Human Neil2 DNA glycosylase - crystal and inhibitor screening, and mutational analysis

Simeng Zhang

Department of Bioscience

Faculty of mathematics and Natural sciences

UNIVERSITY OF OSLO

Spring 2016

Acknowledgments

A journalist from a Norwegian science magazine once made a speech in a science conference. I remember she made a joke about the modesty of scientists. She said most scientists are so modest that always emphasize the contributions of others. They act like they expect that if they were to use 'I found…' instead of 'we found…' when talking about their work, they would be looked down upon. But I think that sometimes you should take the full credit for your work. However, this thesis would not have been possible if not for the generous help of a number of people.

First of all, I would like to express my gratitude to my supervisor, Dr. Bjørn Dalhus. He has been an amazing supervisor to me, and has helped me enormously both in the lab work and while I was writing my thesis. He encouraged me when I doubted if I should continue with the study. He is a really nice and funny person. I feel very fortunate that he has been my supervisor.

Next, I would like to sincerely thank to my colleagues at IKB. I will always remember the kindness of Pernille and Lene, who helped me greatly with purifying proteins, all the while acting like I was "the big boss". Thanks must go to Øyvind as well. He might not have realized it at all, but he encouraged me a lot during my lab work. I appreciate the time that spent with these lovely colleagues.

Also, I really need to thank my mother for supporting her "constantly studying" daughter all the time, even though she's never quite figured out what it is that I am studying. She's also helped me out tons financially, so that I've been able to focus on my studides without worrying about how to survive in this "the world most expensive country". Finally, I would also like to thank my dear André, who has remained steadfast by my side, giving me a feeling of home in this foreign land.

Abbreviations

AP site	abaic (apurinic/apyrimidinic) site
BER	base excision repair
Da	dalton
DMSO	dimethyl sulfoxide
DSB	double strand break
DSF	differential scanning fluorimetry
DTT	dithiothreitol
EDTA	ethylendiaminetetraacetic acid
HEPES	4-(2-hydroxyrthy)-1-piperazineethanesulfonic acid
HR	homologous recombination
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl-β-D-thiogalactopyanoside
MES	2-(N-morpholino)ethanesulfonic acid
MMR	mismatch repair
MST	microscale thermophoresis
Neil	Nei-like endonuclease
NER	nucleotide excision repair
NHEJ	non- homologous end joining
OD	optical density
Ni-NTA	nickel-nitrilotriacetic acid
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
ROS	reactive oxygen species
rpm	revolutions per minute
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SOC	super optimal broth with glucose
Sp	spiroiminodihydantoin
TFA	thermofluor assay
THF	(3-hyfroxytetrahydrofuranyl-2)-methyphosphate
T_m	melting temperature
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
UDG	uracil DNA glycosylase
UV	ultraviolet
WT	wild type
β-ME	β-mercaptoethanol

Abstract

Base excision repair (BER) is the major pathway for repairing oxidative DNA damages. This pathway is initiated by different DNA glycosylases. DNA glycosylases can recognise and remove DNA base lesions, and are also potential drug targets in cancer therapy. Nei-like 2 (Neil2) protein, a glycosylase in the Fpg/Nei superfamily, is a trifunctional enzyme, exhibiting glycosylase activity and AP lyase activity at both the 3' and 5' termini of AP sites. However, there is no available crystal structure for human Neil2 at present. In this thesis, different crystallisation conditions were used in trying to obtain crystals of full-length human Neil2, a truncated human Neil2 as well as Neil2-DNA complexes. In addition, different full-length human Neil2 mutants were used to study the potential active sites that are responsible for Neil2 AP lyase activity. We show that residues Pro2, Lys50 and Lys51 are all essential for Neil2 AP lyase activity. At last, 44 compounds were selected by docking to the known structure of human Neil1, which is a homolog protein of Neil2. The affinity between those compounds and the wild type full-length human Neil2 was studied experimentally. Two potential inhibitors of Neil2 were found by the AP lyase activity assay. The K_d values for the two inhibitors, ligands 20 and 34, are 13±2.5μM and 92±40μM, respectively. The IC_{50} value is 80μM for both inhibitors.

Table of contents

1. Introduction **1**
1.1 DNA damage *1*
1.2 DNA damage response and repair *3*
1.2.1 DNA damage response 3
1.2.2 DNA repair mechanism 3
1.3 BER pathway and DNA glycosylases *6*
1.3.1 Base Excision Repair 6
1.3.2 DNA glycosylases 6
1.4 Function and structure of the Neil protein family *8*
1.4.1 Introduction to Neil proteins 8
1.5 Cancer and DNA repair *11*
1.5.1 DNA repair pathway and cancer 11
1.5.2 BER pathway and cancer therapy 11
1.6 Introduction of selected experimental methods *13*
1.6.1 Protein structure determination 13
1.6.2 ThermoFluor Assay (TFA) 15
1.6.3 nano Differential Scanning Fluorimetry (DSF) 16
1.6.4 MicroScale Thermophoresis (MST) 17
1.6.5 Activity assay for inhibitor screening 18
1.7 Aim of study *19*

2. Methods **20**
2.1 Expression and purification of Neil proteins *20*
2.1.1 Transformation of plasmids of human Neil2 truncated variants 20
2.1.2 Expression of Neil proteins 20
2.1.3 Purification of Neil proteins 21
2.2 Cross-linking of Neil proteins with abasic site containing DNA *23*
2.3 Crystallisation screening of Neil proteins *24*
2.3.1 Creation of Neil - AP site analogue complexes 24
2.3.2 Crystallisation screening of Neil2 proteins and Neil2-DNA complexes 24
2.3.3 Crystallisation screening of Neil1 and cross-linked Neil1-DNA complexes 26
2.3.4 Diffraction study of Neil crystals 27
2.4. Stability study of full-length human Neil2 *28*
2.5 Screening for ligands binding *29*
2.5.1 nano DSF 29
2.5.2.MST 29
2.6. Activity assay of full-length human Neil2 *32*
2.6.1 Preparation of DNA substrate 32
2.6.2 Development of plate assay with full-length human Neil2 32
2.6.3. Screening of wild type and mutant hNeil2 33
2.6.4. Screening of five ligands with WT human Neil2 34

3. Results and discussion **35**
3.1 Expression and purification of Neil proteins *35*
3.1.1 Purification of Neil proteins 35
3.1.2 Expression of human Neil2 truncated variants 36
3.1.3 His-tag removal of human Neil2 (1-324) by TEV protease 37
3.2 Cross-linking of Neil proteins with DNA *38*
3.3 Crystallisation of Neil proteins *40*
3.4 Thermofluor shift assay for stability study of Neil2 *42*
3.5 nanoDSF for inhibitor screening of Neil2 *44*
3.7 Activity assay *52*
3.7.3 AP lyase assay of WT hNeil2 with ligands 53

4. Summary and future work **55**

1. Introduction

1.1 DNA damage

DNA, the genetic information carrier, plays a crucial role in all known living cells. The stability and proper functioning of DNA molecules are therefore essential to maintain the normal function of cells. However, DNA damage still occurs frequently. DNA damage frequently involves a chemical structure change in the DNA strand, such as a single/double strand break, a missing or an alteration of a single or multiple base. DNA damages can be classified broadly into two categories based on the damage sources: exogenous and endogenous damages. The exogenous derivation refers to environmental induced damages, while endogenous damage results from metabolic or hydrolytic internal cell processes. The exogenous DNA damages, like the UV and ionising radiation damage, or a chemically induced damage, are often linked directly to different type of cancers (Clancy, 2008). The internally factor formed DNA damages have a higher frequency compared with the environmental factors (Jackson and Loeb 2001). The endogenous DNA damages are quite frequent, but still pose a serious problem for all living cells. The endogenous DNA damages include DNA oxidation, DNA alkylation, DNA hydrolysis and hydrolytic deamination, as well as DNA mismatches caused by errors in replication (De Bont & Van Larebeke, 2004). In the following part, I will discuss a few of these types of DNA damages in more detail.

The most common type of DNA damage is oxidative DNA damages which is an inevitable damage that occurs because of metabolic processes or induced by oxidising chemicals. Oxidative DNA damages occur at a rate of 10,000 per cell per day in a human cell due to the endogenous cellular processes (Ames, Shigenaga et al. 1993). The reactive oxygen species (ROS) can be generated by normal cellular metabolism, ionizing or ultraviolet radiation (Cooke, Evans, Dizdaroglu, & Lunec, 2003), and also as a response to chronic inflammation (Mittal, Siddiqui et al. 2014). For instance, one type of the ROS, superoxide, can be formed by transferring electrons to oxygen molecules. Superoxide radicals may interact with and induce different modifications to the DNA molecules (**Figure 1.1 A**). Another typical ROS is the hydroxyl radical which is the neutral form of the hydroxide ion. The major base damage formed by a hydroxyl radical is the 8-oxoguanine as shown in **Figure 1.1 B**. These alterations of bases affect the ability of the bases to form proper hydrogen bonds with their normal partner and results in errors in base pairing. 8-oxoguanine is a good example of a base mutation due to an oxidative damage. The oxidized guanine forms a Hoogsteen base pair with adenine (Hoogsteen, 1963), leading to a G:C to T:A base pair mutation.

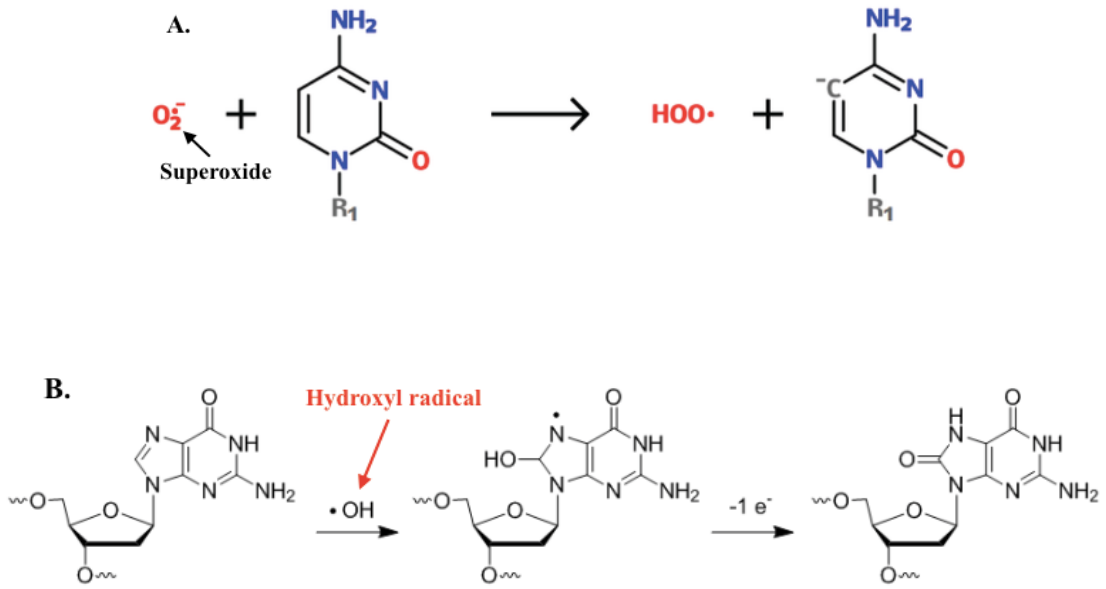


Figure 1.1 A. A superoxide molecule deprotonates the 5-carbon atom of a cytosine base. *The picture was adapted from Franka J. Rang (2014).* ***B. A hydroxyl radical reacts with a guanine molecule.*** *The final product of this reaction is the 8-oxoguanine. The reaction scheme was adapted from http://biologicalchemistry.tumblr.com (2015)*

1.2 DNA damage response and repair

1.2.1 DNA damage response

Cells have many different mechanisms to respond to both exogenous and endogenous DNA damages. There are four major response mechanisms in the cell (**Figure 1.2**). Besides the DNA repair response that have been introduced previously (Doublié, Bandaru, Bond, & Wallace, 2004), the activation of a cell cycle checkpoint prevent the damaged DNA being used as a template by replication. Alternatively, the transcriptional program activation change the transcriptional profile of the cell (Sancar, Lindsey-Boltz, Ünsal-Kaçmaz, & Linn, 2004). The last response is apoptosis which may be activated if the damage is too large to be repaired. These four responses ensure that the integrity of the DNA molecule is protected to sustain the cell stability. However, if any of these responses fail to respond to a damaged DNA, an accumulation of damages could happen and finally lead to genomic instability and cancers.

