2$ , C and S whereas second plane comprises of points C, S and the atom interacting with S (H/O/N).  $\delta$  ranges from  $-90^\circ$  to  $+90^\circ$  (Figure 2).



**Figure 2:** Definition of geometrical parameters  $d$ ,  $\theta$  and  $\delta$ .

To calculate directional parameters mentioned above following formulae were used:

1. Formula for calculating distance between two points:

$$P1 = (x1, y1, z1)$$

$$P2 = (x2, y2, z2)$$

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2}$$

Formula in python:

```
p1 = np.array([x1, y1, z1])
p2 = np.array([x2, y2, z2])
d = round(np.sqrt(np.sum((p1 - p2)**2)),2)
```

2. Formula for calculating angle between two vectors:

$$P1 = (x1, y1, z1)$$

$$P2 = (x2, y2, z2)$$

$$P3 = (x3, y3, z3)$$

$$v1 = (x1-x2, y1-y2, z1-z2) = (a, b, c)$$

$$v2 = (x3-x2, y3-y2, z3-z2) = (d, e, f)$$

$$v1 = \sqrt{(a)^2 + (b)^2 + (c)^2}$$

$$v2 = \sqrt{(d)^2 + (e)^2 + (f)^2}$$

$$dp = v1 \cdot v2 = (a \times d) + (b \times e) + (c \times f)$$

$$\cos \theta = \frac{dp}{v1 \times v2}$$

$$\theta = \cos^{-1}(\cos \theta)$$

Formula in Python:

```

p1 = np.array([x1, y1, z1])
p2 = np.array([x2, y2, z2])
p3 = np.array([x3, y3, z3])
v1 = p1 - p2
v2 = p3 - p2
def Angle(v1, v2):
    len_a = np.sqrt(np.sum((a)**2))
    len_b = np.sqrt(np.sum((b)**2))
    dot = np.dot(a,b)
    cos = dot / (len_a*len_b)
    theta = round(math.degrees(np.arccos(cos)))
    return theta

```

3. Formula of calculating torsional angle:

$$P1 = (x1, y1, z1)$$

$$P2 = (x2, y2, z2)$$

$$P3 = (x3, y3, z3)$$

$$P4 = (x4, y4, z4)$$

$$v1 = (x1-x2, y1-y2, z1-z2) = (a, b, c)$$

$$v2 = (x2-x3, y2-y3, z2-z3) = (d, e, f)$$

$$v3 = (x3-x4, y3-y4, z3-z4) = (g, h, i)$$

$$n1 = v1 \times v2 = \begin{matrix} x & y & z \\ a & b & c \\ d & e & f \end{matrix} = (a \cdot f - c \cdot e, c \cdot d - a \cdot f, a \cdot e - b \cdot d) = (j, k, l)$$

$$n2 = v2 \times v3 = \begin{matrix} x & y & z \\ d & e & f \\ g & h & i \end{matrix} = (d \cdot i - f \cdot h, f \cdot g - d \cdot i, d \cdot h - e \cdot g) = (m, n, o)$$

$$v1 = \sqrt{(a)^2 + (b)^2 + (c)^2}$$

$$v2 = \sqrt{(d)^2 + (e)^2 + (f)^2}$$

$$dp = v1.v2 = (a \times d) + (b \times e) + (c \times f)$$

$$\cos \varphi = \frac{dp}{v1 \times v2}$$

$$\varphi = \cos^{-1}(\cos \varphi)$$

Formula in Python:

```
p1 = np.array([x1, y1, z1])
p2 = np.array([x2, y2, z2])
p3 = np.array([x3, y3, z3])
p4 = np.array([x4, y4, z4])
def Dihedral(p1, p2, p3, p4):
    q1 = p2 - p1
    q2 = p3 - p2
    q3 = p4 - p3
    c1 = np.cross(q1,q2)
    c2 = np.cross(q2,q3)
    n1 = c1/np.sqrt(np.dot(c1,c1))
    n2 = c2/np.sqrt(np.dot(c2,c2))
    u1 = n2
    u3 = q2/(np.sqrt(np.dot(q2,q2)))
    u2 = np.cross(u3,u1)
    a = np.dot(n1,u1)
    b = np.dot(n1,u2)
    phi = round(np.degrees(-math.atan2(a,b)),1)
    return phi
```

### Python Scripts:

All the required information about secondary structural elements was extracted from PDB coordinate files. Simplified python programs are given below. In addition, each defined function in the program is explained in simple language. Based on logic and algorithm applied in these two programs all other programs used at various stages of analyses were developed. Script-1 is the python program that found Ch-bond based on distance criteria. Whereas Script-2 is the python program that calculated defined angular parameters  $\theta$  and  $\delta$ . For Script-1, input was list of PDB files while for Script-2 input was the output of Script-1.

## Script-1

```
#Importing required Python modules
import os
import numpy as np
import pandas as pd
import multiprocessing as mp

#Defining path of the folder where PDB files are stored
path = r'Paste_Path_Here'

#Making list of PDB files needed for further analysis
files = []
with open('txt_file_with_PDBlist.txt', 'r', encoding='utf-8-sig') as
txt:
    for line in txt:
        files.append(line)

#Defining first function that will extract coordinate data from PDB
file in form of data frame
def Pick_up(file):
    filename = os.path.join(path,file)
#head is the list of the names that will be given to the columns in
the pdb files data
    head = ['MODEL', 'ATOM NO', 'ATOM ID', 'Confmn', 'RESIDUE', 'RES
SEQ', 'RES INSERT','X AXIS', 'Y AXIS', 'Z AXIS', 'NR', 'Atm']
#numb is the list of columns in pdb data which are not numeric
    numb = ['MODEL', 'ATOM ID', 'Confmn', 'RESIDUE', 'RES SEQ',
'NR', 'Atm']
#spaces is the list of intervals in terms of spacing in pdb files to
define the width of each column
    spaces = [(0, 6), (6, 11), (12, 16), (16, 17), (17, 20), (21,
22), (22, 26), (31, 38), (38, 46), (46, 54), (54, 78), (76, 78)]
#based upon information in lists defined above we are going to
extract atoms
#coordinate data in form of data frame named df
    df = pd.read_fwf(filename, colspecs=spaces, names=head)
    df = df[df['MODEL'] == 'ATOM'].drop(['MODEL', 'NR'], 1)
    for j in set(df.columns) - set(numb): df[j] =
pd.to_numeric(df[j])
#if some part of the data frame is left blank then it will be
replace with 0 using function 'replace'.
    df = df.replace(np.nan,'0')
    df = df.loc[ df['Confmn'] == '0']
#To make separate data frame of atom Sulfur and Oxygen we can sort
mother data frame df into two data frame named SD and O by picking
up atom with Atom ID 'SD' and 'O'.
    SD = df.loc[ df['ATOM ID'] == 'SD']
    O = df.loc[ df['ATOM ID'] == 'O']
```

```

#OutPuts of the function Pick_up(file) are going to be two data
frames named SD and O
    return SD, O

#Defining second function that will calculate distance between all
possible pairs of Sulfur and Oxygen.
def Contact(SD, O, file):
#contact is going to be output list that will contain list of
positive Sulfur Oxygen contacts
    contacts = []
#for loop that will make all possible pairs of Sulfur and Oxygen
    for i1, row1 in SD.iterrows():
        for i2, row2 in O.iterrows():
#to exclude Sulfur and Oxygen from same residue and next consecutive
residue following if criterion is applied
            if (row1['RES INSERT'] + 1 < row2['RES INSERT'] or
row1['RES INSERT'] - 1 > row2['RES INSERT']):
#p1 and p2 are the 3d coordinates of the Sulfur and Oxygen
respectively
                p1 = np.array([row1['X AXIS'],row1['Y AXIS'],row1['Z
AXIS']])
                p2 = np.array([row2['X AXIS'],row2['Y AXIS'],row2['Z
AXIS']])
#calculating distance (d Å) between Sulfur and Oxygen
                d = round(np.sqrt(np.sum((p1 - p2)**2)),2)
#if distance d is less than or equal to 3.32 Å that pair of Sulfur
and Oxygen will be considered for further geometric calculations
else will be excluded
                    if d <= 3.32:
#details of the Sulfur and Oxygen following distance criterion will
be put in a output list lst that is later appended into the main
list contacts
                        lst = [file, 'SD', row1['RES SEQ'], row1['RES
INSERT'], 'O', row2['ATOM ID'], row2['RESIDUE'], row2['RES SEQ'],
row2['RES INSERT'], 'd', d]
                        str1 = '\t'.join(str(e) for e in lst)
                        contacts.append(str1)
    return contacts
#final function that will run all above defined functions
sequentially
def Finale(file):
    Result = []
    SD, O = Pick_up(file)
    contacts = Contact(SD, O, file)
    [Result.append(l) for l in contacts]
    return Result

