IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY PURIFICATION (IMAC) OF HIS-TAGGED CTCF-ZN (ZINC FINGERS DOMAIN) PROTEIN:

A COMPARISON BETWEEN IMINODIACETIC ACID (IDA) - BASED SEPHAROSE AND NITRILOACETIC ACID (NTA) - BASED AGAROSE

by

MAISARAH BT AB SAMAD

(105105)

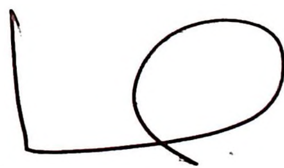
Dissertation submitted in partial fulfilment of the requirement for the degree of Bachelor of Health Sciences (Biomedicine)

May 2013

CERTIFICATE

This is to certify that the dissertation entitled "IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY PURIFICATION (IMAC) OF HIS-TAGGED CTCF-ZN (ZINC FINGERS DOMAIN) PROTEIN: A COMPARISON BETWEEN IMINODIACETIC ACID (IDA) - BASED SEPHAROSE AND NITRILOACETIC ACID (NTA) - BASED AGAROSE" is the bonafide record of research work done by Ms MAISARAH BINTI AB SAMAD during the period from September 2012 to May 2013 under my supervision.

Supervisor,



....

Associate Professor Dr. Shaharum Shamsuddin
Lecturer
School of Health Sciences
Universiti Sains Malaysia
Health Campus
16150 Kubang Kerian
Kelantan, Malaysia

Date: 15/6/13

ACKNOWLEDGEMENT

"In the name of Allah, Most Gracious, Most Merciful"

"Over the knowledgeable, Allah the Most Knowledgeable"

All praises and gratitude is to Allah, the Lord to whom every single creature in the heaven and the earth belongs to. Thank Allah for giving me the strength and patient during this challenging time. May peace and blessings be on the leader of all creation, the prophet Muhammad S.A.W, his family and companion.

I would like to express my sincere thanks and appreciation to the following people and institutions for making this study possible.

My deepest gratitude goes to Associate Professor Dr. Shaharum Shamsuddin for his advice, constructive comment and wise suggestion in supervising me throughout the duration of the study and I am indebted. Many thanks to research assistant, Madam Nurul Aini Samsuddin, post-graduate students; Tee Chee Wei and Niswathul Haania Zain Ali and other labmates for their guidance, advice and also their willing to share with me their information, experience and skills.

I would also like to thank my family and friends for their encouragement and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENT i

TABLE OF CONTENTS ii

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF ABBREVIATIONS ix

ABSTRACT x

ABSTRAK xi

CHAPTERS

1.0 INTRODUCTION 1

1.1 Problem statement 1

1.2 Objective 3

1.3 Rationale of study 3

1.4 Thesis overview 4

2.0 LITERATURE REVIEW 6

2.1 Protein Purification: Immobilized Metal Affinity Chromatography (IMAC) 6

2.1.1 Introduction to IMAC 6

2.1.2 The principle of IMAC 7

2.1.3 Divalent cation: The transition metal ion 10

2.1.4 Metal chelating agents: The multidentate ligands 11

2.1.5 Iminodiacetic acid (IDA) and Nitriloacetic acid (NTA) 13

2.1.6 The stationary phase 15

2.1.7 The mobile phase 16

2.1.8 Advantages and disadvantages of IMAC 17

2.2 CTCF Transcription factor 20

2.2.1 General overview of CTCF: A multivalent transcription factor 20

2.3 Application of IMAC: The purification of His-tagged proteins 23

2.3.1 Histidine and its outstanding features 23

2.3.2 Polyhistidine tag: An affinity tag for recombinant protein 24

3.0 MATERIAL AND METHODS 27

3.1 Recombinant DNA methods 27

3.1.1 DNA materials and vectors 27

3.1.2 Extraction of the plasmid DNA with the QIAprep Spin Miniprep Kit... 30

3.1.3 Nucleic acid concentration determination by using NanoDrop 2000/2000c Spectrophotometer 31

3.1.4 Restriction enzyme digestion of plasmid DNA 32

3.1.5 Agarose gel electrophoresis of plasmid DNA 33

3.2 Microbiological methods 34

3.2.1 Media and solutions 34

3.2.2 Bacterial strains 35

3.2.3 Preparation of competent *Escherechia coli (E. coli)* cells 36

3.2.4 Freezing and storage of *Escherechia coli (E. coli)* cells 36

3.2.5 Transformation 37

3.3 Bacterial protein expression 37
3.3.1 Small scale of recombinant protein expression (mini-preparation) 37
3.3.2 Large scale of recombinant protein expression (midi-preparation) 38
3.4 Protein analysis 39
3.4.1 Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis (SDS-PAGE) 39
3.4.2 Coomassie Brilliant Blue Staining 41
3.4.3 Western blot hybridization 41
3.4.4 Examination of Protein Concentration (BioRad) 43
3.5 Protein purification – Immobilized Metal Affinity Chromatography (IMAC) .. 43
3.5.1 Preparation of IDA-Sepharose 43
3.5.2 Column purification of His-tagged protein 44
3.4.3 Batch purification of His-tagged protein by Ni-NTA Agarose 45
4.0 RESULTS 46
4.1 Characterization of CTCF-C (Carboxyl terminal domain), CTCF-N (Amino terminal domain) and CTCF-Zn (Zinc finger domain); truncated domains of human CTCF using pET expression system 46
4.1.1 Concentration of plasmid DNA after mini-prep plasmid extraction 46
4.1.2 Double digestion of CTCF-C, CTCF-N and CTCF-Zn plasmids with restriction enzyme 47
4.2 Expression of CTCF-Zn (Zinc finger domain) of human CTCF in bacterial expression system 48

4.3 Purification of the bacterially expressed CTCF-Zn (Zinc finger domain) of human CTCF by IMAC 49

4.3.1 Iminodiacetic acid (IDA) – based IMAC purification 49

4.3.2 Nitriloacetic acid (NTA) – based IMAC purification 52

5.0 DISCUSSION **55**

6.0 CONCLUSION **64**

7.0 REFERENCES **65**

LIST OF TABLES

Table 1: The master mix recipe for double digestion 32

Table 2: Standard media and their ingredients/reagents 34

Table 3: List of bacterial strains and their features 35

Table 4: The recipe for preparation and resolving gel 39

Table 5: Stock solution for SDS-PAGE preparation 40

Table 6: Stock solution for Western blot preparation 42

Table 7: List of antibodies used 42

Table 8: Stock solution of buffers used in this study 45

Table 9: Concentration of plasmid DNA 46

LIST OF FIGURES

Figure 2.1: Structures of Nitriloacetic acid (NTA) chelating agent and transition metal Ni^{2+} forming coordination compound 8

Figure 2.2: The principal mechanism involves protein adsorption (binding) and desorption (elution) in IMAC 9

Figure 2.3: Structure of various chelating ligands in IMAC 12

Figure 2.4: The IDA ligand coordinates the Ni2+ with three valencies and another three valencies are available for interaction with the imidazole rings of histidine residues 13