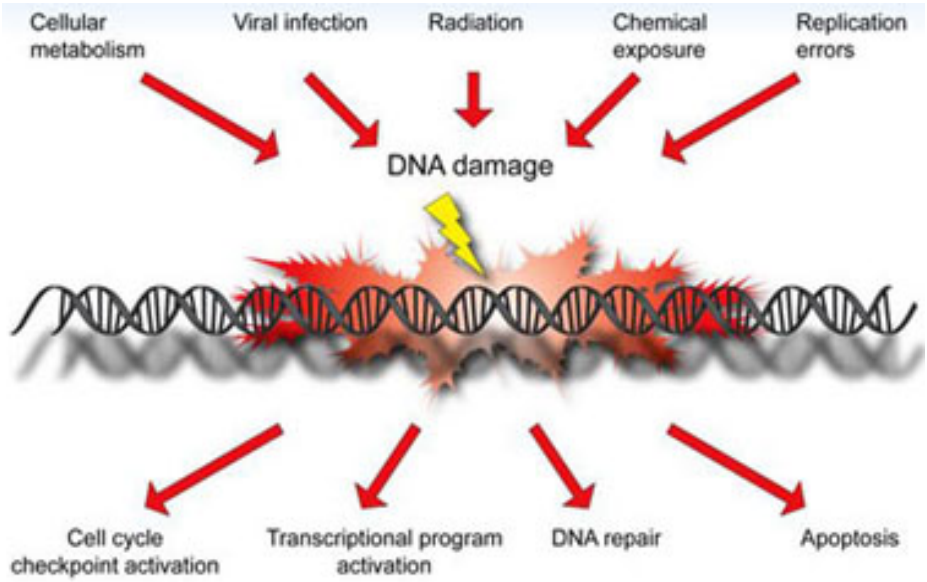


Figure 1.2 DNA damage responses. *Figure adapted from ScanTech (2014)*

1.2.2 DNA repair mechanism

DNA damages such as those described above can potentially result in adverse mutations that in turn may cause pathological conditions. Some normal cellular processes, such as DNA replication during the S-phase of the cell cycle, can also be a threat to the genetic material of the cell, because the cell's own replication apparatus is not foolproof. However, most DNA damages can be repaired by the cell itself. Cells have several repair systems that can deal with the different types of DNA damages. There are more than 130 repair genes in the human genome associated with these repair processes (Wood, Mitchell, Sgouros, & Lindahl, 2001). These repair systems can be divided into four major groups based on their mechanism of repair: direct repair, excision repairs, recombination repair and non-homologous end-joining repair (Krebs, Lewin, Goldstein, & Kilpatrick, 2013). The different DNA damages and their corresponding repair pathways are shown in **Figure 1.3**.

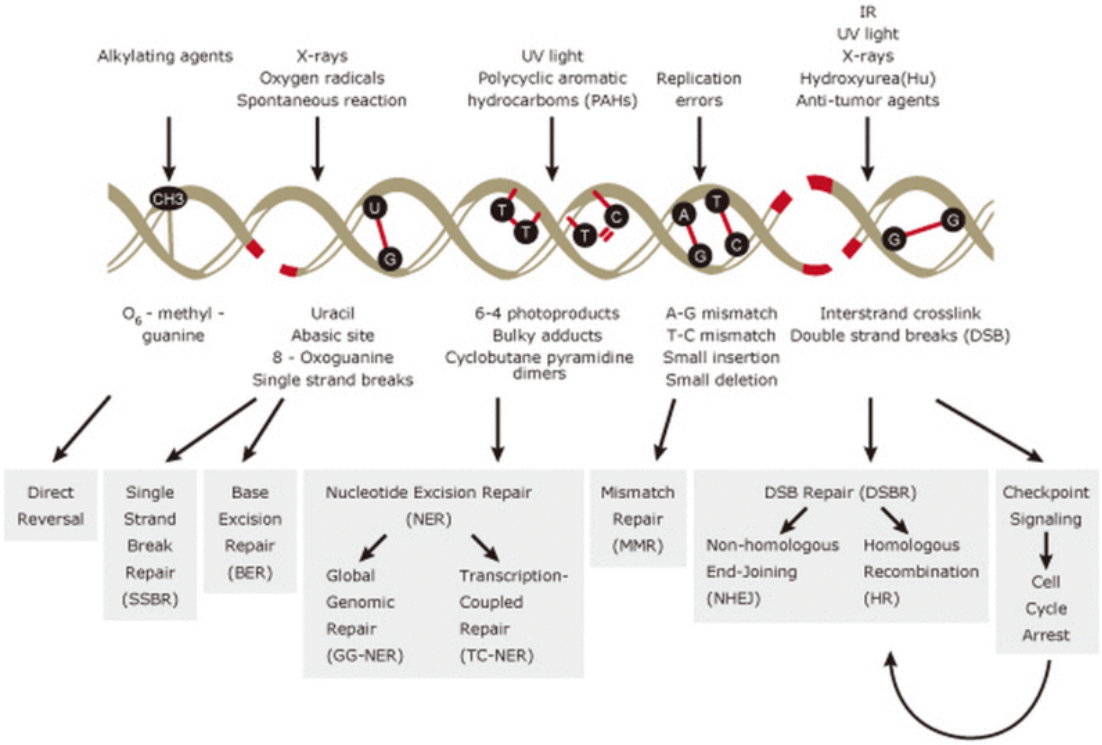


Figure 1.3 The different DNA damages and their repair pathways *(GeneTex)*

Direct repair constitutes the simplest repair pathway and includes the reversal or simple removal of the DNA damage. This mechanism does not break the phosphodiester DNA backbones and it is dependent only on a single enzyme. A good example is the photoreactivation. This process is controlled by a light-dependent enzyme, which works only in prokaryotes and plants: photolyase (Lucas-Lledó and Lynch 2009). When exposed to UV-light, successive pyrimidines and thymine pairs in particular, can form crosslinks where the two bases are covalently linked. The photolyase enzyme can reverse the abnormal covalent bond formed between the adjacent pyrimidine bases. In addition, direct repair enzymes can also reverse some methylated bases using DNA demethylases.

Excision repair pathways include base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR), all of which repair the damaged DNA by removing and replacing the erroneous bases or nucleotides. Put simply, the BER and MMR pathways repair non-bulky lesions in DNA. The former recognises and removes damaged bases and the latter corrects mismatched base pairs. In contrast, the NER pathway works on bulky DNA lesions which are usually caused by UV light or chemical carcinogens (Le May, Egly et al. 2010). All the three pathways repair single-strand DNA damage. BER will be discussed in more detail in **Section 1.3**.

Recombination repair, also known as homologues recombination repair (HR), constitutes a pathway that can repair DNA double-strand breaks, especially during DNA replication. When a replication fork encounters a lgap in a single strand, it can result in a double strand break (DSB) (Krebs et al., 2013). There are two versions of HR pathway repairing DSBs found in

bacteria (Kuzminov, 1999). In eukaryotes, there are several sub-pathways of HR. Besides the most common DSB repair, which HR is initiated by DSBs and results in a heteroduplex DNA (Szostak, Orr-Weaver, Rothstein, & Stahl, 1983), two further variations of the HR theme in synthesis-dependent strand annealing (SDSA) and break-induced replication (BIR) sub-pathways were also proposed (Li & Heyer, 2008).

Non-homologous end-joining repair (NHEJ) also works on double-strand breaks in DNA by ligating the free ends directly (Moore & Haber, 1996). When a sister chromatid or homologues chromosome is not available to use a template to repair, NHEJ will be initiated instead of HR pathway (Krebs et al., 2013). There are a few steps in this repair process. A heterodimer protein named Ku first binds to the DSB ends and the DNA-Ku complexes are juxtaposed. Then the ends and the gap are bridged by some processing enzymes. Finally the DSB ends are ligated by a specialised DNA ligase (Weterings & Chen, 2008).

1.3 BER pathway and DNA glycosylases

1.3.1 Base Excision Repair

Base excision repair (BER) is one of the DNA excision repair pathways in cells that repairs DNA lesions both from intracellular and environmental sources. The BER pathway was discovered through the identification of *E.coli* uracil DNA glycosylase (Lindahl, 1974). As the name implies, the basic mechanism of this pathway is to remove damaged bases from the DNA molecules by replacing the nucleotide and rejoin the DNA strand by multiple processes. Every step in the BER pathway is controlled by different enzymes. In general, BER is initiated by DNA glycosylases. The different glycosylases specifically recognise and remove the damaged bases on a DNA strand with a formation of abasic sites (AP site). These sites can be cleaved by an AP endonuclease (APE). Sites created by mono-functional DNA glycosylases can be cleaved by the APE directly, but sites created by multi-functional glycosylases must first be cleaved at the 3' strand of DNA by a β/δ-elimination, as shown in **Figure 1.4** (Hans E. Krokan, 2014). The steps following after the backbone cleavage can be divided into the short patch repair and long patch repair. The former is a single nucleotide replacement. It needs DNA polymerase β to remove the deoxyribose-5'-phosphate generated by the APE (J.K. Horton, 2008) and then the ligation is completed by DNA ligase I. The long patch version usually involves the replacement of 2-10 nucleotides. The DNA polymerase β/δ/ε, interacting with the proliferating cell nuclear antigen (PCNA) clamp, adds new nucleotides to the cleaved strand. Then the structure-specific flap endonuclease 1 (FEN1) removes the old nucleotides from the strand and DNA ligase finishes the repair by ligation of the gap (Krokan & Bjørås, 2013).

1.3.2 DNA glycosylases

As mentioned above, DNA glycosylases initiate the first step of the BER pathway. A glycosylase can recognise a specific damaged base and remove it from the DNA strand, leaving the deoxyribose and the phosphate parts still in the DNA backbone. The resulting site is an apurinic/apyrimidinic site (AP site or abasic site). The DNA glycosylase family can be divided into three groups based on their functions (**Figure 1.4**). For human cells, there are eleven DNA glycosylases in total. The monofunctional group includes UNG[1], SMUG1[2], MPG/AAG[3], TDG[4], MUTYH[5] and MBD4[6]. This class of enzyme has only the glycosylase activity. The bifunctional group includes NTH1[7] and OGG1[8] which both belong to the Nth glycosylase family. Besides the glycosylase activity, this class of enzyme can also cleave the DNA at the 3' position of the AP site, known as an AP lyase activity (Hegde, Hazra, & Mitra, 2008). However, even if OGG1 is classified into the bifunctional group, the dominant activity is the monofunctional glycosylase activity (Morland, Luna, Gustad, Seeberg, & Bjørås, 2005). The

[1] UNG: Uracil DNA N-glycosylase
[2] SMUG1: Single-strand selective Mono-functional Uracil DNA Glycosylase 1
[3] MPG and AAG: N-methylpurine DNA glycosylase and 3-alkyladenine DNA glycosylase
[4] TDG: Thymine-DNA glycosylase
[5] MUTYH: MutY DNA glycosylase
[6] MBD4: Methyl-CpG-binding domain protein 4
[7] NTH1: Endonuclease III-like protein 1
[8] OGG1: 8-Oxoguanine glycosylase

last group is the trifunctional glycosylases which has the glycosylase activity, plus both 3' and 5' lyase activity. Three endonuclease VIII-like enzymes (Neil1, Neil2 and Neil3) belong to this family. The functions of these three homologs will be the focus of in the next section.