```



```
#To speed up the programme we will use multiprocessing module here
which will deploy defined number of CPU cores to this programmes
allowing us to analyse multiple files at ones
```

```
result = []
if __name__ == "__main__":
    pool = mp.Pool(processes=45)
    result[:] = pool.map(Finale, files)
```

## Script-2

```
#Importing required Python modules
```

```
import os
import numpy as np
import pandas as pd
import multiprocessing as mp
import math
```

```
#Defining path of the folder where PDB files are stored
```

```
path = r'Paste_Path_Here'
```

```
#Making list S-O contacts that followed distance criterion
```

```
contacts = []
with open('TxtFileWith_S-O_Contact_List.txt', 'r', encoding='utf-8-
sig') as txt:
    for line in txt:
        contacts.append(line)
```

```
#Defination of python function with formula to calculate angle
```

```
def Angle(v1, v2):
    len_a = np.sqrt(np.sum((a)**2))
    len_b = np.sqrt(np.sum((b)**2))
    dot = np.dot(a,b)
    cos = dot / (len_a*len_b)
     $\theta$  = round(math.degrees(np.arccos(cos)))
    return  $\theta$ 
```

```
#Defination of python function with formula to calculate dihedral
angle
```

```
def Dihedral(p1, p2, p3, p4):
    q1 = p2 - p1
    q2 = p3 - p2
    q3 = p4 - p3
    c1 = np.cross(q1,q2)
    c2 = np.cross(q2,q3)
    n1 = c1/np.sqrt(np.dot(c1,c1))
    n2 = c2/np.sqrt(np.dot(c2,c2))
    u1 = n2
    u3 = q2/(np.sqrt(np.dot(q2,q2)))
```

```

u2 = np.cross(u3,u1)
a = np.dot(n1,u1)
b = np.dot(n1,u2)
phi = round(np.degrees(-math.atan2(a,b)),1)
return phi

#function that will convert phi values in range of -90 to +90
def PhiConversion(phi1):
    if 90 >= phi1 >= -90:
        phi = phi1
    elif phi1 > 90:
        phi = 180 - phi1
    elif phi1 < -90:
        phi = -180 - phi1
    return phi

#Defining first function that will extract coordinate data from PDB
file in form
#of data frame
def Pick_up(file):
    filename = os.path.join(path,file)
#head is the list of the names that will be given to the columns in
the pdb
#files data
    head = ['MODEL', 'ATOM NO', 'ATOM ID', 'Confmn', 'RESIDUE', 'RES
SEQ', 'RES INSERT','X AXIS', 'Y AXIS', 'Z AXIS', 'NR', 'Atm']
#numb is the list of columns in pdb data which are not numeric
    numb = ['MODEL', 'ATOM ID', 'Confmn', 'RESIDUE', 'RES SEQ',
'NR', 'Atm']
#spaces is the list of intervals in terms of spacing in pdb files to
define the
#width of each column
    spaces = [(0, 6), (6, 11), (12, 16), (16, 17), (17, 20), (21,
22), (22, 26), (31, 38), (38, 46), (46, 54), (54, 78), (76, 78)]
#based upon information in lists defined above we are going to
extract atoms
#coordinate data in form of data frame named df
    df = pd.read_fwf(filename, colspecs=spaces, names=head)
    df = df[df['MODEL'] == 'ATOM'].drop(['MODEL', 'NR'], 1)
    for j in set(df.columns) - set(numb): df[j] =
pd.to_numeric(df[j])
#if some part of the data frame is left blank then it will be
replace with 0
#using function 'replace'.
    df = df.replace(np.nan,'0')
    df = df.loc[ df['Confmn'] == '0']

```

```

#To make separate data frame of atom Sulfur and Oxygen we can sort
mother data
#frame df into two data frame named SD and O by picking up atom with
Atom ID
#‘SD’ and ‘O’.
    SD = df.loc[ df['ATOM ID'] == 'SD']
    CG = df.loc[ df['ATOM ID'] == 'CG']
    CE = df.loc[ df['ATOM ID'] == 'CE']
    O = df.loc[ df['ATOM ID'] == 'O']
#OutPuts of the function Pick_up(file) are going to be two data
frames named SD
#and O
    return SD, CG, CE, O

#defining function Repick(SD, CG, CE, O, kst) that will extract
coordinates of
#the atoms required for further calculating angles  $\theta$  &  $\phi$ 
def Repick(SD, CG, CE, O, kst):
#Atoms is list of coordinates required for further calculations
    Atom = []
    for i1, row1 in SD.iterrows():
        if row1['RES SEQ']==kst[2] and row1['RES
INSERT']==int(kst[3]):
            lst = ['SG', row1['X AXIS'], row1['Y AXIS'], row1['Z
AXIS']]
            Atom.append(lst)
    for i1, row1 in CE.iterrows():
        if row1['RES SEQ']==kst[2] and row1['RES
INSERT']==int(kst[3]):
            lst = ['CE', row1['X AXIS'], row1['Y AXIS'], row1['Z
AXIS']]
            Atom.append(lst)
    for i1, row1 in CG.iterrows():
        if row1['RES SEQ']==kst[2] and row1['RES
INSERT']==int(kst[3]):
            lst = ['CG', row1['X AXIS'], row1['Y AXIS'], row1['Z
AXIS']]
            Atom.append(lst)
    for i1, row1 in O.iterrows():
        if row1['RES SEQ']==kst[5] and row1['RES
INSERT']==int(kst[6]):
            lst = ['O', row1['X AXIS'], row1['Y AXIS'], row1['Z
AXIS']]
            Atom.append(lst)
    return Atom