Figure 2.5: NTA occupies four of the six coordination sites in the coordination sphere of Ni2+ leaving two free sites for interaction with the imidazole ring of 6X His-tag 14

Figure 2.6: CTCF is located at 67,596,310 - 67,673,088 bp on chromosome 16q22.1 locus 20

Figure 2.7: Structural features of CTCF protein showing the three domains of CTCF 22

Figure 2.8: Chemical structures of histidine and imidazole 23

Figure 3.1: The plasmid map of vector pET16bSH3 was illustrated based on plasmid map by Balakrishnan (2008) 27

Figure 3.2: The pET-16b vector is an expression vector engineered with an N-terminal His-tag sequence and followed by a Factor Xa sites and three cloning sites 28

Figure 3.3: The figure illustrates the three consecutive distinct domains of CTCF in bacterial expression system 30

Figure 4.1: Restriction enzyme Nde I and Age I digestion of CTCF truncated domains 47

Figure 4.2: Expression of human CTCF-Zn in E. coli BL 21 (DE3) with and without the induction of IPTG 48

Figure 4.3: IDA-IMAC purification CTCF-Zn protein 50

Figure 4.4: IDA-IMAC purification CTCF-Zn protein 51

Figure 4.5: NTA-IMAC purification CTCF-Zn protein 53

Figure 4.6: NTA-IMAC purification CTCF-Zn protein 54

LIST OF ABBREVIATIONS

bp	Base pair
Co^{2+}	Cobalt (II) Ion
Cu^{2+}	Copper (II) Ion
DNA	Deoxyribonucleic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
IPTG	Isopropylthio-β-galactoside
kDa	Kilo Dalton
M	Molar
mL	Mililiter
Ni^{2+}	Nickel (II) Ion
RNA	Ribonucleic acid
rpm	Revolutions per minute
UV	Ultra violet
V	Volt
x g	Relative centrifugal force
Zn^{2+}	Zinc (II) Ion

ABSTRACT

Since its introduction, Immobilized Metal Affinity Chromatography (IMAC) has massive development by devising several variants of protocols and metal affinity-based techniques, seeking the best protein separation technique to produce high quality and quantity purified protein. This study utilized two different types of metal affinity ligands with the aim of comparing the effectiveness of purification technique for His-tagged CTCF-Zn (Zinc finger domain) protein. Iminodiacetic acid (IDA) and Nitriloacetic acid (NTA) were used to chelate the Ni^{2+} respectively forming the immobilized metal affinity ligand matrices. Different protocols were used in which column purification was applied in the IDA-based IMAC and batch purification for the latter. By modifying the imidazole concentration in the washing and elution buffers, same buffers were used for both IDA-based and NTA-based IMAC. Through the SDS-PAGE and western blot analysis, the IDA-based IMAC demonstrates higher His-tag protein recovery and yield compared to NTA-based IMAC. Furthermore, the high protein retention featured by IDA caused low protein loss in the flow through and after washing. In contrast, NTA-based IMAC resulted in low protein affinity and more His-tagged protein escaped before the elutions. The level of purity of the protein was relatively similar for both methods as the His-tagged protein purity was seen to be influenced by the structure of the protein and the imidazole concentration of the buffers.

ABSTRAK

Sejak diperkenalkan, kromatografi keafinan ketakmobilan logam telah melalui penambahbaikan yang luas dengan memperkenalkan pelbagai variasi protokol dan teknik afiniti logam dengan tujuan untuk mencari teknik pengasingan protein yang terbaik untuk menghasilkan protein tulen yang mempunyai kualiti dan kuantiti yang tinggi. Kajian ini menggunakan dua jenis ligan afiniti logam dengan tujuan untuk membandingkan keberkesanan teknik purifikasi terhadap protein CTCF-Zn (mandala jejari zink). Asid iminodiasetik (IDA) dan asid nitriloasetik (NTA) masing-masing digunakan untuk mengkelat Ni^{2+} dan membentuk matriks ligan afiniti ketakmobilan logam. Protokol yang berbeza digunakan iaitu dengan mengaplikasikan purifikasi kolum dalam kromatografi berasaskan IDA manakala purifikasi kelompok kepada kromatografi berasaskan NTA. Dengan memanipulasi kepekatan imidazol dalam penimbal pencuci dan penimbal elusi, penimbal yang sama digunakan dalam kedua-dua teknik kromatografi. Melalui analisis SDS-PAGE dan pemedapan Western, purifikasi menggunakan kromatografi berasaskan IDA menunjukkan perolehan protein His-tag yang lebih tinggi jika dibandingkan dengan kromatografi berasaskan NTA. Tambahan pula, penahanan protein dalam IDA menyebabkan kurang protein yang terdapat dalam aliran terus dan selepas pencucian. Manakala, kromatografi berasaskan IMAC menunjukkan afiniti protein yang rendah dan banyak protin His-tag hilang sebelum elusi. Tahap ketulenan protein adalah lebih kurang sama untuk kedua-dua teknik kerana ketulenan protein His-tag dilihat bergantung pada struktur protein dan juga kepekatan imidazole di dalam penimbal.

1.0 INTRODUCTION

1.1 Problem statement

Immobilized Metal Affinity Chromatography (IMAC) is a well-known protein purification technique that has been widely used for characterization of natural, recombinant and modified proteins, including surface topography of histidine residues and analysis of protein-metal ion interaction (Cheung et al., 2012). IMAC is applicable under a wide range of conditions and has the ability to incorporate in a significant number of purification procedures (Chaga, 2001). Versatility of IMAC has become a significant feature which permits this technique to become standard to begin protein purification study. IMAC has been extensively studied especially in describing the use of this principle including the variety of the mobile and stationary phases that can be applied in this chromatography technique.

Diverse chemicals such as chaotropics salts, organic solvents, detergents and reducing agent which are compatible to IMAC matrices has brought a significant complexity for the ordinary user to select the appropriate method to facilitate the specific needs of the purification of protein interest. Sulkowski (1989) quoted that "there is a need to develop chromatographic protocols providing for the resolution of proteins displaying multiple histidine residues on their molecular surfaces". Nevertheless, it is hard to propose simple generic protocols for IMAC due to several factors stated by Chaga (2001). Parameters such as versatility of the chelating ligands, variability of the ligand densities, metal ion type, ability of adsorbent to recharge, sample requirements, mode of operation, development of chromatographic process, metal ion leakage and regeneration of adsorbents become the challenges that influence the development of generic protocols of IMAC.

Since being introduced by Porath et al. in 1978, IMAC has well developed in various parameters including the chelating ligand used in the IMAC. IMAC was initiated by utilizing the tridentate Iminodiacetic acid (IDA) to chelate the divalent cation and later been improved by Hochuli et al. (1987) with his invention, the Nitriloacetic acid (NTA), a tetradentate ligand. Besides IDA and NTA, there are many choices of chromatographic resin for IMAC with respective chelating group including tetradentate carboxymethylated aspartic acid (CM-Asp) anand pentadentades such as Tris (carboxymethyl) ethylenediamine (TED).