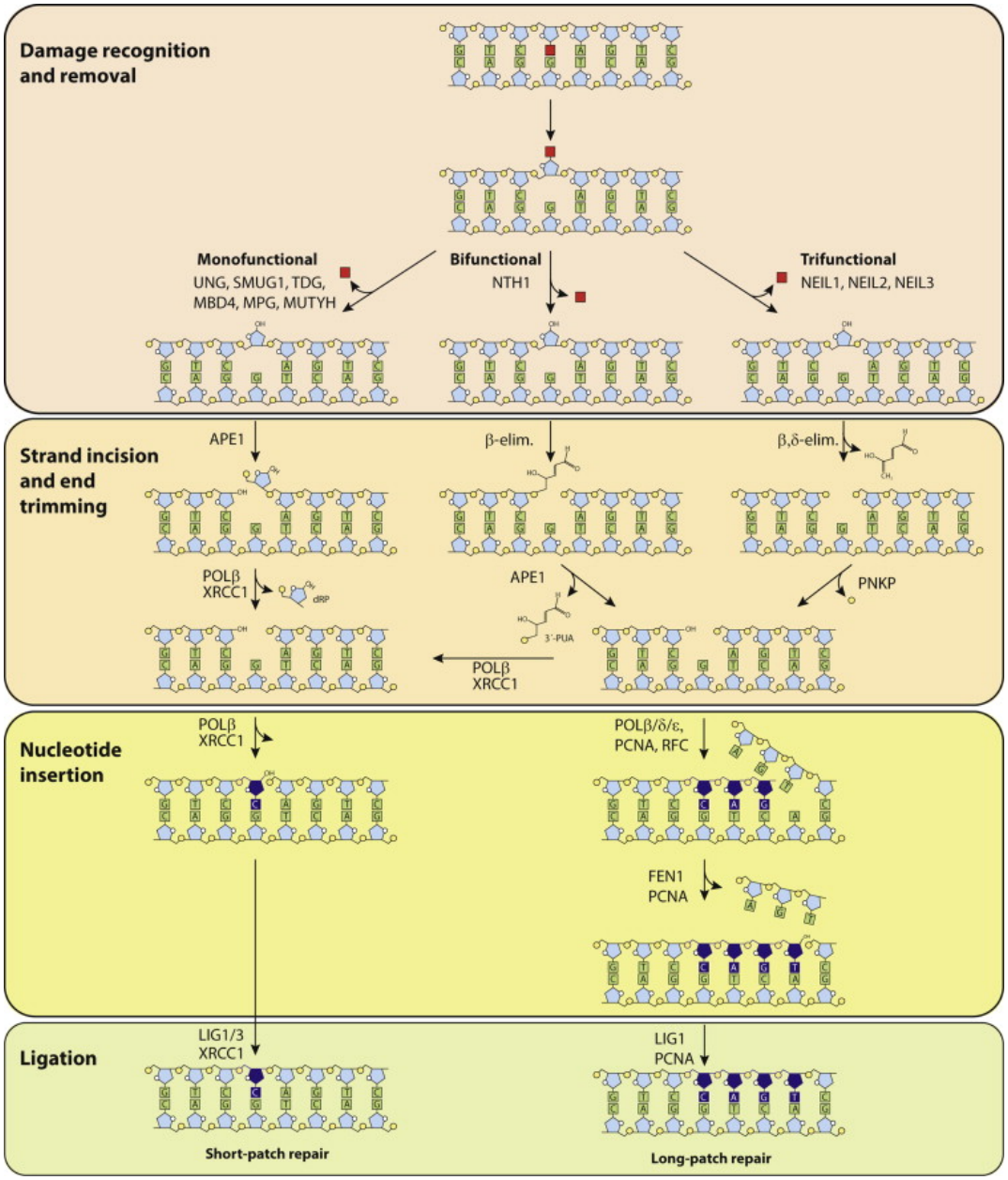


Figure 1.4 The multiple steps of DNA base repair by the BER pathway. *The first step is the damaged sites recognition and removal by DNA glycosylases. The second step is the strand incision and end trimming by AP endonuclease. The last two steps, nucleotide insertion and ligation, have two sub-pathways. One is the short patch repair (left), the other is the long patch repair (right). The two patches are controlled by different types of DNA polymerase and ligase. The red boxes in the figure refer to the damaged base and the dark blue boxes denotes the new bases (Hans E. Krokana, 2014).*

1.4 Function and structure of the Neil protein family

1.4.1 Introduction to Neil proteins

The two DNA glycosylase superfamilies, the Nth family and the Fpg/Nei family, mainly recognise oxidative DNA base lesions. The three homologs endonuclease VIII-like enzymes Neil1, Neil2 and Neil3 belong to the Fpg/Nei superfamily. These three proteins were successively identified and characterised in 2002 (Hazra et al., 2002). The sequence comparisons of the three Neil proteins with Fpg and Nei proteins are shown in **Figure 1.5**. There are some conserved structural motifs among the three Neil proteins. The most conserved motif is the helix-two-turn-helix (H2TH) motif which is involved in DNA binding. Except for Neil3, the other two Neil proteins have the PE-helix motif at N terminal which is a Pro2 followed by a Glu3. Both of the amino acids are active residues. Pro2 is the Schiff-base forming residue in the AP lyase step (Zharkov, Rieger, Iden, & Grollman, 1997) while Glu3 is a proton donor. Neil3 has a Val2 instead of Pro2, it also form a helix structure at N terminal with Glu3. Val2 in Neil3 is the Schiff-base site that is important for DNA binding (Takao et al., 2009). The zinc-finger motif is found in both Neil2 (Das et al., 2004) and Neil3. There are two zinc-finger domains in Neil3, the RanBP-type and the RanBP2 type. Neil1 only has a zincless finger domain which is also important for DNA binding (Doublié et al., 2004).

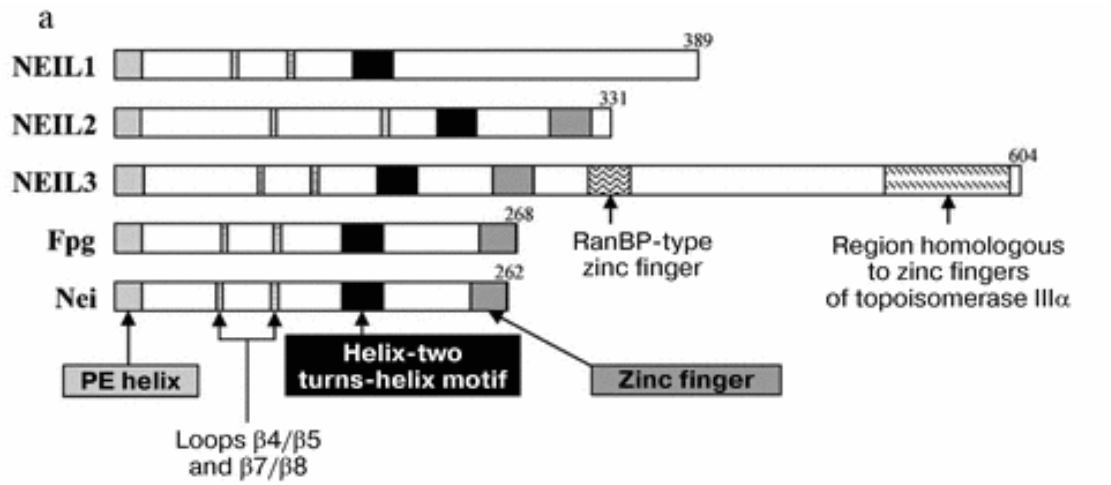


Figure 1.5 Location of characteristic structural motifs in DNA glycosylases from the Fpg/Nei superfamily. *Fpg and Nei are from E. coli, while Neil1, Neil2 and Neil3 are from human. The numbers at each end refer to the number of residues of each protein. Some similar structural motifs are also marked, like the H2TH motif and the Zinc finger motif (Grin & Zharkov, 2011).*

In general, all of the three Neil proteins have glycosylase activity but with different substrate specificities. Neil1 prefers double-strand and bubble structure DNA with 5-hydroxyuracil (5ohU) and thymine glycol (Tg). In addition, it can also process purine lesions, such as spirominodihydantoin (Sp) and guanidinohydantoin (Gh) (Prakash, Carrol et al. 2014). Neil2 prefers bubble structure than double-strand DNA. Neil1 is also active with single-strand DNA while Neil2 has a much lower activity with single-strand DNA (Dou, Mitra, & Hazra, 2003). Neil3 prefers single-strand DNA and double-strand DNA (Takao et al., 2009). Neil1 and Neil2 are trifunctional DNA glycosylases, that besides the glycosylase activity, also have the AP lyase

activity. They can cleave the DNA backbones by both β- and δ- elimination. In contrast, Neil3 is considered having mainly a monofunctional glycosylase but also with a very low activity as AP lyase that incises damaged DNA by β elimination (Krokeide et al., 2013).

As shown in **Figure 1.5**, Neil3 has a much longer sequence than Neil1 and Neil2, which sets it more apart from the other two. So in the next section, the similarity of only Neil1 and Neil2 are compared.

1.4.2 Comparison of Neil1 and Neil2

Even though Neil1 and Neil2 have a similar length, the similarity between these two enzymes are still low. **Figure 1.6** shows the sequence alignment of full-length human Neil1 and Neil2. Comparing the sequences, there are some conserved residues between the two proteins. Compared with Neil2, Neil1 has a large gap in the middle and a long C-terminal tail. Until now, only the human Neil1 structure has been determined (**Figure 1.7**).

Figure 1.6 The sequence alignment of full length human Neil1 and Neil2. *The residues highlighted with dark grey are identical between the two sequences while the ones highlighted with light grey are similar. The sequences were obtained from the UniProt database (the accession numbers for Neil1 and Neil2 are Q96FI4 and Q969S2 respectively). The alignment of the two proteins was computed by the Clustal Omega program provided by UniProt (UniProt).*

The structure of Neil1 shows the protein is composed of two domains connected by a linker. There is a wide pocket between the two domains which is for substrates or ligands binding (Doublié et al., 2004). Based on the sequence similarity, the N-terminal domain of both Neil1 and Neil2 comprises an α-helix followed by a few β-strand structure and the C terminal domain

is mostly helical, which is why Neil2 can be considered also having two domains with a cleft between them. Comparison with *E.coli* Nei and Fpg shows that the DNA binds in this cleft between the two domains. Since the structure of human Neil2 is still unknown, the similarity in sequence of the two proteins allows us to use theNeil1 structure to infer properties of Neil2.

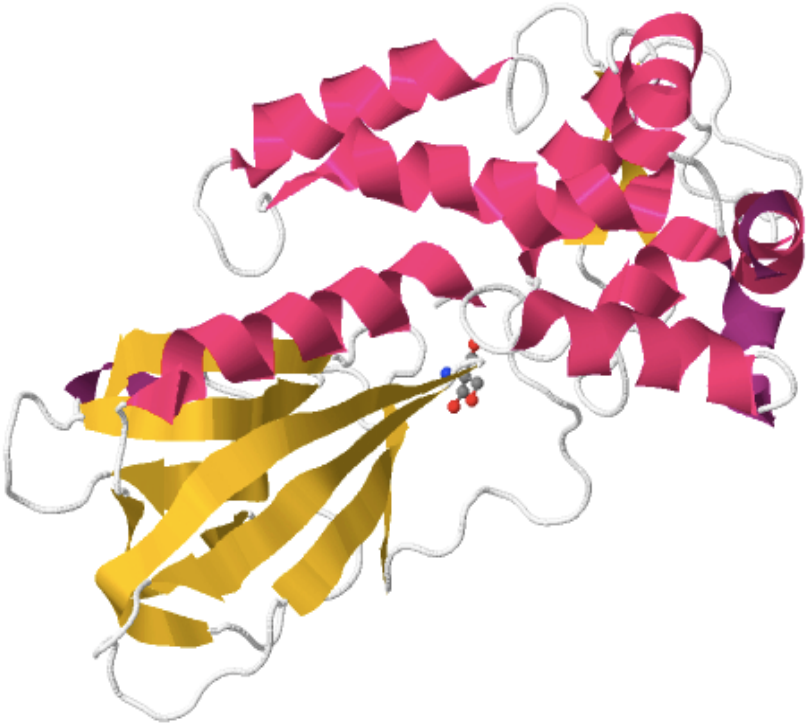


Figure 1.7 Crystal structure of human Neil1. *PDB code: 4NRV, figure adapted from Prakash, Caroll et al (2014).*

The mutagenesis study of Neil1 suggests some important residues with respect to activity: Pro2, Glu3, Lys54 and Arg277. The site mutation of Pro2 and Glu3 leads to loss of both the glycosylase and lyase activities (Bandaru, Sunkara, Wallace, & Bond, 2002). The deletion of Pro2 and the mutation of Lys54 result in the loss of glycosylase activity (Dou et al., 2003). And the mutation of Arg277 strongly reduces glycosylase activity but has little effect on lyase activity. A mutagenesis study on Neil2 shows that Lys50 is an important residue for enzyme activity (Bhakat, Hazra, & Mitra, 2004). Moreover, mutations of some residues, such as Cys291 and His295, in the zinc-finger domain of Neil2 also lead to the loss of activity (Das et al., 2004).

1.5 Cancer and DNA repair

1.5.1 DNA repair pathway and cancer

Different DNA lesions trigger different DNA repair pathways. The repair pathways are controlled by their specific proteins. Deficiencies in those pathways contribute significantly to the onset of cancers (Hoeijmakers, 2001). In addition, DNA repair pathways, as shown in **Figure 1.8**, can enable tumour cells to survive DNA damage that is induced by common cancer therapy. Therefore, inhibitions of these pathways, or to be more specific, inhibitors of specific key proteins in each pathway, may improve the drug efficiency when used in combination with DNA-damaging chemotherapeutic drugs (Tell and Wilson 2010).