```

```

#defining function direction that will calculate angles  $\theta$  &  $\phi$  using
coordinates
#from list Atom
def direction(Atom):
    SD = np.array([Atom[0][1],Atom[0][2],Atom[0][3]])
    CG = np.array([Atom[1][1],Atom[1][2],Atom[1][3]])
    CE = np.array([Atom[2][1],Atom[2][2],Atom[2][3]])
    O = np.array([Atom[3][1],Atom[3][2],Atom[3][3]])
    C = (SD + CG + CE)/3
    v1 = C - SD
    v2 = O - SD
#Using predefined angle function to calculate angle  $\theta$ 
     $\theta$  = Angle(v1, v2)
    p1 = CE
    p2 = C
    p3 = SD
    p4 = O
#Using predefined angle function to calculate dihedral angle  $\phi$ 
     $\delta$ 1 = Dihedral(p1, p2, p3, p4)
     $\delta$  = PhiConversion( $\phi$ 1)
    return  $\theta$ ,  $\delta$ 

#defining function that will check if calculated angles  $\theta$  &  $\phi$  are in
the defined
#range or not
def Chalc( $\theta$ ,  $\delta$ , kst):
    res = []
    if 115 <=  $\theta$  <= 155 and -50 <=  $\delta$  <= 50:
        lst = kst + [' $\theta$ - $\delta$ ' ,  $\theta$ ,  $\delta$ ]
        str2 = '\t'.join(str(e) for e in lst)
        res.append(str2)
    else:
        pass
    return res

#final function that willrun all above defined functions
sequentially
def Finale(contact):
    Result = []
    contact = contact.split()
    file = kontakt[0]
    SD, CG, CE, O = Pick_up(file)
    Atom = Repick(SD, CG, CE, O, kontakt)
     $\theta$ ,  $\delta$  = direction(Atom)
    res = Chalc( $\theta$ ,  $\delta$ , kst)
    [Result.append(l) for l in res]
    return Result

```

```
#to speed up the programme we will use multiprocessing module here
which will
#deploy defined number of CPU cores to this programmes allowing us
to analyse
#multiple files at ones
result = []
if __name__ == "__main__":
    pool = mp.Pool(processes=45)
    result[:] = pool.map(Finale, contacts)
```

## Materials and Methods of Experimental Study:

### Cloning of MetRS and its Y260F mutant:

From genomic DNA of *E.coli* K12 strain, we amplified coding region of truncated MetRS (amino acid from 1 to 551). This amplification was performed using primers mentioned in Table 1. This amplified gene was subsequently cloned into an expression vector pRSF using digestion-ligation protocol as mention on NEB traditional quick guide. NcoI and XhoI restriction sites were used for this purpose. This ligated product was transformed in the *E. coli* NEB turbo electrocompetent cells by electroporation. After transformation, the positive clone was screened on LB agar plates containing kanamycin as selection marker. The positive clones were checked by full gene sequencing. Y260F mutant of MetRS was amplified from pRSF plasmid coding for wild type construct. For this purpose we used primer 3 mentioned in Table 1. This amplified mutant gene was subsequently cloned in to pRSF vector using restriction free cloning technique.

**Table 1:** Primers Used for cloning wt and mutant Y260F in vector pRSF.

Primer Name	Primer Sequence (5' -> 3')
MetG Forward Primer	AACTTTAATAAGGAGATATACCATGGGCACTCAAGTCGC GAAGAAAATTCTGGTGA
MetG Reverse Primer	CAGCAGCGGTTTCTTTACCAGACTCGAGTTATTTTACTT CTTCTTTAGAGGCTT
MetG <sup>Y260F</sup> Forward Primer	TGGCTGGACGCACCGATTGGCTTTATGGGTTCTTTCAA GAATCTG

### Purification of MetRS and its MetRS<sup>Y260F</sup> mutant:

Wild type and mutant of MetRS were purified using predefined protocols (Serre et al., 2001). Protein was overexpressed in 1L culture of *E. coli* BL21 DE3 at 37°C using 0.01mM IPTG when OD reached to 0.6. Induced cell cultures were harvested by centrifugation past 3 hours of incubation at 37°C. After harvesting cell

culture, pellet was suspended in 150 ml of lysis buffer (20 mM Tris-HCl pH 7.6, 2 mM DTT, 0.1 mM EDTA). All steps of protein purification were carried out at 4°C. Then cells were submitted to ultrasonic disintegration (pulse was on for 1 sec & off for 3 sec with total on time of 3 min at 60% amplitude). Followed by centrifugation of the lysate at 12000 rpm. To eliminate nucleic acids, 3 % w/v of streptomycin sulphate was used. Then supernatant was fractioned by ammonium sulphate precipitation (35-70% w/v). The re-suspended pellet was dialysed against buffer B50 (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA and 1 mM DTT) for 1 hour then loaded on MonoQ 10/100 GL (GE Healthcare) ion exchange column equilibrated with same buffer. Elution was performed by applying a 0 mM to 300 mM linear NaCl gradient. At the end, size exclusion chromatography was performed using Superdex75 column. Potassium phosphate buffer (10 mM potassium phosphate, 2 mM DTT) was used as elution buffer for size exclusion. Finally, wild type MetRS and its mutant were concentrated to 20 mg/ml and stored at -80°.

#### **CD Spectroscopy:**

To confirm if purified MetRS is well folded and whether point mutation Y260F affects folding, we recorded CD absorption spectra from 200 nm to 250 nm. For recording CD absorption spectra, Jasco's J-1500 CD spectrophotometer was used.

#### **Binding Assay:**

As reported earlier MetRS has intrinsic fluorescence and it increases upon ligand binding. For binding assay Horiba's Fluoromax steady state and lifetime benchtop spectrofluorometer was used. Tryptophan residues in wild type MetRS and its mutant Y260F were excited at 295 nm. Intensity of emission maximum at 330 nm was monitored by varying ligand concentration. These binding experiments were carried at 25°C. Protein was diluted to 0.8 µM final concentration in binding buffer (20 mM Tris-HCl pH 7.6, 0.1 mM EDTA, 10 mM DTT) with total reaction volume of 200 µl. All these binding conditions were optimized in lab. In all reaction mixture, protein concentration was kept constant (0.8µM) and methionine concentration was sequentially varied as 0 µM, 10 µM, 50 µM, 100 µM, 200 µM, 500 µM, 700 µM, 1000 µM, 1500 µM, 2000 µM. Data were analysed and plotted in Origin program.

# Results and Discussion

First step in this study was to understand the type interactions made by S with electrophile and nucleophile. The approaching electrophile and nucleophile can be covalently linked to each other or can be part of different molecule. For this purpose, we employed concept of interaction made by S as shown in Figure 1. As the protein structures considered for this study are determined by X-ray crystallography, they, in general, do not have positional information of hydrogens (H). Hence, it was not trivial to determine the nature of interaction i.e. H-bond or Ch-bond. To solve this ambiguity we used range of directional parameters ( $\theta$  and  $\delta$ ) as mention in method section. Our criteria to decide the nature of interaction in proteins were verified in small molecule structural database such as CSD, where position of H were experimentally determined. Based on this we decided following ranges of  $\theta$  and  $\delta$  values (Vishal Adhav, unpublished data).