Selection of chelating ligand must be based on two important criteria which are the stability of the metal ion complex and the availability of free coordination sites for protein binding. Selecting the best chelating ligand and support in IMAC is important since the chelating resin influences the protein retention which later affects the protein yield and purity. Purifying His-tagged recombinant protein has always been associated with the use of Ni-NTA as the chromatographic support. However, the use of other popular chelating ligand, IDA has also become an option in IMAC to determine the adsorption and desorption of His-tagged recombinant protein through IMAC. In addition, NTA-based resin (Martinez & Miranda, 2010) and IDA-based resin (Chernukhin et al., 2000 & 2007; Shamsuddin, 2002; Balakrishnan, 2008) were the common choice as chelating ligand for IMAC purification of CTCF protein, the target protein of this study.

1.2 Objective

The main aim of this study is to compare qualitatively the purification method of Immobilized Metal Affinity Chromatography (IMAC) based on different metal chelating ligand used on His-tagged protein. Two types of metal chelating ligands are utilized which are the tridentate iminodiacetic acid (IDA) and tetradentate nitriloacetic acid (NTA). The purity and the yield of the purified protein are the main criteria to be discussed to compare the two metal chelating ligands in IMAC. Protein purity and yield are the critical factors to be considered as they depend on the affinity of the protein toward the matrices.

1.3 Rationale of study

Through this study, it will help in selecting suitable protocol for IMAC protein purification method by suggesting the appropriate metal chelating ligand to be used in IMAC. This is important to further studying the recombinant proteins especially the His-tagged protein such as human CTCF-Zn (zinc finger) protein. It is crucial to produce protein with high expression and purity levels in order to pursue this study to a higher level.

Broad aspect of proteomic research such as the study of protein-protein interactions, protein modification, bioconjugation and production of antibodies require an efficient purification method which can prepare a high quality and high quantity of material for the study. IMAC is a versatile protein purification method which develops extensively especially in preparation of the adsorbents for both resin and ligand materials as well as optimizing the application protocols. This study will emphasize the

utilization of different ligand materials in IMAC that influence the performance of IMAC in recombinant protein purification.

1.4 Thesis Overview

This thesis is begun by highlighting IMAC as a versatile purification technique with diverse protocol and metal ligands that had been devised since it had been introduced. Due to its diversity, it is crucial in order to find the best IMAC purification techniques. Parameters such as adsorbents and buffers were modified to purify the respective protein with high quality and high quantity. Aiming to find the best IMAC technique to purify the His-tagged CTCF-Zn (Zinc finger domain) protein, IDA-based Sepharose and NTA-based Agarose were used in this study. The IMAC, the CTCF protein and the most extensive application of IMAC which is the purification of His-tagged protein are further discussed in the Chapter 2 (Literature Review).

The materials and methodology were described in Chapter 3 of this thesis. The methodology used in this study comprises four steps. First is the recombinant DNA method in which the recombinant plasmids cloned in *Escherichia coli* (*E. coli*) strain DH5α were extracted and screened via double digestion and agarose gel electrophoresis. For the microbiological methods, competent *E. coli* strain BL 21 (DE3) was prepared for the transformation of the extracted recombinant plasmid. Transformation of the plasmid was performed using the heat shock method. The next step is protein expression. Both mini-scale and large-scale protein expression were conducted to produce high amount of lysate. The expressed protein before and after purification were analyzed via sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), followed by Coomassie Brilliant Blue staining and Western blot assay. The final step in

this study is protein purification. The expressed protein was purified through IDA-based Sepharose and NTA-agarose with respective protocols and buffers.

In Chapter 4, the result includes the SDS-PAGE and Western blot analysis of the purified protein through both IMAC purification methods. For each IMAC method, two types of result are presented based on the different buffers in respect to the purification method conducted.

Chapter 5 covers the discussion based on the results obtained by this study. Comparison was made between IDA and NTA by looking at the aspects of protein purity and protein yield which are affected by the protein affinity towards the metal ligand. Aspects regarding the His-tagged protein and the buffers are also included in the discussion. The thesis is concluded in Chapter 6 with the significance of this study.

2.0 LITERATURE REVIEW

2.1 Protein Purification: Immobilized Metal Affinity Chromatography (IMAC)

2.1.1 Introduction to IMAC

Immobilized metal affinity chromatography (IMAC) which is also known as metal chelate chromatography, metal-ion interaction chromatography or ligand exchange chromatography is another branch of application of affinity chromatography. The concept of IMAC was first introduced by Porath et al. (1975) which focused on its use for the fractionation of protein on solid support based on their differential affinity towards immobilized metal ions.

IMAC is a separation technique which exploits the known affinity transition metal ions such as Zn^{2+}, Cu^{2+}, Ni^{2+} and Co^{2+} that bind covalently to an immobilized metal-chelating substance on a chromatographic support (resin) and serves as affinity ligands for various protein. Amino acid residues such as histidine and cysteine act as probe of the protein by making use their coordination binding and affinity for metal ions.

IMAC is one of the most popular methods for purification of proteins with natural surface-exposed histidine residues and for proteins engineered with histidine tags or histidine clusters (Gaberc & Menart, 2001). As this method has gained broad popularity in recent years, it has been developed as a powerful protein prefractionation technology used in proteomics (Sun, Chiu & He, 2005) especially when rapid purification and substantial purity of the product are needed (Gaberc & Menart, 2001). Thus, research is conducted continuously especially to improve the sensitivity and

resolution of the protein separation. The development of affinity chromatography gives possibilities to the researchers to explore fields such as protein-protein interaction, post translational modification and protein degradation.

2.1.2 The Principle of IMAC

IMAC is a protein purification technique that utilizes the differences in protein-metal ion affinity and the reversible interactions between the protein to be purified and the affinity ligand coupled to chromatographic matrix (Magdeldin & Moser, 2012). The underlying principle of the binding of protein or peptides to metal ions in IMAC arises from the interaction between an electron donating group present on a protein surface and chelated transition metal ions which present one or more accessible coordination steps (Hemdan et al., 1988; Ueda et al., 2003).

The affinity ligand in IMAC is coordinated with a metal ion. The transition metal ions are electron pair acceptors and can be considered as Lewis acids. In order to coordinate with metal ions, the electron donor atoms such as nitrogen (N), sulphur (S) or oxygen (O) are attached to the chelating compounds. The interaction between the metal ion and the multidentate chelators forms metal chelates. The metal ions in the complexes must have free coordination site for subsequent interaction with the biomolecules such as protein. The free sites are usually occupied by water molecules. Adsorption of electron donor groups from the protein or electron donor atom in the side chains of the amino acids towards the chelated metal ion would replace water molecules to occupy the coordination sites (Gaberc & Menart, 2001). Coordination binding between metal ion and chelating compound is illustrated in Figure 2.1.

Consequently, the differential affinity of proteins for immobilized metal ions is derived from the coordination bonds between metal ions as electron acceptor and the amino acid side chains exposed on the surface of wild protein molecules or in recombinant protein. The reversible interaction between the protein to be purified and the affinity ligand coupled to chromatographic matrix enables adsorption and desorption of bounded proteins using mild conditions (G.S. Chaga, 2001).