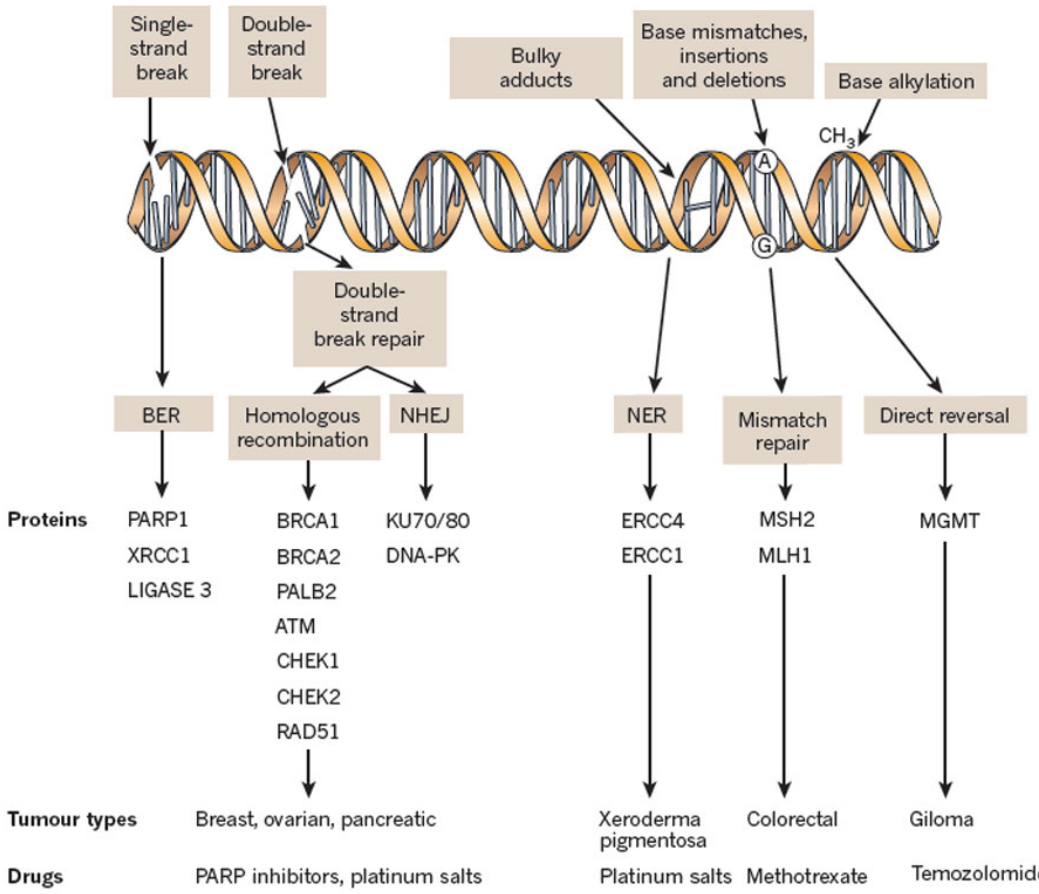


Figure 1.8 DNA repair pathways and related cancer. *The proteins in the figure refer to key proteins that involved in each DNA repair pathway. Defects in those repair pathways may characterise different types of tumour. The drugs that target these defects are shown. Figure adapted from Ashworth (2012).*

1.5.2 BER pathway and cancer therapy

Since DNA base lesions are often mutagenic and thus have the possibility to cause cancers, the BER pathway is considered as a cancer therapy target (Tell and Wilson 2010). The inhibition of BER enzymes increases the sensitivity of cells to some treatments, like ionizing radiation and chemotherapeutics, so that the cancers can be killed (Wallace, Murphy et al. 2012). This is

one of the basis for used of BER inhibitors for cancer therapy. The other concept is to use BER inhibitors to treat cancers which are due to mutations in other DNA repair pathways. If the BER can be shut down, the total DNA damage load of the cells would be high and lead to cell apoptosis (Donley et al., 2015). One example for this mechanism is synthetic lethality (**Figure 1.9**). In normal cells, if the essential genes in a repair pathway are inactive, alternative pathways with functional genes are utilised to response to DNA damage (Huhn, Bolck, & Sartori, 2013). Therefore, in those cancers caused by mutations in a DNA repair pathway, inactivation of the alternative pathway will lead to cell death.

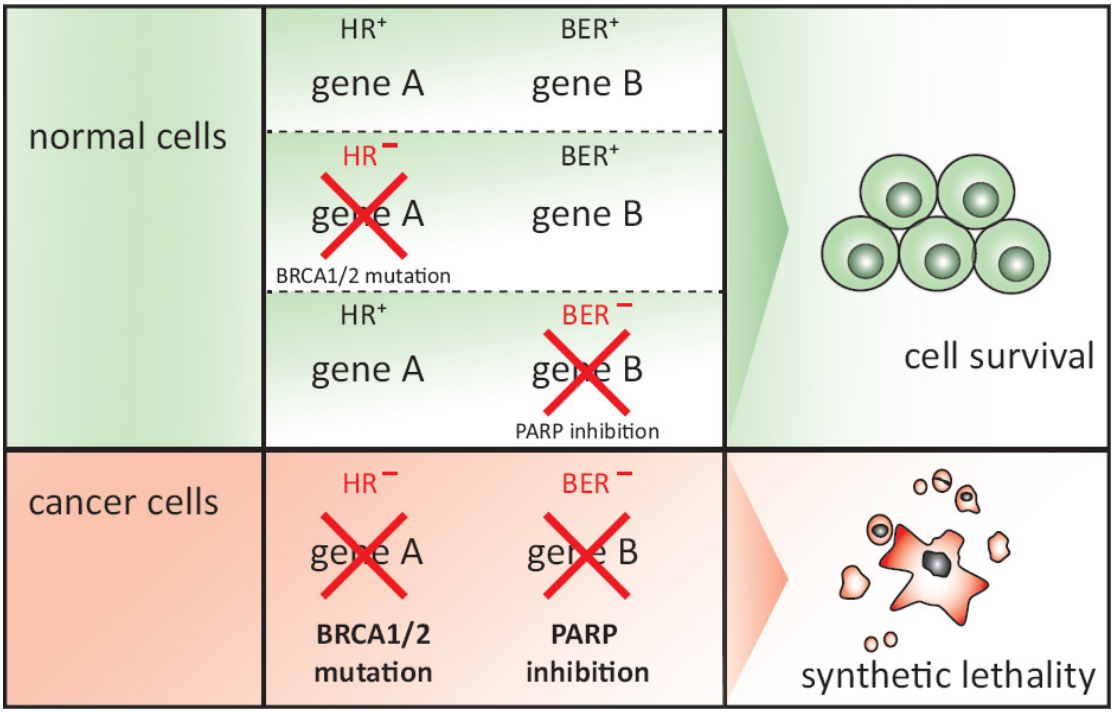


Figure 1.9 The mechanism of synthetic lethality. *Gene A and B are assumed as two essential genes in HR and BER respectively. In cancer cells, when there is mutation in HR pathway, the inhibition of BER pathway leads to cell death (Huhn et al., 2013).*

Enzymes that control BER pathway play important roles in cancer therapy. In the case of DNA glycosylases, the intermediates produced by these enzymes are sometimes more toxic than the lesion itself. For instance, the accumulation of the abasic sites from over-expression of MPG/AAG glycosylase along with down-regulation by APE1[9], Pol β[10] and XRCC1/LigIIIα[11] leads to the accumulation of single- and double-strand breaks, which may also sensitise cells to temozolomide (TMZ) (Wallace, Murphy et al 2012). TMZ is widely used in the clinic to mainly treat glioblastomas. The major TMZ-induced lesion is the N7-methylguanine which is also a substrate of BER pathway (Tentori et al., 1997). So inhibition of MPG/AAG glycosylase reduces the accumulation of toxic intermediates and also increases the efficiency of TMZ-induced cell death.

[9] APE1: AP endonuclease 1
[10] Pol β: DNA polymerase β
[11] XRCC1/LigIIIα: X-ray repair cross-complementing protein 1/ DNA ligase III

1.6 Introduction of selected experimental methods

1.6.1 Protein structure determination

There are many methods to determine a protein structure, such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and cryo-electron microscopy. Among those methods, X-ray crystallography is the most common and widely used method. It is also a method that requires protein crystals. The integral process of X-ray crystallography includes protein crystallisation, X-ray diffraction data collection, electron density map calculation, finally model building and refinement. The process is shown in **Figure 1.10**.

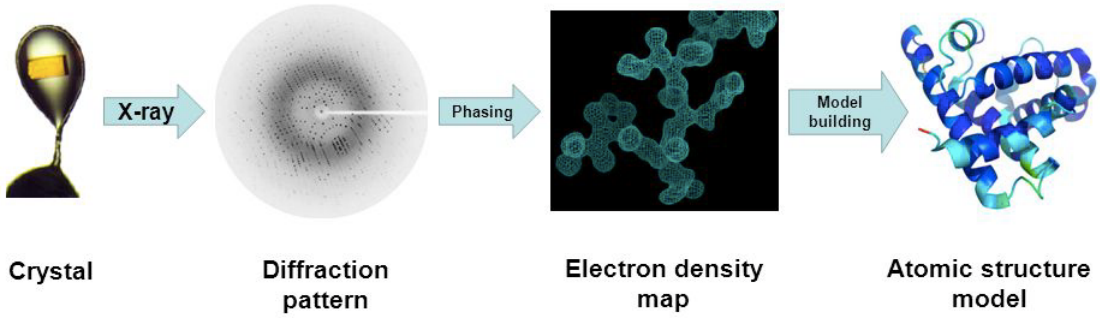


Figure 1.10 Basic process of X-ray crystallography to determine a protein structure. *Figure adapted from (Jordan, 2015).*

Highly purified proteins can form crystals under some specific conditions. This is because individual protein molecules have non-covalent interactions holding them together tightly in symmetrical organised repeats (Rhodes, 2006). Crystallisation conditions vary between different proteins, and in order to obtain protein crystals, many different conditions must be tried. Some proteins may not even form crystals, particularly, proteins with multiple domains or a large flexible region. There are many methods to screen for protein crystallisation conditions, and vapour diffusion is the most commonly used one. This is also the crystallisation method used in this study. Hanging-drop and sitting-drop are the two main ways to perform vapour diffusion crystallisation. Both of the methods uses a droplet of purified protein solution with buffer and precipitant and a relatively larger volume of reservoir solution containing the same buffer and precipitant at twice the concentration. These two parts are sealed in a micro environment (**Figure 1.11**). At first, the concentration of the buffer and precipitant in the drop is lower than in the reservoir. Slowly, because of vapour diffusion of water from the drop and into the reservoir, the concentration of the components in the droplet increases until it reaches an equilibration with the reservoir solution. The protein concentration will also increase, and under favourable circumstances, protein crystals will form. Typical precipitant agents include inorganic salts or polyethylene glycol polymers.

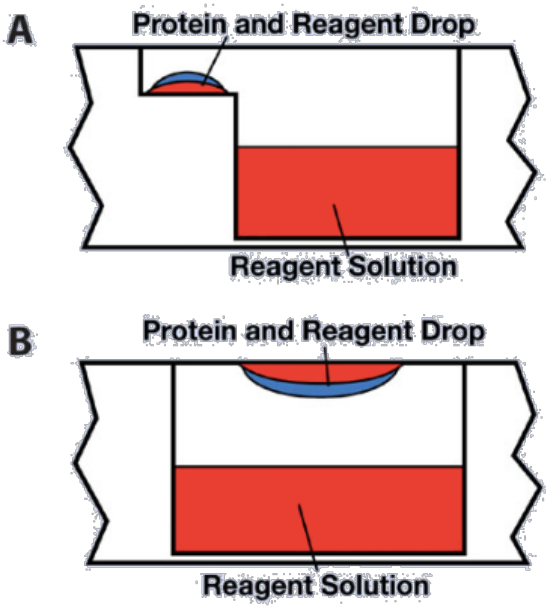


Figure 1.11 A. Sitting-drop crystallisation method B. Hanging-drop crystallisation method. *(Picture was adapted from (Vinson, 2006))*

To get the diffraction pattern of a crystal, it needs to be exposed to X-rays. The obtained crystals are usually transferred to a cryoprotectant to extend their life time in the X-ray beam using liquid nitrogen. Nowadays, an advanced machine is normally used for the crystal data collection: a synchrotron (**Figure 1.12**). A synchrotron can accelerate electrons to emit electromagnetic radiation at an extremely high energy level and that energy is at the X-ray wavelength. The X-rays can be focused into a small and intense beam focused at the protein crystal, which is typically 0.1mm in size. This high quality optical beam allows much shorter exposure times and a higher signal to noise ratio of the diffraction images than classical in-house X-ray instrument (Smyth & Martin, 2000).

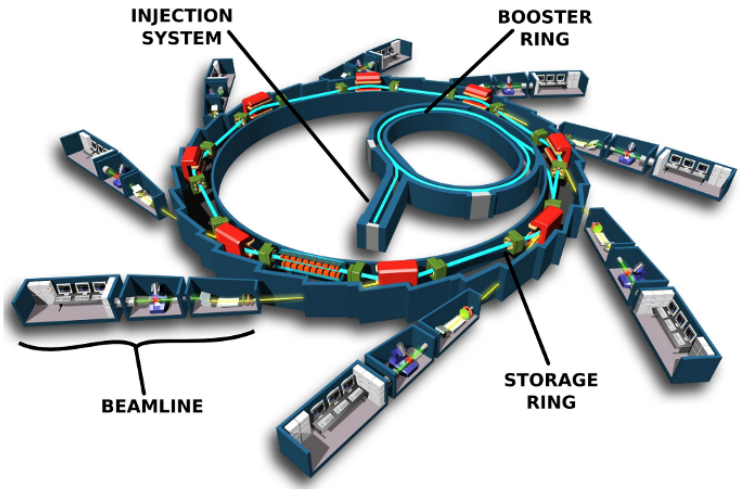


Figure 1.12 The four main components of a synchrotron facility. *(EPSIM 3D/JF Santarelli, 2005)*

In order to calculate the electron density map, two factors are needed. One is the amplitudes of the reflections while the other is the phase angles for the diffracted waves. Amplitudes can be calculated directly from the diffraction pattern, but the phase angles are lost during data collection. Phases can be calculated by way of different methods. The three mostly used methods to solve the phase problem are the multiple-wavelength anomalous dispersion (MAD), the isomorphous replacement (MIR) and the molecular replacement methods (MR) (Smyth and Martin 2000). When both the amplitudes and phases are obtained, the Fourier transform method can be used to calculate the electron density map of the protein. The final step is to improve the quality of the model by refinement. In refinement, the coordinates of atoms are adjusted to give a better fit between model and experimental data. Solvent water molecules and any ligands and ions binding to protein is also added. The best fit is obtained by least squares minimisation and monitored by a residual factor, R-factor.