For Ch-bond (S...O/N interaction):

$$115^\circ \leq \theta \leq 155^\circ \text{ and } -50^\circ \leq \delta \leq +50^\circ$$

For H-bond (S...H-O/N interaction):

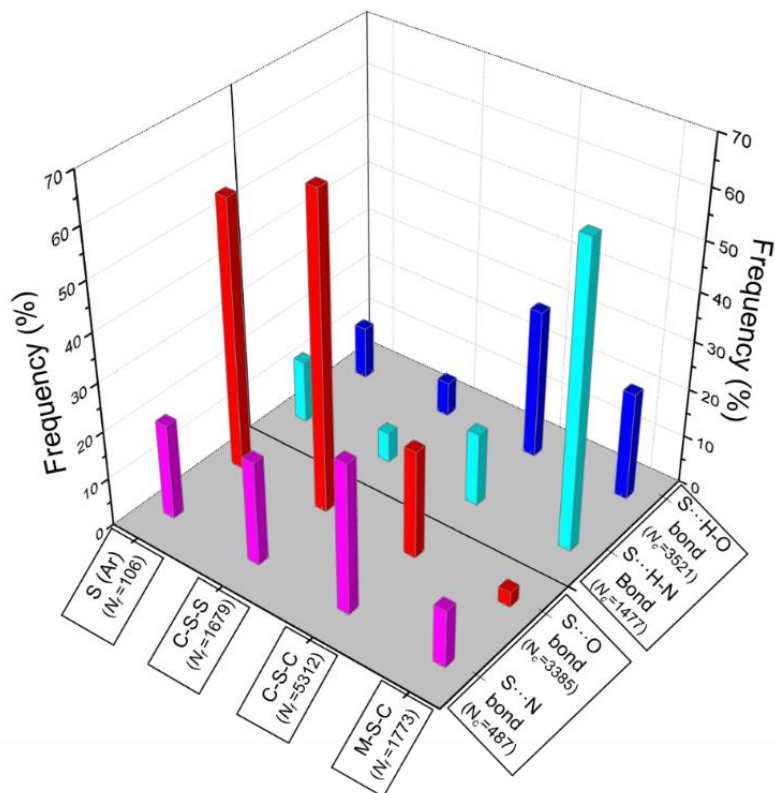
$$90^\circ \leq \theta \leq 140^\circ \text{ and } (-90^\circ \leq \delta \leq -50^\circ \text{ or } +50^\circ \leq \delta \leq +90^\circ)$$

## **What actuate the choice between formation of Ch-bond or H-bond?**

Initially, we asked what factors decide the formation of Ch-bond over H-bond or vice a versa. Towards this aim, we applied distance criteria followed by directional criteria as mention in methods for all interacting S with O/N. Here divalent S can come from a methionine, cystine, metal chelated cysteins or part of an aromatic ring for ligand. The frequency of appearance of S...O/N and S...H-O/N interaction was plotted as a function of S in various chemical environment and shown in Figure 3 and Table 2. From this analysis, it was clear that divalent S from aromatic ring and cystine had very high preference for Ch-bond over H-bond. Whereas, divalent S chelated with metal prefers to form H-bond and rarely forms Ch-bond. Interestingly, divalent S from methionine has almost equal preference to form Ch-bond as well as H-bond. From this differential preference with respect different fragments, we inferred that chemical



nature of divalent S was crucial factor for selection between H- and Ch-bond. These PDB analyses were resemble with CSD performed in lab (Vishal Adhav, unpublished data).



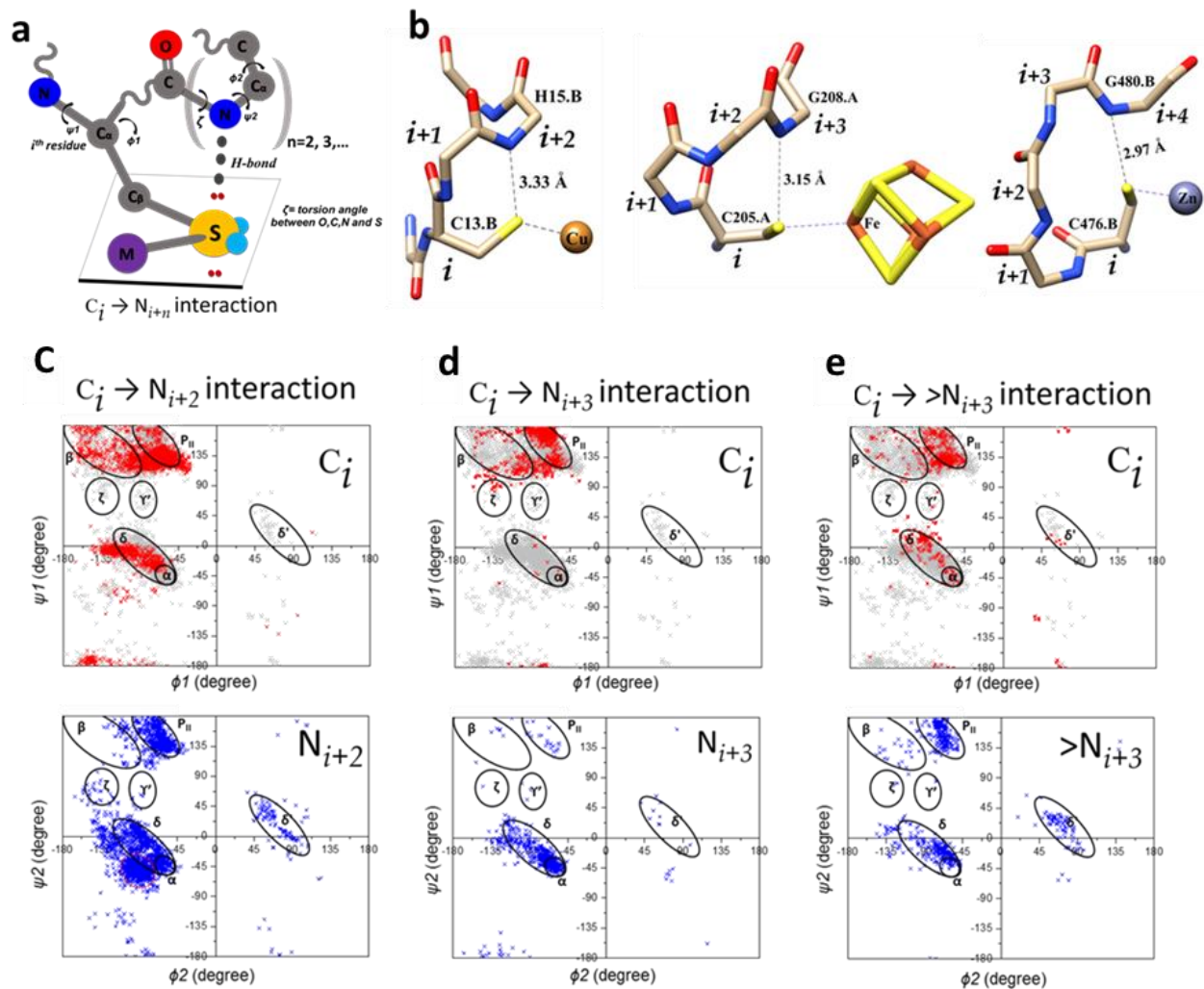
**Figure 3:** Histogram showing relative preference for S...O/N and S...H-O/N interactions by divalent S in various chemical environment. Where, M = Any metal and S(Ar)= S in aromatic rings.