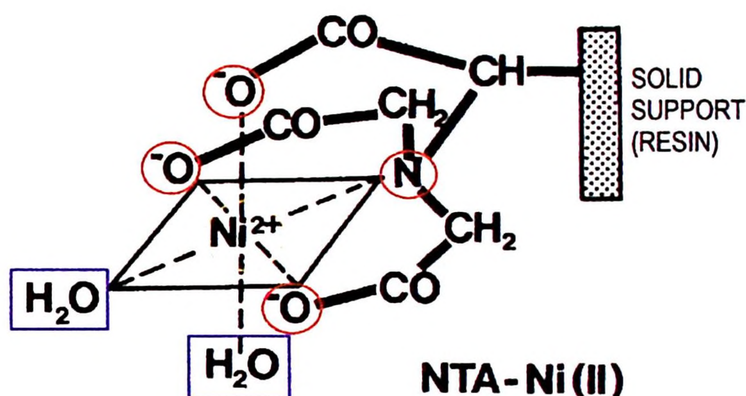


Figure 2.1: Structures of Nitrilotriacetic acid (NTA) chelating agent and transiton metal Ni^{2+} forming coordination compound. (Gaberc & Menart, 2001)

Like typical affinity purification, IMAC involves several steps as depicted in Figure 2.2. First, the sample is applied under conditions that favour maximum binding with the affinity ligand. Then, washing step is conducted to remove unbound substances, leaving the targeted molecule attached to the affinity support. Desorption is done to release and elute the bound protein either by specific or non-specifically means. In specific elution step, the bound protein can be released by using displacer or competitive ligand such as imidazole while the latter alters the media atmosphere such as the ionic strength, pH or polarity (Magdeldin & Moser, 2012).

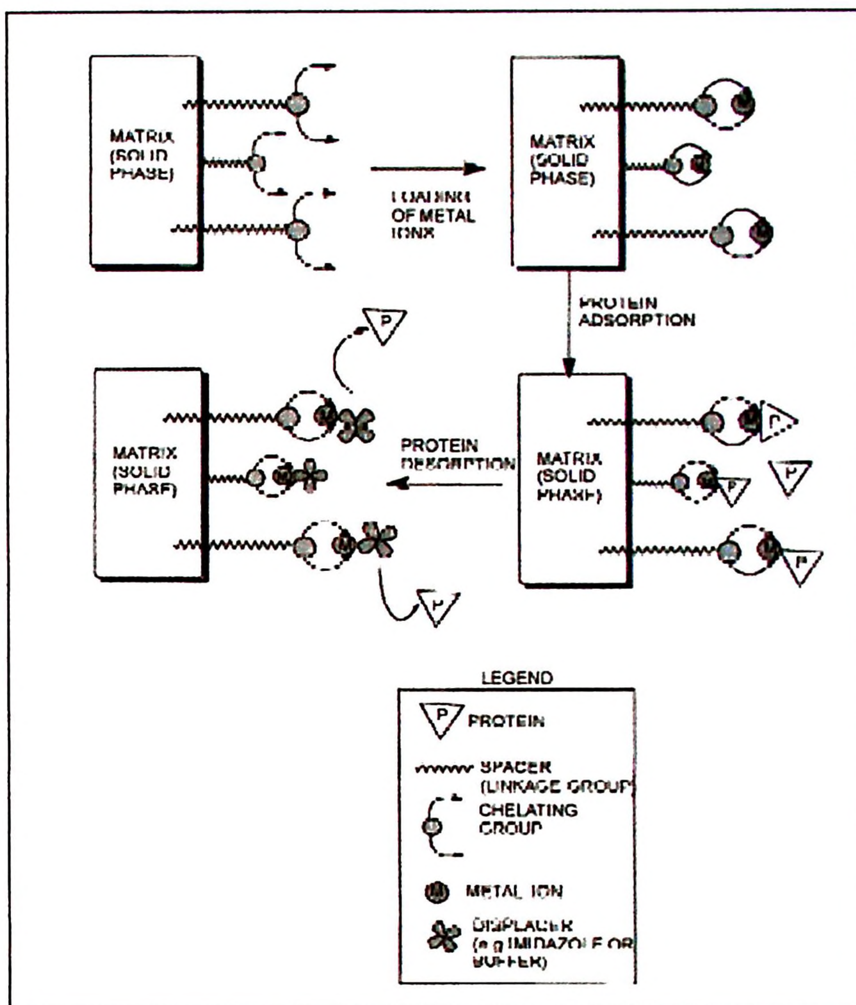


Figure 2.2: The principal mechanism involves protein adsorption (binding) and desorption (elution) in IMAC (Ueda et al, 2003)

IMAC is composed of metal ions and chelating groups forming metal chelates and chromatographic solid support (resin) as stationary phase while the mobile phase is made up by variation of pH, buffers and ionic strength. Various choices of the metal ligands through different selection metal ions and chelating agent affect the selectivity in protein separation.

2.1.3 Divalent cation: The transition metal ion

The chromatographic scheme of IMAC relies on the type of metal affinity ligands especially the metal ion involved in the coordination which largely crucial for the affinity of the protein for the metal chelate. Transition metal ion; Co^{2+}, Zn^{2+}, Cu^{2+} and Ni^{2+} is the most employed metal ions in IMAC. These metal ions are considered as intermediate Lewis acids based on the principles of hard and soft acids and bases (HSAB) as postulated by Pearson (1968). The intermediate metal ions coordinate nitrogen, oxygen and sulphur by providing six coordination number, electrochemical stability under chromatographic condition, borderline polarizability and redox stability (Ueda et al., 2003).

Utilization of intermediate metal ions in IMAC requires mild condition for adsorption and desorption of protein. For instance, adsorption of protein to chromatographic support is achieved at neutral or slightly basic medium which support the non-protonated form of imidazole in histidine residues. Lowering pH gradient in buffers or by ligand exchange with the competitive molecule such as imidazole and ammonium salts is more favourable for the elution compare with the use of strong chelating agents such as EDTA which can completely remove the metal ions and erase the binding capacity of the column (Gaberc & Menart, 2005).

Though having the same properties as transition metal and intermediate Lewis acids, the coordination and affinity of the metal ions for protein retention are varied. In research conducted by Sulkowski (1989), he postulated that immobilized Cu^{2+} would bind a protein with a single exposed histidine while for Ni^{2+} and Zn^{2+}, the metal ions bind a protein with the presence two exposed histidine. Thus, Cu^{2+} binds to the histidine residues more strongly with highest possible binding capacity. Ni^{2+} provides good binding efficiency to the histidine thus making it the most widely available metal ion for

purifying proteins especially the histidine-tagged proteins. Co^{2+} and Zn^{2+} have weaker affinity compared to the previous two. Sulkowski (1989) suggested that Co^{2+} binding requires the presence of at least two adjacent exposed histidines on the surface of a protein (Chaga, 2001). However, Co^{2+} and Zn^{2+} are preferable in for high purity protein purification due to less non-specific interaction. The affinity of metal ion determines the choice of the metal ion to be immobilized on the IMAC ligand and the application of the purification depending on the type of protein to be purified, the recovery yield and specificity of the protein.