1.6.2 ThermoFluor Assay (TFA)

The ThermoFluor Assay (TFA) is a dye- and thermal-based assay to study the thermal stability of proteins. It can be used to determine the melting temperature (T_m) of a protein. The principle of this assay is based on monitoring the changes in protein folding along with a temperature increase. The dye binds to the hydrophobic amino acids in a protein and the increase in fluorescence due to unfolding and exposure of hydrophobic amino acids in the protein core is measured. A typical protein melting curve is shown in **Figure 1.13**. When the temperature is low, the protein folds tightly with hydrophobic residues in the core. As the temperature increases, the protein starts to unfold and those hydrophobic residues are exposed to the dye. With the binding of dye to the protein, the fluorescence signal has a significant increase in a short temperature interval. After reaching the dye-binding saturation, the protein starts to aggregate and the dye dissociate with it, so the fluorescence signal decreases again. The melting temperature (T_m) is defined as the steepest point on the melting curve.

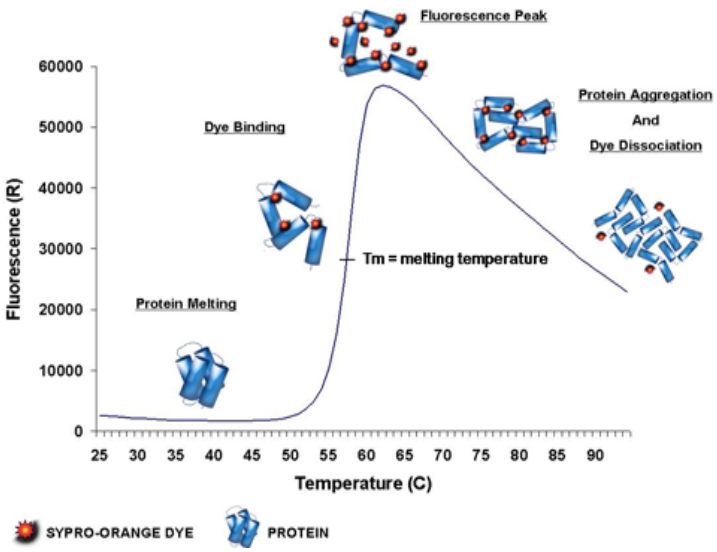


Figure 1.13 Fluorescence-based protein melting curve. *Picture from (Laboratory, 2015).*

1.6.3 nano Differential Scanning Fluorimetry (DSF)

nanoDSF is a very recently developed technology to analyse protein folding as well as its chemical and thermal stability in submicroliter volumes. It is a dye-free method that measures the changes of tryptophan and tyrosine fluorescence in a small volume capillary system. These two hydrophobic amino acids are usually located in the core of a protein. The photophysical properties of these amino acids are very sensitive to the local environment. The fluorescence signal will change as the protein is heated and changing its conformation in a temperature gradient. Hence the transition of a folded protein to its unfolded state can be monitored by detecting the fluorescence intensities of these amino acids. The nanoDSF instrument uses a dual-UV light detector to read the signal from the loaded capillaries at 330nm and 350nm. Because a ligand can bind to a protein and stabilise its structure, nanoDSF can also be used to study the affinity of a ligand for a protein. The change in fluorescence signal as a function of temperature can be analysed at one of the two emission wavelengths or the ratio of the fluorescence intensities (F350/F330) (Lea Martin 2014).

The Prometheus NT.48 nanoDSF instrument from NanoTemper, used in this study, has several strong advantages compared with normal TFA. It is easy to operate and has a fast, accurate data read out. It does not need large amount of samples and the protein concentration range is quite wide. Moreover, the protein samples can be loaded directly into capillaries profiting from its dye-free and flexible buffer approach. On the other hand, since this technique depends on an intrinsic fluorescence change of a protein, it cannot be used if a protein does not contain any tryptophan residues or if a ligand has large fluorescence.

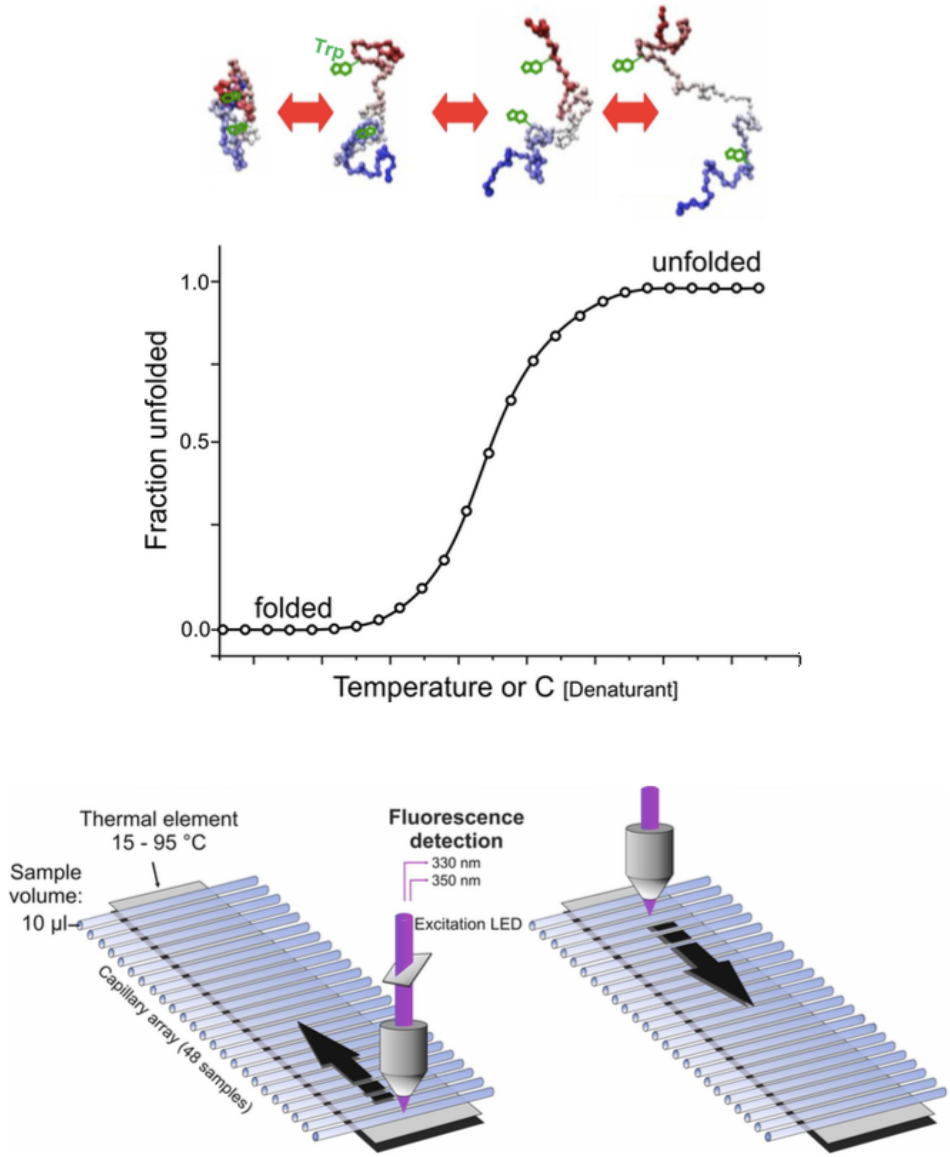


Figure 1.14 A. Principle of capillary-based nanoDSF. *The upper scheme shows the steps of a folded protein turning into an unfolded state. The tryptophanes are in green. They are packed inside of a protein in its native state and exposed to the environment when the protein is denatured. The lower plot shows the transition of a protein from folded to unfolded state as a function of temperature.* ***B. The UV-detector and precision capillary format.*** *48 samples can be measured simultaneously at both 330nm and 350nm.*

1.6.4 MicroScale Thermophoresis (MST)

MicroScale Thermophoresis (MST) is a relatively new technology for studying the interaction between biomolecules. It is based on the detection of diffusion of molecules in a temperature gradient. There are two kinds of MST instruments, the label-free and the standard. Both of them have an infrared laser (IR laser) that induce the temperature gradient and a detector that detects and quantifies the movement of fluorescent molecules. The label-free detector uses fluorescence of protein tryptophanes while the standard detector uses covalently attached fluorophores (Jerabek-Willemsen et al., 2014). The process of a single MST experiment with a single capillary is shown in **Figure 1.15**.

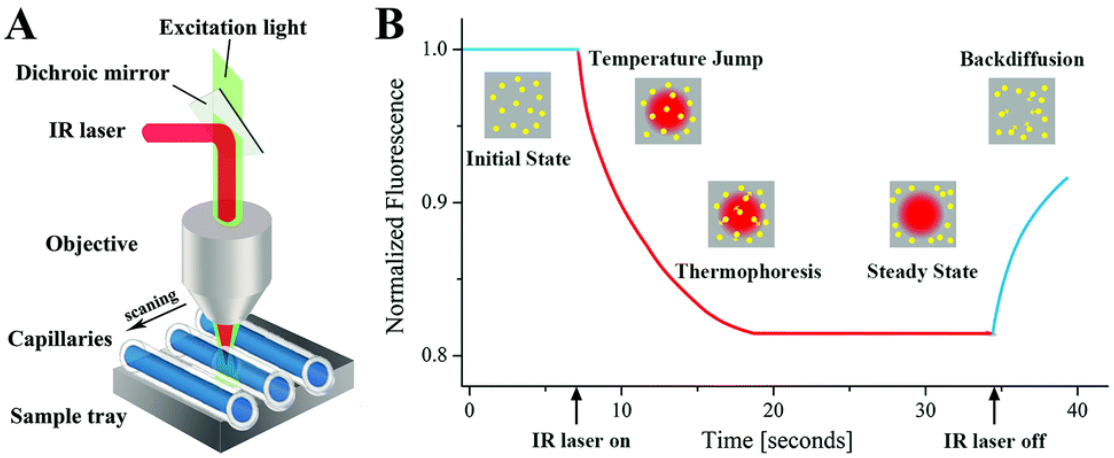


Figure 1.15 A. The IR laser and the capillary format. *The IR laser is used to generate a temperature gradient and a detector is used to measure the fluorescence change in the capillaries.* ***B. A typical MST plot of a single capillary.*** *When the IR laser is turned on, there is a quick temperature jump at first (typically 1-2 seconds) and then a thermophoresis happens as the protein move away from the heated area until the diffusion reaches a steady state. At last, when the IR laser is turn off, there is a back diffusion stage where the protein move toward the initial state (Yanan Liu, 2015).*

1.6.5 Activity assay for inhibitor screening

The activity assay for screening Neil2 inhibitors and mutants in this study is a fluorescence based, real-time kinetic method. The fluorophore is a 5'- carboxytetramethylrhodamine (TAMRA) in a DNA strand with a complementary strand containing a 3'-Black Hole Quencher 2 (BHQ2). The TAMRA fluorescence is quenched in the double strand substrate. When a scission at the TAMRA labeled strand happens, the fluorescence is no longer quenched with the release of the TAMRA containing sequence into solution. A plate reader can be used to detect the fluorescence signal in a proper time interval. The data can be output to software which can plot the real-time kinetics curve using the fluorescence readout against the time. The plate reader has two optical systems, the excitation and the emission. The first system selects a specific wavelength specific for the sample using and the second system collects the emitted light of the sample.

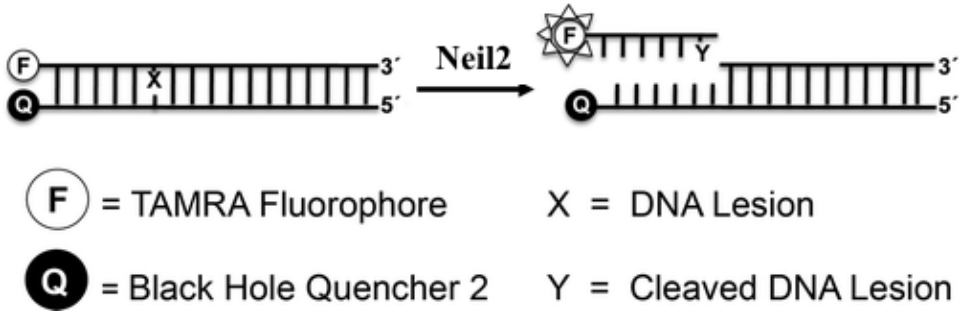


Figure 1.16 Principle of TAMRA fluorescence based activity assay. *Picture adapted from (Donley et al., 2015).*

1.7 Aim of study

The potential of human DNA glycosylase Neil proteins in the BER pathway as a possible drug target in cancer therapy was first discovered in 2002 (Bandaru et al., 2002). However, there is a very limited knowledge about their structural basis for DNA damage recognition and repair, especially for Neil2. Only a viral Neil2 structure has been determined at present (Prakash et al., 2013). Knowing the crystal structure of a protein is very important for studying its enzymatic mechanism. Moreover, the study of mutants of a protein helps to determine the composition of the active site for the protein activity. Until now, only one study has analysed potential active site residues in the zinc finger domain of Neil2 (Das et al. 2004). In addition, small molecule inhibitors of Neil proteins can be used in drug developing in cancer therapy, so there is a need to screen for compounds that bind to human Neil proteins.