**Table 2:** Classification of data based on electronic nature of S.

Fragment	S...H-O contacts (N <sub>c</sub> )	S...H-N contacts (N <sub>c</sub> )	S...O contacts (N <sub>c</sub> )	S...N contacts (N <sub>c</sub> )	Total Fragments (N <sub>f</sub> )
M-S-C	747	870	101	55	1773
C-S-C	2596	540	1815	361	5312
C-S-S	162	59	1391	67	1679
S (Ar)	16	8	78	4	106
Total	3521	1477	3385	487	8870

## **Role of H-bond formed between metal chelated S and backbone amino group**

As noted in previous section, metal chelated cysteine has high preference to form H-bond we sought for its role in protein architecture. One way to address this question was to study the role of this H-bond on backbone conformation of interacting residues using Ramachandran angles  $\phi$  &  $\psi$ . Hence, we sought for structures of metal chelated divalent S from cysteine forming H-bond with protein backbone amino group (refer method section). For this analysis, we referred cysteine chelated to metal as  $C_i$  and H-bond donor residue as  $N_{i+n}$ , where  $n$  can be  $\pm 2, \pm 3 \dots$  and so on. We categorised these H-bond in three groups based on the number of intervening residues between donor (N) and acceptor (C) (refer Table 3). These three groups were, 1)  $C_i$  to  $N_{i+2}$ ; 2)  $C_i$  to  $N_{i+3}$ ; and 3)  $C_i$  to  $>N_{i+3}$ . From the Ramachandran plot in Figure 4, we inferred that backbone conformation of two interacting residues was strongly affected by the number of residues intervening between the two. As in the Ramachandran plot shown in Figure 4, we noted that certain points had a strong preference to populate in certain region. For example, we found that backbone torsion angles of  $C_i$  residue in  $C_i \rightarrow N_{i\pm 2}$  case were primarily occupied in  $P_{II}$ , and  $\delta$ -region in the Ramachandran plot. Interestingly, the same  $C_i$  residue in  $C_i \rightarrow N_{i\pm 3}$   $C_i \rightarrow >N_{i\pm 3}$  cases preferentially fall in to  $P_{II}$  region. Also, backbone torsion angles of  $N_{i\pm 2}$  in case of  $C_i \rightarrow N_{i\pm 2}$ , we found that point were populating  $P_{II}$ ,  $\alpha$ ,  $\delta$  and a region adjacent to the bridge and the  $\alpha$ -helix, which we suggest should be referred to as  $\epsilon$ , in accordance with the nomenclature by Karplus (Hollingsworth and Karplus, 2010) (Refer Fig 4c). For  $N_{i\pm 3}$  in case of  $C_i \rightarrow N_{i\pm 3}$ , points populated  $\alpha$  and  $\delta$  regions of these residues. Interestingly the  $i+3^{rd}$  residues populating  $\alpha$ -region were often at the N-terminus of a helix. This metal chelated divalent S from cysteine forms H-bond with free backbone amino group of the N-terminal residues from  $\alpha$ -helix. These observations clearly showed that H-bond made by S had a strong influence on backbone torsion angle of proteins.



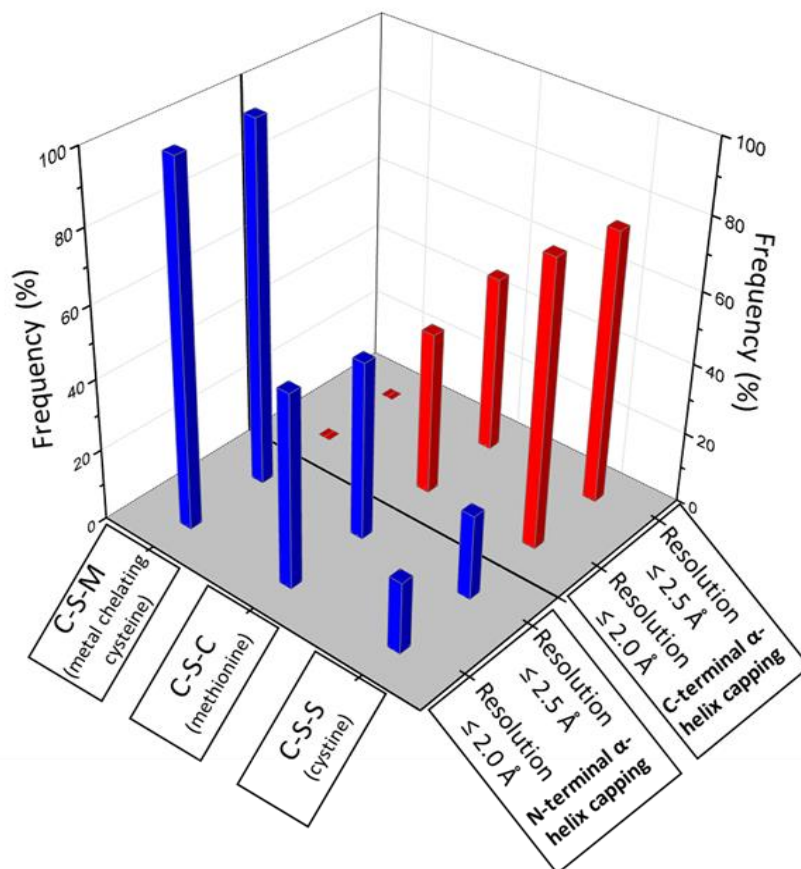
**Figure 4:** (a) A Schematic presentation of S...H-N interaction between  $C_i^{\text{th}}$  and  $N_{i+n}^{\text{th}}$  residues under investigation along with definition of Ramachandran angles and geometry of the interaction (C represent metal chelated cysteine and N represent any H-bond donor residue). (b) Some of the representative examples of S...H-N interaction between  $C_i^{\text{th}}$  and  $N_{i+n}^{\text{th}}$  residues (PDB IDs: 2QIF, 1ORN and 1LAT). (c-d) Ramachandran plot for metal chelating cysteine (9563 C-S-M fragments) shown as gray points and point those satisfied distance  $d_{S \cdots H} \leq 3.6 \text{ \AA}$  and directional criteria ( $90^\circ \leq \theta \leq 140^\circ$  and  $\pm 50^\circ \leq \delta \leq \pm 90^\circ$ ) for S...H-N interaction shown as red. While, for  $N_{i+n}^{\text{th}}$  residues, which act as H-donor in S...H-N interaction shown as blue.

**Table 3.** Classification of PDB data of M-S-C fragments forming H-bond.

Criterion	Set 1	Set 2
Metal-S-C fragments	5301	9563
$C_i \rightarrow N_{i+2}$ interaction	1918	3245
$C_i \rightarrow N_{i+3}$ interaction	498	920
$C_i \rightarrow >N_{i+3}$ interaction	354	709

### Role of S in $\alpha$ -Helix Capping

As mentioned in previous section, a large number H-bond made by S with backbone amino group were at N-terminus of  $\alpha$ -helix. In these cases, interacting cysteines were acting as N-terminus  $\alpha$ -helix capping residue. This prompted us to ask if C-helix can also stabilize the C-terminus of  $\alpha$ -helix i.e. capping C-terminus of  $\alpha$ -helix. The  $\alpha$ -helix capping is important concept in protein folding (Richardson Richardson, D.C, 1988). Residues in  $\alpha$  -helix are held together by H-bonds formed between  $i^{\text{th}}$  residue's backbone Carbonyl group and backbone amino group of the  $i+4^{\text{th}}$  residues. However, residues at termini of the  $\alpha$ -helix have one naked side. Therefore, side chains of these terminal residues provide stability by involving themselves in non-covalent interactions with proximal residues. The C-terminus of  $\alpha$ -helix has free carbonyl group and N-terminus has free amino group.



**Figure 5.** Histogram showing frequency of appearance of S...H-N and S...O interaction at N and C-terminus  $\alpha$ -helix capping.

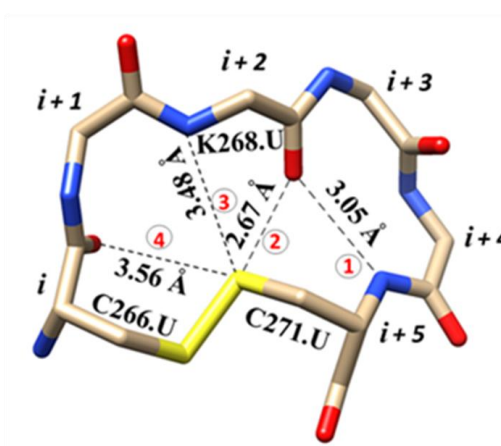
**Table 4.** A summary of PDB analyses performed to investigate role of H-bond and Ch-bond interaction  $\alpha$ -helix capping in proteins.