2.1.4 Metal chelating agents: the multidentate ligands

The aim of any purification method is to achieve higher protein selectivity and specificity by monitoring various parameters including the metal chelating agents in IMAC. In addition to variation of metal ions, usage of different supports with various chelating ligand is important to be considered. Multidentate chelating compounds are selected as they provide the strength of the complex formed by the protein, metal ion and chelating group (Ueda et al., 2003).

The ligand is covalently attached to the chromatographic matrix support using spacers of various lengths and chemistries. The number of coordination bonds that been occupied through the chelation of metal ion to the surface immobilized chelating compound determines the type of chelators used in IMAC. Nonetheless, free coordination sites must be retained in the metal chelate structure in order to coordinate the proteins or solvent molecules. The difference in the number of free coordination sites influences the selectivities and adsorption of target protein. The commercially available matrices usually utilize ligands such as tridentate chelator iminodiacetic acid (IDA), tetradentate chelators nitrilotriacetic acid (NTA), a carboxymethylated aspartic

acid (CM-Asp) and pentadentate N,N,N'-tris(carboxymethyl) ethylenediamine (TED). The structure of the chelators is shown in Figure 2.3. IDA and NTA will be described further in the literature.

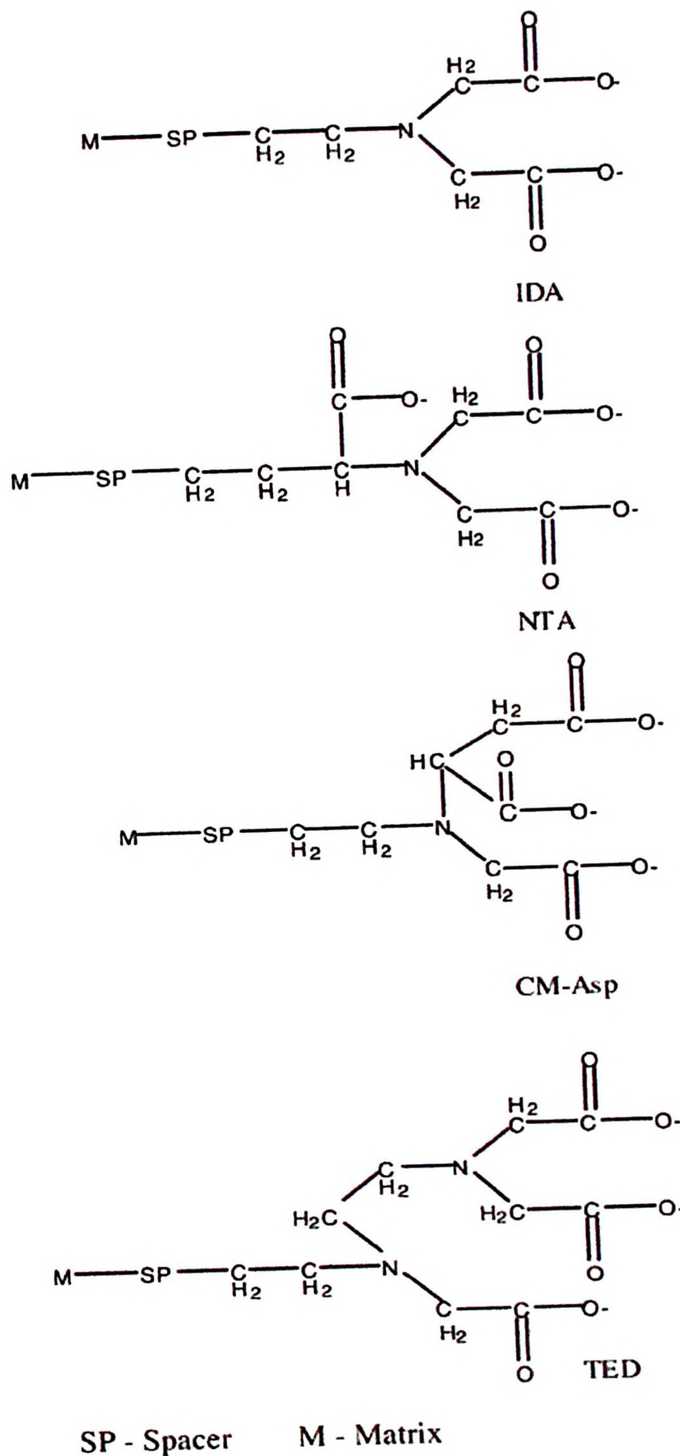


Figure 2.3: Structure of various chelating ligands in IMAC (Chaga, 2001)

2.1.5 Iminodiacetic acid (IDA) and Nitriloacetic acid (NTA)

Tridentate chelator IDA is the standard, most commonly used metal-chelating ligand for immobilization of metal ion in IMAC supports (Gaberc & Menart, 2001; Sun, Chin & He, 2006). IDA chelates transition metals through three coordination sites (Porath et al., 1975). The metal ion binds to the nitrogen atom and the two carboxylate oxygen, leaving three free coordination sites in the octahedral arrangement around a divalent metal ion for protein or solvent molecules. IDA matrices can be charged with various metal ions. Based on the retention strength, the selectivity of IMAC adsorption of divalent metal cations as chelated by IDA follows the order of $Cu^{2+} > Ni^{2+} > Zn^{2+} \geq Co^{2+}$ (Gaberc & Menart, 2001). As the mostly used chelator in IMAC, IDA matrices are available by numerous producers such as Amersham Pharmacia Biotech, Sigma and Merck. The model of interaction between protein and the metal ion in IDA is depicted in Figure 2.4.

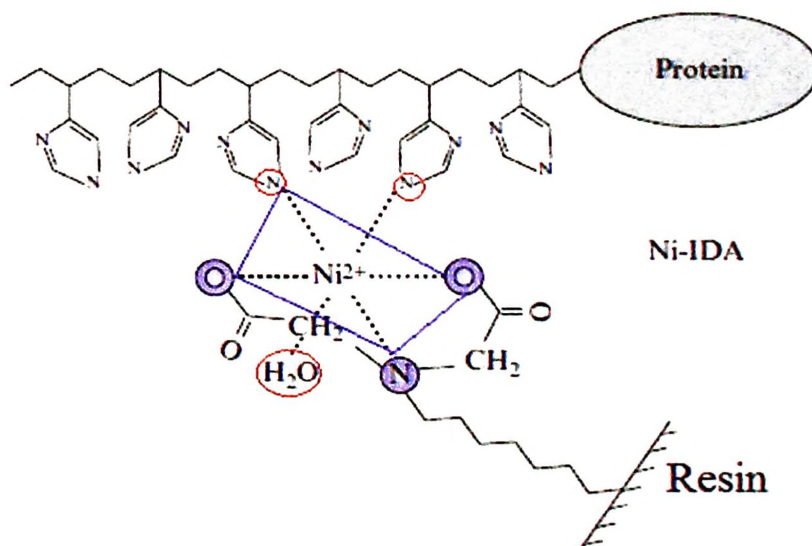


Figure 2.4: The IDA ligand coordinates the Ni^{2+} with three valencies and three free valencies are available for interaction with the imidazole rings of histidine residues. Only two valencies of Ni^{2+} are occupied by the histidine residues as what has been described by Sulkowski (1989). The figure is adapted from Block et al. (2009).