The aims of the present study are summarised below:

a) Obtain the crystal structure of human Neil2. This part includes the crystallisation of free Neil2 protein and the co-crystallisation of Neil2 with DNA.

b) Characterise potential active site residues in Neil2 AP lyase activity

c) Discover small molecule inhibitors of human Neil2

2. Methods

2.1 Expression and purification of Neil proteins

In this study, hNeil1 refers to human Neil1 with 337 residues, hNeil2 refers to full-length human Neil2 with 332 residues and mNeil2 refers to full-length mouse Neil2 with 329 residues. There are three different truncated hNeil2 variants that were also studied, with 250, 290 and 324 residues, respectively. All the truncated Neil2 proteins, constructed in plasmid pET22b, were purchased from GenScript and had to be transformed into *E.coli* cells before further processing. Expression and purification procedures were the same for all proteins. For the recipes of all buffers, see Appendix.

2.1.1 Transformation of plasmids of human Neil2 truncated variants

25µL of BL21-CondonPlus (DE3) -RIPL *E.coli* strain was thawed on ice. 1µL of the respective Neil2 plasmid was added into the *E.coli* cells and incubated on ice for two minutes, followed by a heat-shock treatment in a 42°C water bath for 25 seconds and then put back on ice. The mix was incubated at 37°C for one hour. 10µL and 100µL of the transformed cell culture were applied onto LB-agar plates with 50µg/mL ampicillin and spread evenly all over the plates using glass beads. The plates were incubated at 37°C over night.

2.1.2 Expression of Neil proteins

Small-scale expression test of hNeil2 truncated variants

In order to check if the transformation of the various truncated human Neil2 variants were successful and able to express the protein, a small sample of each cell culture was tested for the expression of Neil2.

Four colonies from a LB-agar plates were inoculated into 4x100mL LB-medium containing 100µg/mL ampicillin. The cell cultures were incubated at 37°C, 180rpm overnight. 1mL of each overnight culture was transferred into a new 100mL LB-medium with ampicillin and incubated at 37°C, 180rpm until the OD600nm value reached ~0.6. 1mL of each cell culture was removed and centrifuged at 4°C, 14,000rpm for 2min. The cell pellets were stored in a -20°C freezer and were used as a negative control (subsequently denoted with "-"). 100µL of 0.25M IPTG were added to the rest of the culture to induce the expression of the protein. The cultures were incubated at 18°C, 180rpm over night. 1mL of the cultures from each bottle was centrifuged at the same condition and the cell pellets were collected in the same manner as the negative control samples.

100µL of a cell lysis buffer (refer to protein cracking buffer in Appendix) containing 15% (w/v) of β-ME and 6% (w/v) SDS was added to each cell pellet tube. All the tubes were mixed by vortexing and kept in a heat block at 95°C for 15 minutes. Then the mixtures were sonicated for 10 seconds with an amplitude of 3 with the small probe.

The cell lysate was analysed by SDS-PAGE to examine the result of Neil2 expression. The electrophoresis was performed using MOPS buffer and 12% NuPAGE gel with 200V for 40 minutes. Two different volumes of the same sample (+) and its negative control (-) were loaded onto the gel as shown below.

Table 2.1 The loading arrangement of one sample on a NuPAGE gel for examining the Neil2 expression. *The standard marker used in this study was the SeeBlue®Plus2 Prestained Standard (1x) from Invitrogen.*

Standard maker	-	+	-	+
5μL	5μL	5μL	10μL	10μL

After having verified that the transformation had succeeded and the clones were able to express the different Neil2 variants, 500μL of 60% glycerol was added to 1mL of the culture to make a glycerol stock, which was stored at -80°C for later use.

Large-scale expression of Neil protein variants

100mL pre-cultures were set up by inoculating a small volume of each Neil protein glycerol stock and incubating at 37°C with shaking overnight. 6x10mL of the pre-culture was transferred into 6x1L autoclaved LB-medium with 100μg/mL ampicillin. The six flask were incubated at 37°C, 180rpm until the OD600nm value of the cell culture reached ~0.4. After the temperature was reduced to ~25°C, 1mL of 0.25M IPTG was added into each bottle. The six bottles were incubated at 18°C, 180rpm overnight. The cell cultures were centrifuged at 4°C, 5500rpm for 30 minutes. The supernatant was discarded while the cell pellets were collected and stored at -20°C.

2.1.3 Purification of Neil proteins

Ni-NTA affinity chromatography

The cell pellets from the 6L cell culture were combined and resuspended in 90mL buffer A and the resulting mixture was vortexed until all the cell pellets were dissolved. The suspension was sonicated 3x30s on ice with 60% of maximum amplitude with the large probe, and then centrifuged at 4°C, 13000rpm for 30 minutes. The cell lysate was collected for Ni-NTA affinity chromatography.

6mL of Ni-NTA agarose (Qiagen; stored in 30% ethanol) was applied to an Econo column (Biorad). The nickel resin was washed with 2x25mL MilliQ water to remove the ethanol, and equilibrated with 2x25mL buffer A. The cell lysate was then applied to the column and passing through the resin by gravity. The flow-through was collected in 50mL Falcon tubes. The nickel resin was washed with 25mL buffer A and 3x5mL buffer B. The bound protein was eluted from the nickel resin by 3x5mL buffer C, containing 300mM imidazole. All the fractions were collected in separate Falcon tubes.

The purification result was checked by SDS-PAGE. 15μL of each fraction from the Ni-NTA step was mixed with 5μL of NuPAGE buffer (4x, Novex) and heated to 70°C for 5min. The denatured samples were loaded onto a 12% Nu-PAGE gel and ran at 200V for 40min.

Ion exchange chromatography
A Hitrap SP XL cation exchange column (GE Healthcare) was used for further purification of Neil proteins. The 5mL column was connected to the Äkta FPLC instrument (GE healthcare) and washed with 20mL MQ water, then washed with 20mL high-salt buffer (buffer D), and finally equilibrated with 20mL low-salt buffer (buffer E).

The protein eluted from the Ni-NTA column, which was in 300mM imidazole, was diluted by buffer E with a volume ratio of 1:2 and then applied to the equilibrated Hitrap column. After washing away the unbound proteins from the column with 20mL buffer E, the Neil protein was eluted by a salt gradient from 50mM to 2M NaCl. The fractions were collected every 1.5mL.

His-tag removal of truncated Neil2
The truncated Neil2 with 324 residues and a C-terminal 6x-His-tag contains a sequence between the protein and the tag that can be recognised by TEV (Tobacco Etch Virus) protease. The protein was first purified by Ni-NTA chromatography as described above. The fractions were treated directly with 3% (w/w) of 0.8μg/μL TEV protease and incubated on ice for at least two days to remove the His-tag. The result was examined by SDS-PAGE.

Size exclusion chromatography
In some cases, hNeil2 WT was further purified by a Superdex75 size exclusion column (GE Healthcare) connected to the Äkta FPLC instrument. It is a single buffer system. The column was equilibrated with 25mL of the running buffer (buffer F) and the hNeil2 in the Hitrap buffer was then applied to a 1mL loop. 0.5mL protein fractions were collected and analysed on SDS-PAGE.

2.2 Cross-linking of Neil proteins with abasic site containing DNA

All the DNA oligos were purchased from (Eurofins). The list of DNA oligos is given in Appendix. The received tubes with DNA were centrifuged for one minute and then MilliQ water was added to each tube so that the resulting DNA concentration was 10mM. 5µL of each AP site containing DNA and its complementary strand DNA solution were mixed together and centrifuged for one minute, followed by annealing at 80°C for five minutes on a PCR machine. The annealed DNA solution was slowly cooled to room temperature and kept in a freezer.

A real AP site was created by treating the annealed DNA containing a single uracil nucleotide with *E.coli* UDG enzyme. Following the UDG product specification from New England Biolabs (NEB), 10µL of 5mM annealed DNA was mixed with 70µL UDG, 200µL 5x UDG reaction buffer and 720µL MQ water. The reaction took place at 37 °C for two hours.

Besides the purchased UDG enzyme, a human UDG protein available in the pET28a expression vector was also expressed and purified by affinity chromatography as described for the Neil proteins. The buffers used for the ion exchange chromatography are given in Appendix.

The Neil protein collected from the ion exchange chromatography with a volume of ~10mL was pooled and transferred to a tube with 12-14kDa MW cut-off membrane (Spectrum Labs). The sealed membrane tube was placed into two litres of cross-linking buffer (buffer G) and dialysed for two hours at 4°C. The protein was then concentrated to a smaller volume around 5mL. The AP site containing dsDNA, prepared using UDG as describe above, was added to the protein with a molar ratio 1.2:1. $NaBH_4$ was then added to the protein-DNA mix with a final concentration of 50mM. The reaction was proceeded on ice for at least 2 days. The result of the cross-linking reaction was checked by SDS-PAGE.

The protein-DNA complex was further purified by ion exchange chromatography. A 5mL Hitrap Q column was used to separate the cross-linked complex from the unbound protein and free unreacted DNA. The low salt buffer was the cross-linking buffer used in dialysis while the salt gradient was still from 50mM to 2M NaCl. The result was checked by SDS-PAGE.

2.3 Crystallisation screening of Neil proteins

2.3.1 Creation of Neil - AP site analogue complexes

Oligos containing a tetrahydrofurane (THF) nucleotide as an AP-site analogue were used to prepare protein-DNA complex for crystallisation. The THF cannot be processed by Neil2, but it can bind non-covalently as a substrate analogue. The annealing procedure was the same as described above. The protein-AP site analogue complexes were generated by adding an annealed DNA directly to a concentrated protein solution with a molar ratio of 1.2:1. The protein solution with DNA was then ready for crystallisation.

2.3.2 Crystallisation screening of Neil2 proteins and Neil2-DNA complexes

The sitting-drop format of vapour diffusion method was used for Neil2 proteins crystallisation screening in this thesis. All the sitting-drop format crystallisation was performed by the Mosquito® crystallisation robot from TTP Labtech. A triple platform sitting-drop 96-well plate (TTP Labtech) was used to dispense the crystallisation drops, and the plate was sealed by a transparent sealing tape. **Table 2.2** shows the overview of the different conditions for screening the Neil2 proteins. For the details of all the kits used in this study, see Appendix. The Neil2 proteins or Neil2-DNA complex solutions were mixed with the reservoir solutions in various ratio as shown in **Table 2.2**.

Table 2.2 Overview of crystallisation screening of Neil2 and Neil2-DNA complexes. *The DNA contain a THF AP site analogue with different bases (T, A or G) in the opposite strand as specified. RT= room temperature. T= temperature.*

Protein	DNA	KITs	Drop scale (nL)	T
0.6mM full-length hNeil2	-	JCSG+, PGA, MEM Gold, Hampton index, Natrix	250 + 250	4°C
	THF:G, 11mer			
	THF:A, 13mer			
0.8mM full-length hNeil2	-	Hampton index, PGA	150 +150	
	THF:A, 13mer			
	THF:G, 13mer, a			
0.3mM full-length hNeil2	THF:A, 13mer	JCSG+, PGA, Wizard cryo	200 + 200	
	THF:G, 13mer, b			
	THF:A, 15mer			
	THF:G, 15mer, a			
0.25mM full-length hNeil2	THF:A, 13mer	PGA, Wizard cryo	100 + 200, 150 + 150, 200 + 100	4°C, RT
	THF:G, 13mer, a			
	THF:A, 15mer			
	THF:G, 15mer, b			
0.1mM full-length mNeil2	-	Hampton index, PGA, Wizard cryo	200 + 200	4°C
	THF:A, 15mer			
	THF:G, 15mer, a			
0.4mM hNeil2 (1-324)	-	PGA, Wizard cryo	250 + 250	4°C
	THF:G, 15mer, b			
	THF:G, 13mer, b			
0.3mM full-length hNeil2	-	PGA, MEM Gold, Natrix	250 + 250	4°C
	THF:G, 15mer, a			
	THF:G, 13mer, a			

2.3.3 Crystallisation screening of Neil1 and cross-linked Neil1-DNA complexes

Both sitting drop and hanging drop methods were used for Neil1 screening. The sitting-drop was performed by the Mosquito® robot as used for the Neil2 screening, while the hanging-drop was performed manually. hNeil1 mentioned in this part denotes to human Neil1 (1-337).