Fragment	Total Number of $\alpha$ -helix capping contacts ( $N_T$ ) <sup>[a]</sup>		N-terminal $\alpha$ -helix capping Contacts ( $N_N$ ) <sup>[b]</sup>		C-terminal $\alpha$ -helix capping Contacts ( $N_C$ ) <sup>[c]</sup>	
	set 1	set 2	set 1	set 2	set 1	set 2
C-S-S	107	167	26	40	81	127
C-S-C	186	316	99	152	87	164
C-S-M	516	930	516	930	0	0

As free Carbonyl Oxygen (O) from the C-terminal is a nucleophile whereas N from free amino group is, an electrophile of the  $\alpha$ -helix so we sought for divalent S mediated interaction at termini of  $\alpha$ -helix. For this purpose, we sought for these contacts in PDB and results are shown in Figure 5 and Table 4. Here, S from metal chelated cysteine (M-S-C), methionine (C-S-C) and cystine (C-S-S) were considered for analyses. We found that each of these three fragments (M-S-C, C-S-S and C-S-C) containing divalent S had varying preferences for N-terminus and C-terminus capping of  $\alpha$ -helices (Refer Figure 5 and Table 4). S from methionine has almost equal preference for capping of both termini of  $\alpha$ -helices. The C-terminus capping was through Ch-bond whereas N-terminus capping through H-bond. Interestingly, S from cystine had a very high preference for capping of C-terminus as compared to N-terminus of  $\alpha$ -helix. However, metal chelated S from cysteine preferentially involved in capping of N-terminus of  $\alpha$ -helix (Figure 5). These results clearly demonstrated that nature has employed Ch-bond and H-bonds to modulated secondary structures of the proteins.

### Role of S in stabilizing $\beta$ -turns

A  $\beta$ -turn is a region of the protein involving four consecutive residues where the polypeptide chain folds back on itself by nearly  $180^\circ$  (Jane S. Richardson, 1981). Four residues in  $\beta$ -turn are defined by their position as  $i^{\text{th}}$ ,  $i+1^{\text{st}}$ ,  $i+2^{\text{nd}}$  and  $i+3^{\text{rd}}$  (Arbor, 1968). Backbone carbonyl group of  $i^{\text{th}}$  residue form H-bond with amino group of the residue  $i+3$ . Four residues in  $\beta$ -turn have well defined geometry i.e. their  $\phi$ - $\psi$  values occupy specific region of the Ramachandran plot.



**Figure 6.** A representative example of CXXXXC motif in case 1 (PDB ID: 2FD6)

**Table 5.** Distribution of CXXXXC motif for various cases.

Criteria	Number of contacts	
	Set1	Set 2
Case 1 Chalcogen bond + H-bond (Type II turn)	46	66
Case 2 (Type I and type II' turns)	05	07
Case 3 Only H-bond (Type I turn)	19	42
Case 4 Neither Chalcogen bond nor bond	97	171

While inspecting role of S mediated Ch-bond in stability of  $\beta$ -turns, we came across a CXXXXC motif in which cysteine at 1<sup>st</sup> position (referred as  $i^{\text{th}}$  residue) and at 5<sup>th</sup> position (referred as  $i+5^{\text{th}}$  residue) are cross-linked through disulfide bond (refer Figure 6). Based on their conformation the CXXXXC motif was classified into three cases (Vishal Adhav, unpublished data). In Case 1 CXXXXC motif has H-bond between  $i+2^{\text{nd}}$  to  $i+5^{\text{th}}$  residues forming a type II  $\beta$ -turn, in addition to this, main-chain carbonyl O of  $i+2^{\text{nd}}$  residue formed a Ch-bond with S of the cysteine at  $i+5^{\text{th}}$  position (Table 5). Case 2 have a type I instead of type II  $\beta$ -turn with a Ch-bond coexisting with the H-bond that stabilizes the turn. In this case, the type I turn is immediately preceded by a type II' turn (Table 5). This Case 2 CXXXXC motif has a unique conformation in which residues from  $i+1^{\text{st}}$  to  $i+3^{\text{rd}}$  form a type I  $\beta$ -turn followed by a type II' turn between residues from  $i+2^{\text{nd}}$  to  $i+5^{\text{th}}$ . In addition to two turns a Ch-bond is also formed between backbone carbonyl O of  $i+2^{\text{nd}}$  residue and disulfide bonded S of the cysteine at  $i+5^{\text{th}}$  position. Whereas Case 3 CXXXXC motif does not have Ch-bond. Conformational stability of these three distinct sub-structures are being studied through computational calculations of their energies (Vishal Adhav, unpublished data).

### Role of S mediated interaction in stability of the regular secondary structures:

After investigation of the role of S in secondary structures such as capping of  $\alpha$ -Helix and  $\beta$ -turns from CXXXXC motif, we were interested to know if apart from helix capping divalent S could also provide stability to the regular secondary structures through Ch-bond with the backbone O and N. From earlier investigation, we found that methionine can form Ch-bonds with backbone carbonyl O of the residues residing intra  $\alpha$ -helix and  $\beta$ -sheet region. We analyzed PDB for interactions made by S of methionine or cystine with backbone carbonyl O or amino N of regular secondary structures. Interestingly we found that a higher number of Ch-bond between divalent S and backbone carbonyl O of internal amino acids in  $\alpha$ -helices in comparison to those in  $\beta$ -strands (Table 6). S...N interactions were noted with both intra-helical and  $\beta$ -strand regions. In all these cases, the interactions with methionine were more than the interactions with cystine. As expected in  $\alpha$ -helix, backbone N-H group is engaged in H-bond with  $i-4^{\text{th}}$  carbonyl O so as expected N-H...S interaction was not observed in intra-helical regions. The role of these interactions in stabilizing protein structure is being investigated in laboratory (Vishal Adhav, unpublished data).

**Table 6:** A summary of PDB analyses performed to investigate H-bond and Ch-bond interactions in intra- $\alpha$ -helix and  $\beta$ -strand.

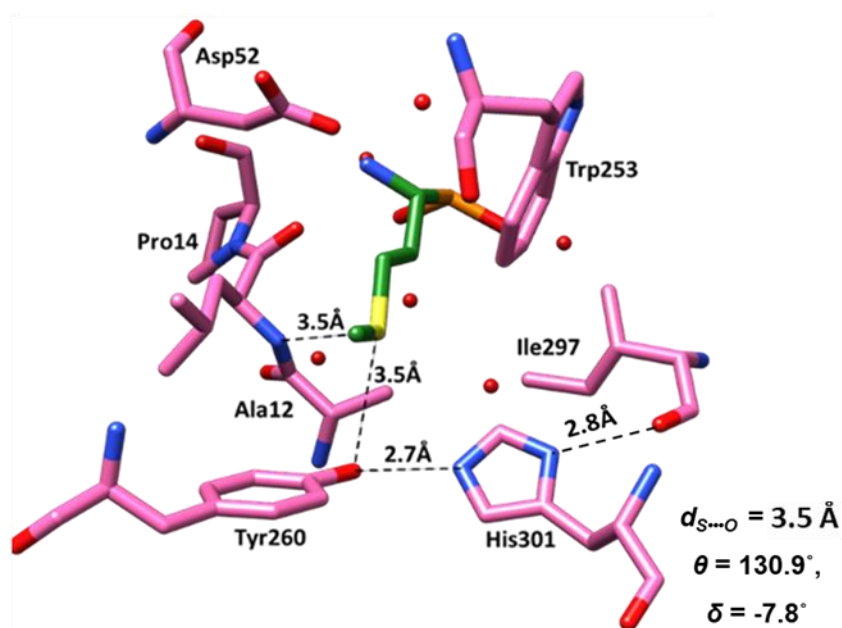
Fragment	S...H-N interaction		S...O interaction		S...N interaction	
	intra- $\alpha$ -helix	$\beta$ - strand	intra- $\alpha$ -helix	$\beta$ - strand	intra- $\alpha$ -helix	$\beta$ - strand
C-S-S (Cys)	0	13	267	93	0	10
C-S-C (Met)	0	242	104	152	39	40