NTA is an improved chelating ligand invented by Hochuli et al. (1987). Since its invention, NTA contributes tremendously in the purification of oligohistidine extended (His-tagged) polypeptides. The tetradentate NTA binds the metal ion (Ni^{2+}) with extra carboxylate oxygen creating four valencies between the ligand and Ni^{2+}. The formation of metal chelate leaves two free coordination sites for interaction with the protein the imidazole rings of histidine residues. Unlike IDA, Ni-NTA matrix is patented by Qiagen and is usually charged with Ni^{2+}. Figure 1.6 illustrates the interaction between NTA ligand, Ni^{2+} and the histidine residues of targeted protein.

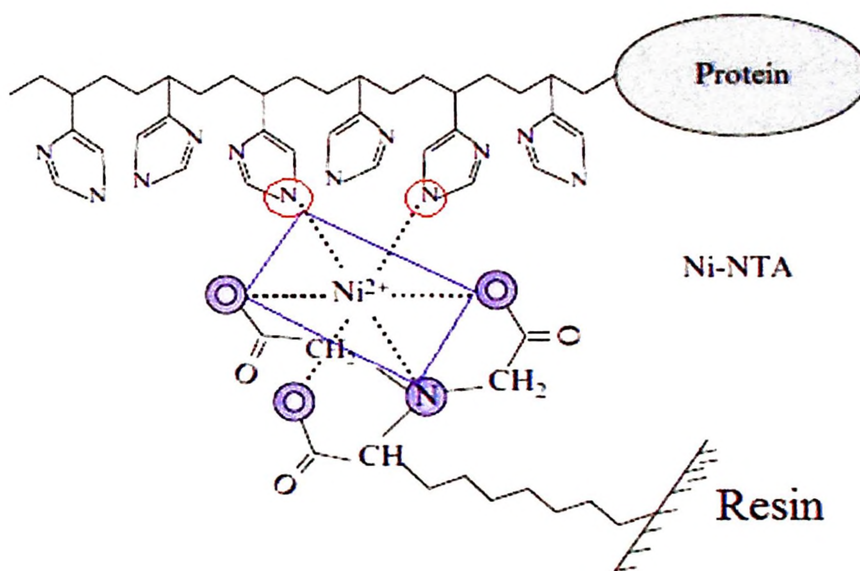


Figure 2.5: NTA occupies four of the six coordination sites in the coordination sphere of Ni2+ leaving two free sites for interaction with the imidazole ring of 6x His-tag (Hochuli et al., 1987). The figure is adapted from Block et al. (2009).

Based on the review by Chaga (2001), IDA has the highest protein adsorption which is due to the presence of more free coordination sites compared to other chelating ligands. This gives stronger protein retention power and higher binding affinities for protein (Ueda et al., 2003; Gaberc & Menart, 2005). Tetradentate NTA exhibits lower protein binding due to the loss of one coordination site, leaving two free sites for interaction with protein.

Although the IDA and NTA have differences n protein retention ability, the protein recovery is usually similar between IDA and NTA-based chromatohraphy (Block et al., 2009). However, IDA has poor metal chelating strength causing high risk of metal leaching from the ligand (Hochuli et al., 1989) and it was proved in the study conducted by Block et al., (2009). As the metal ion is bound weakly to the tridentate matrix, metal leaching from the matrix during purification affect the protein recovery resulting with low protein yields and impure products (Bornhorst & Falke, 2000).

NTA is extensively used for immobilization applications. By having superior metal chelating strength due to extra carboxylated oxygen, it is more stable than tridentate IDA and thus, the leaching of the metal in NTA is low (Block et al., 2009; Knecht et al., 2009). The same study by Block et al. (2009) also demonstrated that the purity of the His-tagged protein purified from the IDA is lower compared to NTA-based purification.

2.1.6 The stationary phase of IMAC

The affinity support is also known as matrix, resin or solid chromatographic support. Classical stationary phases include porous support materials and soft-gel matrices such as agarose, polymethacrylate, polyacrylamide, cellulose, cross-linked dextran and silica. In IMAC, the chelating agent is coupled covalently to the selected resin and being immobilized on the surface. By manipulating the chelating resin, it contributes to the alteration of protein retention and the interactions between the protein-metal chelates.

For any affinity purification, selection of a support material is based on several factors such as chemical inertness, chemical stability, mechanical stability, pore size and particle size (Magdeldin & Moser, 2012). Review by Ueda et al. (2003) had listed some physicochemical features for an ideal support for IMAC. The support must be easy to derivatize and have little or no nonspecific binding. Selection of appropriate matrix material will reduce the unspecific and residual interactions. Other than that, good physical, mechanical and chemical stability are qualities that must be displayed by the resin. The matrix should resist from degradation by all enzymes and microbes, buffers and solvents that will be used within the column.

By having high porosity, the resin provides easy ligand accessibility. The support should allow the use of high flow-rates and the column can be regenerated without degeneration of the matrix. In addition, it provides a stable gel bed with no shrinking or swelling during the chromatographic run. The critical factor that is important to be considered is the matrix should allow enough exposure of the metal ion to interact with the protein (Yip & Hutchens, 1994).

2.1.7 The mobile phase of IMAC

The mobile phase consists of wide range of conditions through alteration of pH, ionic gradients, ionic strength, use of detergents and displacer. The composition of the mobile phase affects the capacity and the selectivity of the adsorbents and plays vital role in adsorption and desorption of targeted proteins from the immobilized metal affinity ligand. Establishment of the mobile phase depends on the type of metal ion that being used; either hard metal ions or intermediate ones. With hard metal ions, the

elution can be carried out by rising pH gradients and by addition of competitors such as organic acid and phosphate to the mobile phase (Chaga, 2001).

IMAC charged with intermediate metal ions requires mild conditions for protein purification. The ionic interaction between sample and matrix and between proteins is reduced by increasing salt concentration. Salt such as sodium chloride is added into the buffers (washing and elution buffers) in concentration range within 0.1 – 1.0 M to prevent non-specific binding. Neutral or slightly basic pH values (7 or 8) and high ionic strength are applied to the buffer to favour the binding process. Adding denaturant such as urea and Guanidine HCL is important to solubilise protein under denaturing condition. Imidazole is also preferable to be added into the buffers and competes with the imidazole ring in the histidine residues to elute the protein. The concentration of the displacer is varied depending on the buffer used.

Adsorption of protein in IMAC is conducted at a given pH in which the electron donor groups on the protein surface are partially protonated (Ueda et al., 2003). Adsorbed proteins are usually eluted by decreasing the pH causing the protonation of the nitrogens of the histidine side chains and consequent disruption of the coordination bond between the amino acids and the immobilized metal ions (Chaga, 2001). Specific elution step can also be done via ligand exchange by the displacer or via extraction of the metal ion by a stronger chelator like EDTA.