Table 2.3 Overview of sitting-drop crystallisation conditions of Neil1 and Neil1-DNA complex

Protein	DNA	KITs	Drop scale (nL)	T
0.3mM hNeil1	-	PGA, Wizard cryo, Hampton index, MEM Gold	250 + 250	4°C
	AP:G, 10mer			
	AP:G, 12mer			
0.5mM hNeil1	-			
	AP:G, 11mer			
	AP:G, 12mer			

The hanging-drop method was used in optimising hNeil1 crystallisation screening. A 48-well silicone-coated plate was employed to perform the crystallisation. 50μL of a home-made kit library was added to each well as the reservoir solution. 2x2μL protein solution was placed onto a cover glass slide (Hampton Research) and an equal volume of the reservoir solution was added to each protein drop. The glass slide was flipped over and placed carefully onto each corresponding well. The experiment was performed manually in a cold room at 4°C. The composition of the home-made kit library is shown in **Table 2.4**.

Table 2.4 The recipe of the Kit for Neil1 screening optimisation. *All the wells are with 200mM NaCl. Some wells are with two buffers the two buffers were mixed with equal volume.*

	1	**2**	**3**	**4**	**5**
A	4%(w/v) PEG4000, MES pH6.5	6%(w/v) PEG4000, MES pH6.5	7.7%(w/v) PEG4000, MES pH6.5	10%(w/v) PEG4000, MES pH6.5	15%(w/v) PEG4000, MES pH6.5
B	4%(w/v) PEG4000, MES pH6.5, Tris pH7.0	6%(w/v) PEG4000, MES pH6.5, Tris pH7.0	7.7%(w/v) PEG4000, MES pH6.5, Tris pH7.0	10%(w/v) PEG4000, MES pH6.5, Tris pH7.0	15%(w/v) PEG4000, MES pH6.5, Tris pH7.0
C	4%(w/v) PEG4000, Tris pH7.0	6%(w/v) PEG4000, Tris pH7.0	7.7%(w/v) PEG4000, Tris pH7.0	10%(w/v) PEG4000, Tris pH7.0	15%(w/v) PEG4000, Tris pH7.0
D	4%(w/v) PEG4000, Tris pH7.0, Tris pH7.5	6%(w/v) PEG4000, Tris pH7.0, Tris pH7.5	7.7%(w/v) PEG4000, Tris pH7.0, Tris pH7.5	10%(w/v) PEG4000, Tris pH7.0, Tris pH7.5	15%(w/v) PEG4000, Tris pH7.0, Tris pH7.5
E	4%(w/v) PEG4000, Tris pH7.5	6%(w/v) PEG4000, Tris pH7.5	7.7%(w/v) PEG4000, Tris pH7.5	10%(w/v) PEG4000, Tris pH7.5	15%(w/v) PEG4000, Tris pH7.5

2.3.4 Diffraction study of Neil crystals

The obtained crystals of DNA-crosslinked hNeil1 were soaked for a few seconds in a cryoprotectant solution containing 25% ethylene glycol and 75% reservoir solution, and then flash-frozen in liquid nitrogen. X-ray diffraction experiments were done at the ESRF synchrotron in Grenoble, France.The synchrotron beamlines produce high intensity X-rays which allows the use of very small crystals. The diffraction data of Neil crystals obtained in this study was collected on the ESRF beamline ID23-1. The diffraction data was collected at 100K with a wavelength 0.97702Å, exposure time 0.5s and oscillation range 1°.

2.4. Stability study of full-length human Neil2

Thermal Fluor Assay (TFA), also known as Differential Scanning Fluorimetry (DSF), was used to study the thermal stability of mouse Neil2 WT, human Neil2 WT and hNeil2 (1-324). Neil2 in this part refers to these three versions of Neil2. All the reactions were processed on a 96-well PCR plate and the plate was sealed by a MicroAmp optical adhesive film (Life Technologies). The reaction plates were slowly heated to 95°C while measuring the change in fluorescence of the reporter Protein Thermal Shift^{TM} dye as it binds to the unfolding protein. The data was collected on a Real-Time PCR instrument from Life Technologies and analysed by StepOnePlus^{TM} software. The melting point temperature (T_m) was determined by analysing the first derivative of the fluorescence curves.

In principle, each reaction followed the recipe provided by Life Technologies. All the components are shown in **Table 2.5**. Besides the prescribed buffer (Life Technologies), a home-made buffer composed of 70% buffer D and 30% buffer E was also used in the assay. This mix was used to get a similar condition to which Neil2 was eluted from the Hitrap column. The 1000x Protein Thermal Shift^{TM} dye (SYPRO Orange, Life Technology) was diluted to 8x by MQ water right before each experiment. Three replicates of each reaction were prepare for all sets of experiments.

Table 2.5 The composition of each TFA reaction

COMPONENT	VOLUME
Protein thermal shift^{TM}/Home-made buffer	5.0 µL
Protein + MilliQ water	12.5µL
Diluted protein thermal shift^{tm} dye (8x)	2.5 µL
Total volume for each reaction	**20.0µl**

In order to study the effects of different amounts of protein on the melting temperature of Neil2, three different concentrations of Neil2 proteins were used in each experiment which was 5µM, 10µM and 20µM in 12.5µL volume. Moreover, 100% pure DMSO was added to some of the 5µM Neil2 reactions to study the stability of Neil2 proteins under the presence of organic compound. 2%, 5%, 10% and 20% (v/v) of DMSO in 20µL volume were tested with Neil2 in the assay. At last, 0.02% (w/v) of Tween solution which is a detergent additive was added to each reaction at one plate to test if it can stabilise the Neil2 protein.

Before preparing the reaction samples in the PCR plate, Neil2 in the reaction buffer was first analysed on a Dynamic Light Scattering (DLS) instrument to make sure the buffer condition for the protein was good.

2.5 Screening for ligands binding

Since there is no crystal structure of human Neil2 available, the same compounds resulting from a previous computational docking study of the human Neil1 homolog was used in the present study (You, 2014). The NCI diversity set IV library containing 1596 compounds was screened using the crystal structure of human Neil1 (PDB: 1TDH). According to You's study, the 300 highest ranking compounds from the Schrödinger Glide docking module were manually inspected in PyMol, and 79 compounds were finally selected for experimental testing. The 79 compounds were ordered from the NCI Drug Synthesis and Chemistry Branch, Chemotherapeutic Agents Repository. All the compounds were dissolved in pure DMSO to a concentration of 10mM.

2.5.1 nano DSF

A Differential Scanning Fluorimetry instrument (Prometheus NT.48) from the company NanoTemper was employed to study the possible binding of the compounds to full-length hNeil2. 44 of the 79 compounds were selected based on an inspection of the docking solution using PyMol, due to the capacity limitation of the instrument. The purified Neil2 from ion exchange affinity chromatography was concentrated to ~0.1mM and diluted to 0.02mM by buffer D. 1 μL 10mM ligand was added to 39μL 0.02mM hNeil2 solution and mixed well. The capillaries were loaded by dipping directly into the prepared protein-ligand solution and applied to the instrument. The initial fluorescence scans were performed at 20°C, and displayed the emission at wavelength 330nm and 350nm to test if all the samples were within the optimal concentration range. The full thermal unfolding experiment was run from 15°C to 95°C with the temperature increasing by 1°C/min. The melting temperature (T_m) was determined by the fluorescence signal only at 330nm. The experiment with the 44 ligands was performed only once and all the samples were without replicates due to capacity limitations. Another run was performed based on the results of the first run. 8 of the 44 compounds with promising T_m profiles were selected and diluted by pure DMSO to different concentrations. Due to the limitation of the maximal DMSO concentration tolerated by the protein solution, the highest ligand concentration was 250μM and decreased successively by 50% to the lowest 31.25μM. Each diluted ligand was mixed with 0.02mM Neil2 with final 2.5% DMSO in the solution. The experiment settings were the same as the first run.

2.5.2.MST

Following the results from nanoDSF, the same eight compounds were chosen for further testing on a MicroScale Thermophoresis (MST) instrument from NanoTemper. Similar to nanoDSF, MST is also a good and sensitive method to study the affinity between ligands and proteins. Both a label-free MST (Monolith LF, NanoTemper Technology) and a standard MST instrument for labelled proteins (Monolith NT.115, NanoTemper Technology) were used in this study. The protein solution can be analysed directly by label-free MST. It uses the intrinsic fluorescence change of the protein, which is dominated by the presence of tryptophan residues. The NT.115 MST instrument on the other hand, requires that the target protein is labeled by a blue or red dye that modify lysines or cysteines in the protein.

Label-free MST

A test capillary scan was first performed on a label-free MST instrument to make sure the intrinsic fluorescence of all eight ligands was close to the background level. 1μL 5mM of each ligand in pure DMSO was diluted in 39μL PBS buffer to 125μM. Equal volume of the diluted ligand and 1μM hNeil2 was mixed together and was loaded directly into 10μL capillaries. The capillaries were placed onto the tray and a capillary scan was run with 20% LED power.

In order to ensure the concentration of hNeil2 and ligands were optimal, another pre-test on label-free MST was done by using different protein concentrations. 125nM, 250nM and 500nM of hNeil2 were prepared in PBS buffer. One of the ligands with highest concentration (125μM), PBS buffer and MilliQ water were used in the test as controls.

Standard MST

Before being applied to the MST instrument, the purified Neil2 was labeled. The protein labelling kit with blue-NHS dye was provided by NanoTemper Technology. The labelling procedure can be divided into three parts. The first part was the buffer exchange. 100μL Neil2 protein was added to a 0.5mL centrifuge filter (Amicon® Ultra-0.5) and centrifuged at 1500g for 1min. 300μL of the Labelling Buffer (NanoTemper Tech.) was then added to the tube and the tube was centrifuged again at 1500g for 1min. This step was repeated for 3 times, and the protein was diluted in PBS buffer. The second part was the labelling. The protein concentration in the labelling buffer was measured to ensure it was in the range of 2-20μM. In this case, the hNeil2 was adjusted to 5μM by labelling buffer. The blue-NHS dye was dissolved completely in 30μL pure DMSO. And 3μL of the dye was diluted to 100μL by labelling buffer to a concentration of 15μM. The Neil2 protein and the diluted dye was mixed in a 1:1 volume ratio, 200μL in total. The mixture was incubated on ice in the dark for one hour. The last step was the purification. The gravity flow column B (NanoTemper Tech.) was equilibrated and washed by 3x3mL of buffer H which was used as the assay buffer for MST analysis. The 200μL labelled Neil2 solution was applied to the column, followed by adding 300μL of the buffer H. After all the solution entered the column bed, 600μL of the buffer H was added to the column and the eluate collected in 100μL fractions. The two collected fractions with apparently blue colour were analysed by a UV spectrophotometer, NanoDrop.

After determining the labeled Neil2 concentration, the protein was diluted to 100nM because of the sensitivity of the MST instrument. In addition, in order to prevent protein aggregation, 0.1% (v/w) Pluronic F-127 was added to the labeled sample. Dilution series of the seven ligands (without ligand 3) were made from the highest 250μM and decreased sequentially 16 times with 2.5% DMSO. Equal volume of Neil2 and each of the ligand series was mixed well by an electronic pipet. The standard capillaries were dipped directly into the prepared protein-ligand series and loaded to the instrument. Seven full MST experiments were run individually at 20% and 40% MST power, with a 50% LED power. The experiments were performed at 20°C. Ligands 14, 20, 34 and 43 were analysed again with higher concentrations. The ligand dilution series in this second run had a concentration of 1.5mM and diluted 16 times with 5% DMSO. The experimental setting was the same as for the first run series. Moreover, to increase the reliability of the collected data, another replicative series was performed with the same

concentration range of the four ligands. For this last run, ligand 40 which known as an inhibitor of human Neil1 was also tested in the experiment.

In some case, a SDS-denaturation test (SD-test) was performed to identify the reason of fluorescence loss. The protein-ligand samples of the highest and lowest ligand concentrations were centrifuged at 15,000g for 10min. 10µL of the supernatant was transferred to the tube containing 10µL of a 2xSD mix (4% SDS, 40mM DTT) and incubated at 95°C for 5min. After the tubes cooled down to room temperature, the samples were loaded to standard capillaries and run a capillary-scan to measure the fluorescence intensities.

2.6. Activity assay of full-length human Neil2

2.6.1 Preparation of DNA substrate

The DNA substrate for activity assay was purchased from Midland Centified Reagent Company (MCRC). One of the DNA oligos was modified at the 5' terminal by Carboxytetramethylrhodamine (TAMRA), while the opposite strand was tagged by a Black Hole Quencher (BHQ-2) at the 3' terminal. The sequences of the two DNA oligos are,

5'-Tamra-TCACC(dU)TCGTACGACTC-3'
5'-GAGTCGTACGAAGGTGA-(BHQ-2)-3'

The two lyophilized oligonucleotides were dissolved in MilliQ water to a concentration of 1mM. 2μL of each oligonucleotide stock solution was added to 2μL 10x annealing buffer (100mM Tris-HCl pH7.5, 1M NaCl, 10mM EDTA) and 14μL MilliQ water. The prepared oligo solution was heated to 60°C for 10 minutes and cooled down slowly to room temperature on the bench.

In order to prepare an AP site in the DNA substrate, 5μL of UDG enzyme (5000 units/mL, NEB), 5μL of 10x UDG buffer (NEB) and 20μL MilliQ water were added to the annealed DNA solution. The mixture was incubated at 37°C for at least 2 hours. The DNA solution was diluted by annealing buffer to a final concentration of 200nM.