### Methionyl-tRNA synthetase (MetRS) utilizes S...O interaction for recognition of methionine.

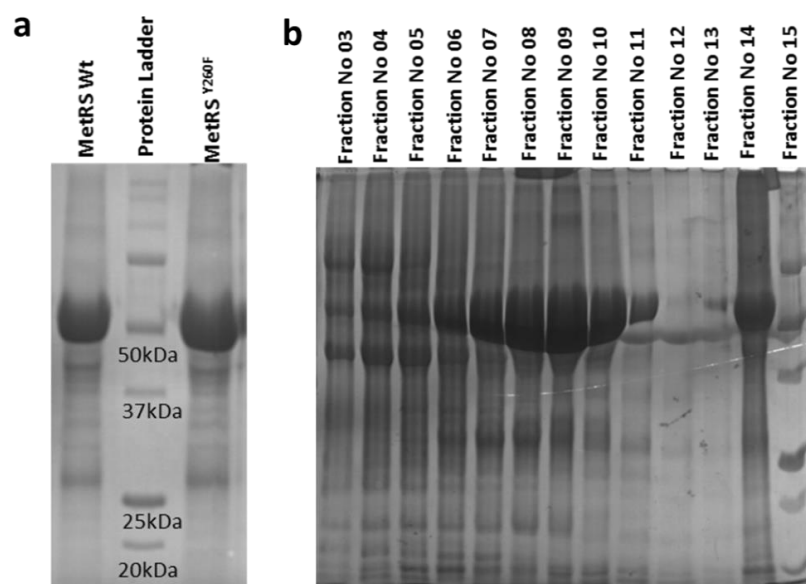
While, investigating the protein-ligand complexes, we identified presence of S...O interaction in structure of MetRS bound to methionine (PDB ID 1PFU). From structural analyses it was clear that MetRS recognizes S mainly from two



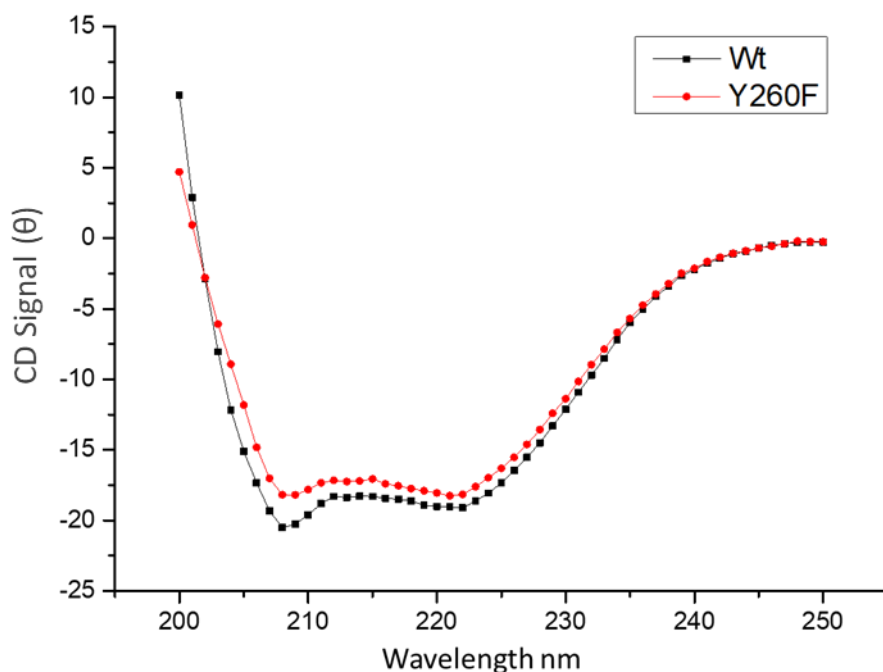
interactions such as H-bond with backbone amino group and S...O with –OH group of Try260 (Refer Figure 7). The distance between S and O was 3.5 Å along with  $\theta$  and  $\delta$  values indicated presence of S...O (refer Figure 7). In addition, from structure, it was clear that H of Tyr260 was involved in H-bond with N of His301. Based on the analyses we asked, if S...O interaction has any role in recognition of methionine.



**Figure 7.** Methionine Phosphinate (green) bound active site of Methionyl tRNA Synthase (MetRS).

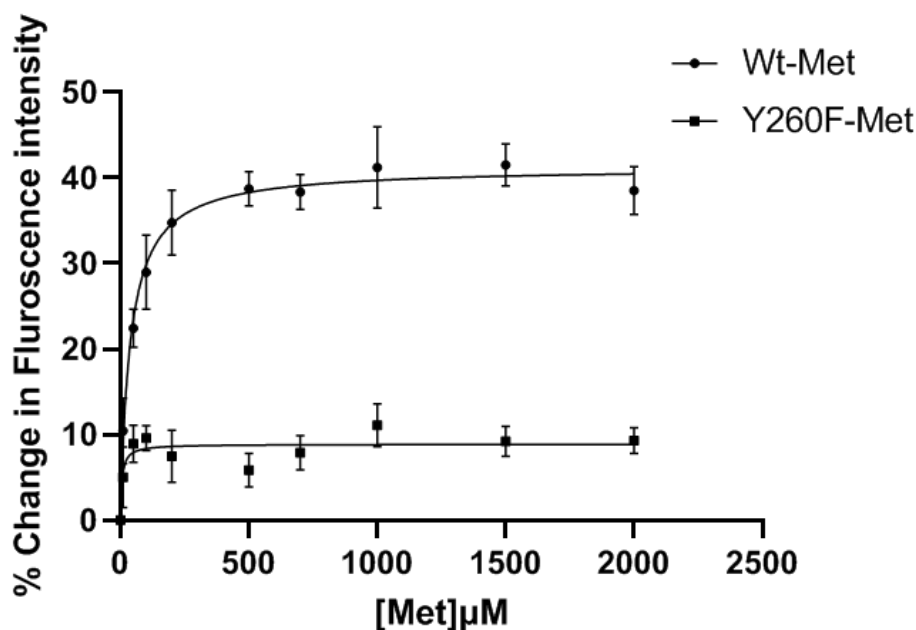


**Figure 8 (a)** SDS PAGE gel showing purity of protein purified **(b)** SDS PAGE Gel showing which collected fraction has maximum protein eluted after passing through anion ion exchange column.



**Figure 9.** Circular Dichroism (CD) spectra of MetRS, Wild type and its mutant (Y260F)

In order to understand the importance of S...O interaction we disrupted this interaction by mutating Tyr260 residue to Phe. In addition, we purified MetRS and its Y260F mutant as describe in method section. As Tyr260 residue is highly conserved (Vishal Adhav, unpublished data), we suspected the folding of the mutant. For this purpose, we recorded CD spectra of wild type MetRS and its mutant. From CD spectra, it was clear that wild MetRS has only  $\alpha$ -helices and when compared to the mutant we found that the mutation has not affected the overall secondary structure of the protein (Refer Figure 9).