2.1.8 Advantages and disadvantages of IMAC

The specificity of the interaction between the protein of interest and the ligand is well known as the main feature of affinity chromatography. IMAC exhibits some remarkable advantages as well as limitations associated with the chromatography technology and adsorption principle. Meanwhile, metal affinity applied by IMAC has

high ligand stability which permits high protein loading on the chromatography matrices. The high to medium capacity in IMAC depends on the affinity of the protein towards the ligands. Other separation technique such as bio-affinity chromatography has low ligand stability resulting with low protein loading and capacity. In addition, purification of highly expressed protein for IMAC purification presents high protein recovery with 90% recovery of the tagged protein in a single purification step (Bornhorst & Falke, 2000).

Besides that, IMAC can be applied under a wide range conditions. It requires mild conditions in most protein desorption which allows the protein to retain their biological activity. IMAC is applicable in relatively harsh conditions. Through IMAC, protein purification can be performed in biologically relevant pH range from pH 4 to 10 and strong denaturing reagents such as urea and guanidium-HCl are applicable for IMAC purification. Hence, IMAC is a stable method which can be used under denaturing conditions and in situ refolding (Gaberc & Menart, 2001). This allows insoluble protein and his-tagged protein to be purified efficiently in the presence of urea or guanindium-HCl. The versatility of IMAC permits significant number of purification procedures in the purificiation of wild type and recombinant protein.

Another advantage of IMAC is the stability of the affinity matrix resin which is unaffected by protease or nuclease activities in the extract. It can purifies crude cell lysate and can be used effectively as initial purification step. The purification by IMAC can be pursued with additional chromatographic method for further purification.

The immobilization of polyhistidine-tagged proteins on IMAC matrices is applicable for protein-protein interaction studies due to relatively high affinity of histidine tags for IMAC matrices. Proteins can be resersibly immobilized based on the

protein affinity towards IMAC. By utilizing the high affinity his-tagged proteins for matrix compared to most antibodies and antigens, it can be immobilized on the column and act as affinity purification of specific antibodies (Gaberc & Menart, 2001). In addition, IMAC provides a rapid purification and medium to high protein selectivity with low cost in comparison to other affinity chromatography methods.

IMAC is also accompanied by several disadvantages. IMAC is limited by the potential damage on the protein due to the redox (reduction and oxidation) reactions in the column. As reviewed by Ueda et al. (2003), there was a case where a protein was damaged by IMAC on a Cu^{2+}-IDA column due to oxidative proteolysis catalysed by Cu^{2+} after being activated by the reducing agents. Cu^{2+} has the potential to mediate site-specific cleavage of proteins. Several amino acids including histidine and cysteine can induce metal-catalyzed oxidation reactions which are responsible in production of reactive radical intermediates that cause damage to various proteins. In addition, sample and buffers with dissolved oxygen can provide reactive oxygen species or metal-catalyzed reactions inside the IMAC column. Contamination of the final product with oxidative radicals such as hydroperoxy compounds might be occurred after extensive elution with detergents such as Tween 20 and Triton X-100.

Metal leaching from the column is one of challenges faced by IMAC. Type of chelating compound involves in IMAC affect the strength of metal binding for the column. Tetradentate NTA ligand has superior metal binding and low metal leaching compared to tridentate IDA ligand. Metal contamination in final product is undesirable especially in therapeutic proteins. Besides that, single step IMAC purification does not assure high quality of product requested in biopharmaceutical proteins. Most common metal ions such as Ni^{2+} or Co^{2+} are considered carcinogens and Ni^{2+} is suggested to be

associated with DNA damage. Thus, additional chromatographic steps and protocols are needed to refine the contaminated protein and produce a metal free final product.

2.2 CTCF transcription factor

2.2.1 General overview of CTCF: A multivalent transcription factor

CTCF, or CCCTC binding factor, CTCF is a single polypeptide chain containing 727 amino acids which can be subdivided into three domains namely an N-terminal region, a central domain containing 11 zinc fingers and a C-terminal region (Ohlsson et al., 2001) as illustrated in Figure 2.7. Zlatanova & Caiafa (2009) described that the three distinct domains of CTCF also provide interaction platforms for various proteins, including the CTCF itself. CTCF protein exhibits 100% identity of all 11 ZFs between avian and human amino acid sequences besides being as a unique protein with no known family members (Klenova et al., 2001). According to Klenova et al., (1997), the endogenous CTCF migrates as a 130 kDa protein on SDS–PAGE. However, the open reading frame (ORF) of the CTCF cDNA encodes only a 82 kDa protein. In various cell types the CTCF protein is encoded in a 4.1 kb mRNA, with the longest ORF of the cDNA is 2184 bp. Figure 2.6 presents the genomic location of CTCF gene.

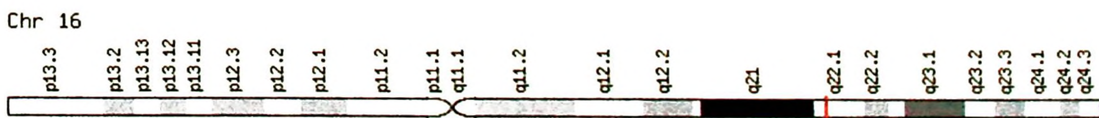


Figure 2.6: CTCF is located at 67,596,310-67,673,088 bp on chromosome 16q22.1 locus (Source by Genecards)

CTCF was identified as a transcriptional repressor that binds with three repeats of the CCCTC sequence in the human and chicken c-myc oncogene (Lobanenkov et al., 1990; Fillipova et al., 1996) and lysozyme gene (Baniahmad et al., 1990). It is described as a multivalent transcription factor (Filippova et al., 1996) which recognizes multiple target sites and regulates a number of different target genes in a variety of regulatory functions such as promoter repression and activation, hormone responsive gene-silencing and enhancer-blocking and/or boundary elements (Ohlsson et al., 2001). The evolutionarily conserved zinc finger (ZF) phosphoprotein has ability to bind through combinatorial use of its 11 ZFs to ~50 bp target sites that possess remarkable sequence variation which allows CTCF to be uniquely versatile. CTCF is also recognized to be involved in X-chromosome inactivation and gene imprinting.

The CTCF gene is cell-cycle regulated and it is highly expressed at S-G2 phase (Klenova et al., 1998). Human CTCF gene is found to be localized near the cancer-associated hot spot on chromosome 16q22.1 locus (Figure 2.6), a region commonly detected in breast and prostate cancer. Hence, CTCF gene has been suggested by Filipova et al., (1998) as a tumour suppressor gene which regulates *H19* and *Igf2* genes and promoters of *PLK* and $p19^{ARF}$ genes. These four CTCF target genes are noticed to be overexpressed and/or deregulated in cancers (Klenova et al., 2001).

As a transcriptional regulatory protein, CTCF has a variety of proteins that interact with it. The CTCF protein partners can be divided into several functional groups which are (i) the group of DNA-binding proteins including the Y-box-binding protein (YB1) (Chernukhin et al., 2000), (ii) chromatin proteins, (iii) important multifunctional proteins such as poly[ADP-ribose] and Topo II and (iv) the miscellaneous proteins (Zlatanova & Caiafa, 2009). By utilizing different binding partners in different biological perspectives, CTCF is able to perform numerous

functions either a multiplicity of functions or fine-tuning of function through directly or indirectly interactions.

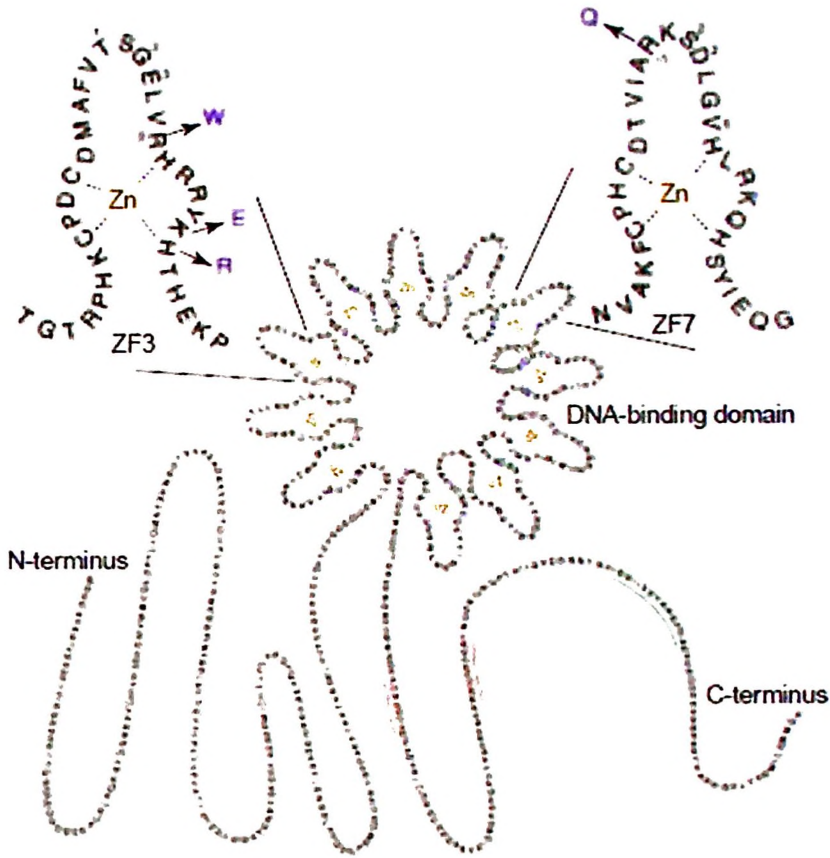


Figure 2.7: Structural features of CTCF protein showing the three domains of CTCF (Ohlsson et al., 2001)

1.3 Application of IMAC: The purification of His-tagged proteins

1.3.1 Histidine and its outstanding features

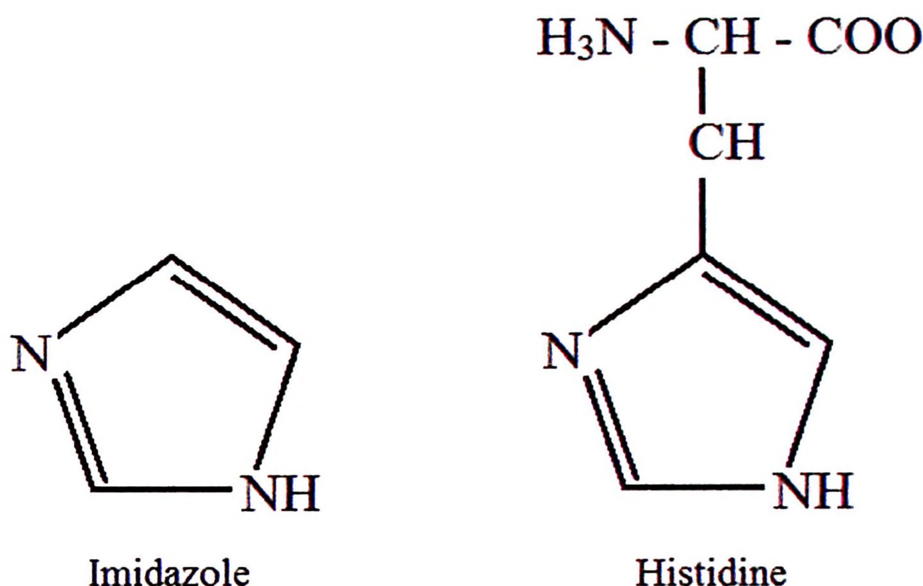


Figure 2.8: Chemical structures of histidine and imidazole

Histidine (C6H9N3O2) is an amino acid with an imidazole side chain which acts as electron donor groups and forms coordination bonds with the immobilised transition metal (Figure 2.7). Histidine exhibits the strongest interaction with the immobilised metal-ion matrices (Terpe, 2003). Early study of IMAC by Porath et. al (1975) states that strong interactions with metal ions are the most crucial aspect in protein adsorption in IMAC. Histidine is an amino acid with the strongest affinity for metal ion and has the ability to form stable complexes with Zn^{2+} and Cu^{2+} in aqueous solution. In addition, histidine possesses strong retention behaviour of protein molecule which is influenced by the spatial distribution of histidine residues over a protein surface and their accessibility (Hemdan et al., 1989; Ueda et al., 2003).

Histidine is one of the amino acids in protein in which the residues are relatively rare with a total of about 2% of amino acid content of globerular protein with only half of them are exposed on wild protein surface (Ueda et al., 2003). Purification of protein was first conducted by Porath et al. (1975) with wild proteins containing native-surface exposed histidine residues. The use of IMAC was limited with naturally occurring metal binding proteins in the beginning.

2.3.2 Polyhistidine tag: An affinity tag for recombinant protein

The method of protein purification with histidine residues was introduced by Hochuli et al., (1987), in which nitrilotriacetic acid (NTA) was used as adsorbent. In the pioneer work of Hochuli et al., (1988), a poly-His-tag was introduced to dihydrofolate reductase and the protein was successfully purified with Ni^{2+}-NTA-matrices. Today, affinity tags especially the His-tag technology has become prevailing and crucial tools for the rapid protein isolation and purification through the immobilization of recombinant proteins to the respective ligands (Terpe, 2003; Knecht et al., 2008). As antibiodies against the His-tag has been developed, the protein indentification of the recombinant protein has become possible.

Polyhistidine-tags have outstanding features in protein purification due to the high selective binding between the tag and chelated metal ion, relatively low interference of other compound during purification (Ueda et al., 2003) and high affinity which result in a high degree of purity of the target protein preparation (Block et al., 2009). As a consequence, the background of proteins that bind weakly to the resin can be easily washed away under relatively stringent conditions without affecting the binding of the tagged proteins. The high binding constant between the target protein and