**Figure 10.** Graph showing comparison in change Fluorescence intensity with respect to increasing ligand (Met) concentration for wild type and its Y260F mutant.

As reported previously, this enzyme uses an induced fit mechanism while recognizing methionine. During this process, Trp at active site undergoes a large conformation change. As a result, fluorescence intensity of bound form increases compared to native form. Hence, probing intrinsic fluorescence of this enzyme it is trivial to study binding of methionine (Mellot et al., 1989). Hence, large increase in amount of fluorescence can directly be correlated to the stronger the affinity for methionine. Previously, it is shown that upon binding of Methionine to MetRS fluorescence value increase by >30% from its initial value. We also performed similar experiment and found that intensity in case of wild type MetRS was increases by 38.5%. Interestingly, in case of MetRS<sup>Y260F</sup> mutant the fluorescence change was only 9.3% (Refer Figure 10). This observations showed that disrupting the S...O interaction resulted in decrease in affinity for methionine by four-fold. As discuss previously, S made two interaction with protein H-bond with backbone amino group and S...O bond with Tyr260. A large decrease in affinity upon disrupting S...O interaction provided a direct experimental evidence to importance this interaction in biomolecules.

## Discussion

The overall aim of the thesis was to study the properties of and physical relevance of divalent S mediated interactions in biomolecules. As discussed previously, divalent S can interact with electrophiles as well as nucleophiles. However, it is not trivial to find the circumstances to identify nature of interaction made by S. Here, we addressed this issue of the selectivity by investigating the PDB data and arrived at following conclusions (Vishal Adhav, unpublished data),

1. Metal chelated S can only interact with electrophiles.
2. In proteins, S from disulfide bridge have high potential to interact with nucleophiles.
3. S from aromatic ring, such as ligand also interact with nucleophiles preferentially compared to electrophiles.
4. In case of methionine, electrophile and nucleophile have almost equal probability to interact with S.

This rules suggested that the chemical environment of divalent S is a primary factor to decide the selectivity for electrophile and nucleophile by S. However, in case of methionine, both interactions had almost equal probability of occurrence. Here, we do not consider other factor such as chemical nature of the interacting electrophiles and nucleophiles. It is also possible that, location of methionine in protein structure can also affect selectivity. Along with this, we did not consider the presence of simultaneous interaction made by S. All these secondary factors were neglected during this study, which can also possibly answer the selectivity in case of methionine.

Apart from this, the precise role of these S mediated interactions on protein architecture is also unclear. Hence, we investigated the effect of S mediated interactions on regular and non-regular secondary structural elements. Interestingly, H-bond made by S was predominantly at N-termini of  $\alpha$ -helices while Ch-bond was found at the C-termini. Along with this, we found that the Ch-bond also involved in enhancement of the stability of  $\beta$ -turns. We found the high conservation of Ramachandran angles in presence of Ch-bond in  $\beta$ -turns. In case of regular secondary structures, we found presence of large number of Ch-bonds, which we hypothesise, increases the overall stability these elements, as these Ch-bonds play

a similar role of conventional interaction such as H bonds. This proposal is being further investigated computationally (Vishal Adhav, unpublished data).

Apart from these statistical investigations, we also performed experimental analyses to investigate physical relevance of Ch-bond in biomolecules. We selected a classic enzyme, MetRS that is crucial for protein synthesis in all form life as a model system. We applied our criteria of Ch-bond as discussed in results section to find out the presence of S...O interaction between ligand (methionine) and MetRS. From biophysical & biochemical experiment we clearly showed that disruption of this S...O interaction reduces the recognition of the methionine by four-fold. In conclusion, this study will not only help understanding how divalent S containing ligands/drugs interact with their target molecule but also to help to design new ligands/drugs or will help to modulate the binding affinity of existing ones.

# References:

- Arbor, A. (1968). Stereochemical Criteria for Polypeptides and Proteins. V. Conformation of. *Biopolymers* 6, 1425–1436.
- Craik, C. (2008). 基因的改变 NIH Public Access. *Bone* 23, 1–7.
- Deniziak, M.A., and Barciszewski, J. (2001). Methionyl-tRNA synthetase. *Acta Biochim. Pol.* 48, 337–350.
- Dill, K.A., Maccallum, J.L., and Folding, P. (2012). The Protein-Folding Problem , 50 Years On. 1042–1047.
- Hollingsworth, S.A., and Karplus, P.A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol. Concepts* 1, 271–283.
- Iwaoka, M., Takemoto, S., and Tomoda, S. (2002). Statistical and theoretical investigations on the directionality of nonbonded S...O interactions. Implications for molecular design and protein engineering. *J. Am. Chem. Soc.* 124, 10613–10620.
- Jane S. Richardson (1981). The anatomy and taxonomy of protein structure 190 203.
- Kilgore, H.R., and Raines, R.T. (2018). N→π\* Interactions Modulate the Properties of Cysteine Residues and Disulfide Bonds in Proteins. *J. Am. Chem. Soc.* 140, 17606–17611.
- Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., and Baker, D. (2012). Principles for designing ideal protein structures. *Nature* 491, 222–227.
- Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S., Konno, M., and Blanquet, S. (1999). Crystal structure of Escherichia coli methionyl-tRNA synthetase highlights species-specific features. *J. Mol. Biol.* 294, 1287–1297.
- Mellot, P., Mechulam, Y., Le Corre, D., Blanquet, S., and Fayat, G. (1989). Identification of an amino acid region supporting specific methionyl-tRNA synthetase: tRNA recognition. *J. Mol. Biol.* 208, 429–443.

- Murray, J.S., Lane, P., Clark, T., Riley, K.E., and Politzer, P. (2012).  $\sigma$ -Holes,  $\pi$ -holes and electrostatically-driven interactions. *J. Mol. Model.* *18*, 541–548.
- Nick Pace, C., Martin Scholtz, J., and Grimsley, G.R. (2014). Forces stabilizing proteins. *FEBS Lett.* *588*, 2177–2184.
- Politzer, P., Murray, J.S., Clark, T., and Resnati, G. (2017). The  $\sigma$ -hole revisited. *Phys. Chem. Chem. Phys.* *19*, 32166–32178.
- Richardson, J.S., and Richardson, D.C. (2002). Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 2754–2759.
- Richardson, D.C., J.S. (1988). Helix capping proline. *Science* (80- ). *240*, 1648–1652.
- Rosenfield, R.E., Parthasarathy, R., and Dunitz, J.D. (1977). Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur. 1. Electrophiles and Nucleophiles. *J. Am. Chem. Soc.* *99*, 4860–4862.
- Serre, L., Verdon, G., Choinowski, T., Hervouet, N., Risler, J.L., and Zelwer, C. (2001). How methionyl-tRNA synthetase creates its amino acid recognition pocket upon L-methionine binding. *J. Mol. Biol.* *306*, 863–876.
- Taylor, P., Pal, D., and Chakrabarti, P. (2012). Non-hydrogen Bond Interactions Involving the Methionine Sulfur Atom Non-hydrogen Bond Interactions Involving the Methionine Sulfur Atom †. *J. Biomol. Struct. Dyn.* *19*, 115–128.
- Wang, G., and Dunbrack, R.L. (2003). PISCES: A protein sequence culling server. *Bioinformatics* *19*, 1589–1591.