2.6.2 Development of plate assay with full-length human Neil2

To simplify the description, hNeil2 denotes full-length human Neil2 in this section. To determine the best enzyme concentration for performing the assay, the hNeil2 in ion exchange chromatography buffer was diluted to 200nM, 1μM and 2μM, respectively by 2x Reaction Buffer (2xRB, the recipe of 5xRB, see Appendix). *E.coli* endonuclease IV (EndoIV) that was also purified in our lab was used as a positive control with the concentration of 1uM. A 384-wells plate and a Victor 2X plate Reader instrument (Perkin Elmer) was employed to perform the assay. The plate was kept in the reader at 20°C and read every 10 minutes for 10 repeats in total. The excitation wavelength was 534nm and the emission wavelength was 590nm. The reading time of each well was 2 seconds. The experiment setup is shown in **Table 2.6**. The 10⌈L DNA solution was loaded first into the wells and the protein solutions were added on the top of the DNA drops by an electronic pipette (Eppendorf). Every reaction was perpared with 4 parallels.

Table 2.6 The experiment set-up of different concentration hNeil2 activity assay.

No.	Well contents (10μL +10μL)	
A	2xRB	200nM DNA solution
B	1μM EndoIV	
C	200nM Neil2	
D	1μM Neil2	
E	2μM Neil2	

2.6.3. Screening of wild type and mutant hNeil2

In order to study the role of some specific residues that may be involved in the activity of Neil2, six different variants of Neil2 were selected. Those mutants are listed below. All the mutants were available in the same pET22 plasmid as wild type (WT) Neil2. All the mutants and WT hNeil2 are full-length proteins.

Table 2.7 The mutation sites of all the mutants

Name	Mutation site(s)
Mutant A	P2A + K50A
Mutant B	P2A + E3A + K50A
Mutant C	P2A + K50A + K51A
Mutant D	K50A
Mutant E	K51A
Mutant F	K50A + K51A

Expression and purification of Neil2 WT and mutants

The expression and purification procedures of the Neil2 variants were nearly the same as described in **section 2.1**, except that the incubation was shortened from over night to only four hours at 37°C. This was to make sure all the variants could be purified at the same time and kept as fresh as possible.

Assay set-up

After the WT and all the mutant hNeil2 proteins were purified by Ni-NTA and ion exchange chromatography, the protein containing fractions were pooled and diluted to 0.2μM and 1μM by 2xRB. Like described in **section 2.6.2**, only DNA in reaction buffer was used as the background control while the EndoIV with DNA was the positive control. Each samples had 4 parallels. The plate was read ten times at ten-minutes intervals at 20°C.

2.6.4. Screening of five ligands with WT human Neil2

In order to study if the ligands identified as possible hNeil2 binders by MST and nanoDSF can inhibit the activity of hNeil2, four ligands were tested in an activity assay. In addition, another Neil1 inhibitor discovered by You (You, 2014) was also tested. The hNeil2 WT protein used here was successively purified by Ni-NTA, ion exchange and size exclusion chromatography. The five ligands were first diluted by 2xRB, followed by adding the protein solution to a final concentration of 100μM ligand with 1μM protein in the 5% (v/v) DMSO. The mixture was incubated for a few minutes on ice before the DNA substrate (200nM) was added. The assay was run with the same setting as described before.

The two best of the five ligands that showed inhibition ability were further tested to determine IC_{50} values. A concentration gradient of the chosen ligands was made in order to determine the IC_{50} value, which is the concentration of the inhibitor that reduces the protein activity by 50%. A serial-dilution of the ligand in 100% DMSO from 16mM to 125μM was prepared. 4μL of the ligand stock solution was then mixed with the enzyme (3μL, 1mg/mL) and 73μL of 2xRB to a final volume of 80μL. The volume was prepared for 7 replications. The 200nM DNA solution and protein-ligand solution were added successively to a plate. Samples for background and positive control were also included. The assay was performed following the same protocol as before.

3. Results and discussion

3.1 Expression and purification of Neil proteins

3.1.1 Purification of Neil proteins

The methods and buffers used in the protein expression and affinity chromatography purification of hNeil1 (1-337), full-length hNeil2, hNeil2 (1-324) and full-length mouse Neil2 were almost the same. ***Figure 3.1 and 3.2*** are SDS-PAGE results of hNeil2 after purification by Ni-NTA and ion exchange chromatography, respectively. Since the PAGE results are quite similar for all the Neil proteins mentioned above, only the scanned images of hNeil2 are shown here. Ni-NTA based protein purification is used for purifying polyhistidine-containing recombinant proteins. The Neil-tagged proteins bound to the Ni-NTA resin were eluted by competition with high concentration of imidazole. However, the Neil proteins were not totally pure as shown in **Figure 3.1**, there are still many weak bands which indicates the presence of some bacterial proteins in those fractions. The protein used for crystallisation requires higher purity (Chayen & Saridakis, 2008). So further purification steps were performed.

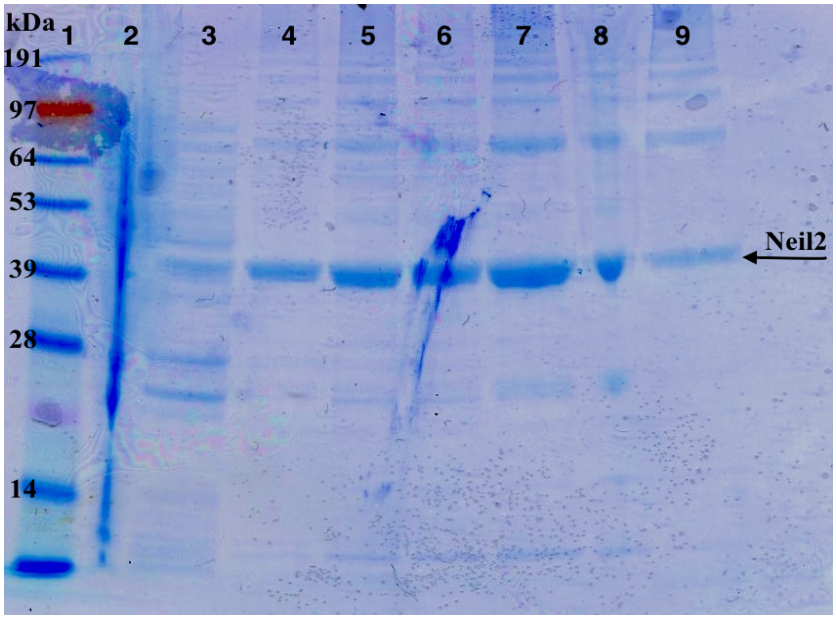


Figure 3.1 SDS-PAGE result of the full-length human Neil2 purified by Ni-NTA chromatography. *Lane 1 is the SeeBlue 1x standard marker. Lane 2 is the flow through containing E.coli and unbound Neil proteins. Lane 3 is the wash fraction by 25mL buffer A. Lanes 4-6 are the wash fractions by buffer B containing 50mM imdazole, while the last three lanes are the elution fractions by buffer C containing 300mM imdazole. The dark blue bands in lanes 4-9 indicate the hNeil2 protein.*

The ion exchange chromatography separates proteins based on the different charges of molecules. The Hitrap SP is a cation column. The Neil proteins in the low salt buffer (buffer D) can bind to the column and be eluted at a specific salt concentration by increasing the salt concentration due to the ionic strength change. The β-ME in the buffer is a reducing agent to stabilise the protein (T. W. Green, 1999). The chromatogram and scanned SDS-PAGE gel of

the purification are shown in **Figure 3.2**.

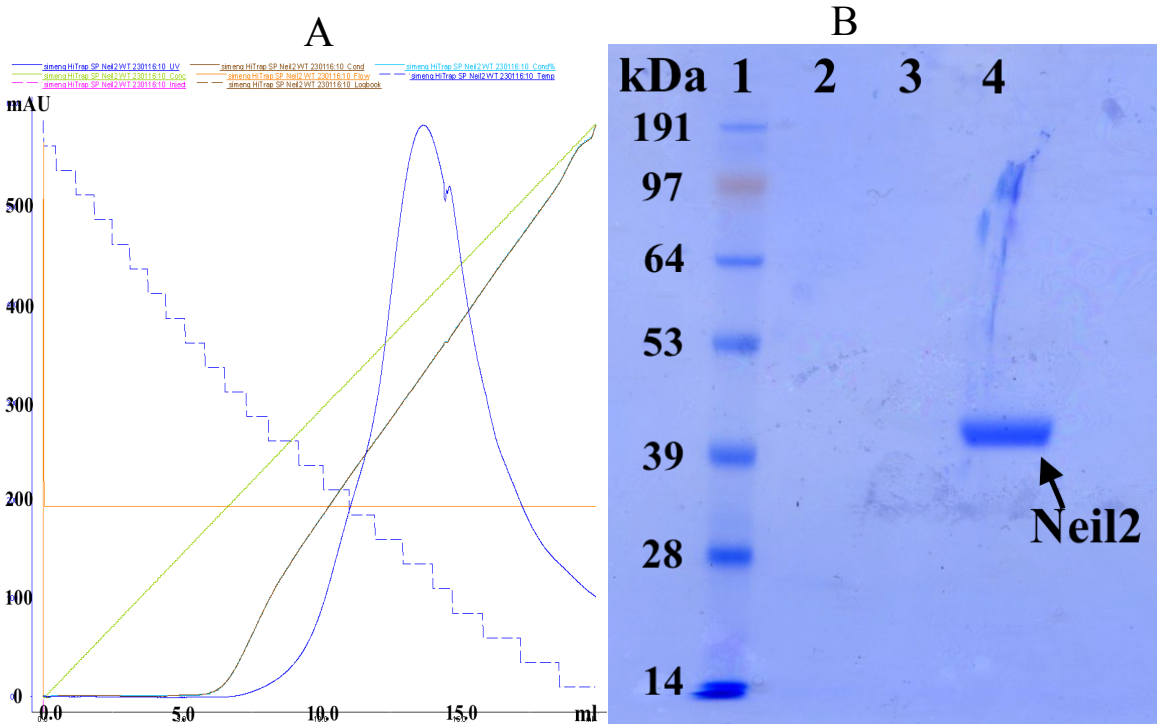


Figure 3.2 A. Chromatogram of hNeil2 purified by ion exchange chromatography. *The blue line is the absorbance at 280nm (unit: mAU) and the peak indicates the presence of hNeil2. The green line shows the increasing percentage of high salt buffer during elution. The x-axis shows the elution volume.* ***B. SDS-PAGE result of ion exchange purification of hNeil2.*** *The 1st lane is the SeeBlue 1x standard marker. The 2nd and 3rd lanes are flow through and wash fraction respectively. The dark blue band in the 4th lane is the purified hNeil2.*

3.1.2 Expression of human Neil2 truncated variants

The two truncated Neil2 variants (with 250 and 290 residues) were first expressed as described for the other Neil proteins. However, there were only very weak bands on the gel after the proteins were purified by Ni-NTA chromatography. A possible explanation for this could be that the protein was not expressed successfully, or the His-tag could stick to the protein and not work properly. Or the proteins formed inclusion bodies in *E.coli* and were not soluble. Thus, a small scale expression test was performed. After incubated over night, the collected cell pellets was dissolved in a cell lysis buffer to crack the *E.coli* cells. The solution was denatured and two different volumes of the solution were loaded to a NuPAGE gel with two volumes. Unfortunately, none of the samples indicated a normal expression of protein based on the results of SDS-PAGE. According to the study by Das and Rajagopalan (Das et al., 2004), there is a zinc finger domain at the C-terminal of hNeil2. The function of a zinc finger domain is to help the macromolecules binding and stabilise the protein folding. The zinc finger domain of hNeil2 includes residues from 284 to 319. However, these two truncated hNeil2 variants exclude part of this domain, which may lead to a low expression efficiency of the proteins. So these two truncated hNeil2 variants were not continued in this study. Therefore, another truncated hNeil2 variant with 324 residues, which also includes the zinc finger domain, was also tried to be expressed in the same way. The expression of the hNeil2 (1-324) was successful and the protein

was determined to be continued in this study.

3.1.3 His-tag removal of human Neil2 (1-324) by TEV protease

The 6x-His-tag is mainly used for protein purification by immobilised metal-affinity chromatography since histidine exhibits the strong interaction with metal ion (Bornhorst & Falke, 2000). So the His-tag at the C-terminal helps the Neil proteins binding to Ni-NTA column. However, the His-tag may affect the protein folding and activity, and the tag can also sometimes interfere with substrate binding or crystallisation. Therefore, it is normally better to remove the tag from the protein. TEV protease is a common used enzyme for cleaving tags in recombinant proteins. It can recognise a linear epitope and cleave at a specific site (Waugh, 2011). The cleavage result of hNeil2 treated with TEV is presented in **Figure 3.3**. Obviously, the gel band of the experimental sample that was treated by TEV protease is positioned lower than the untreated sample on the gel. It indicates that the cleavage was complete.

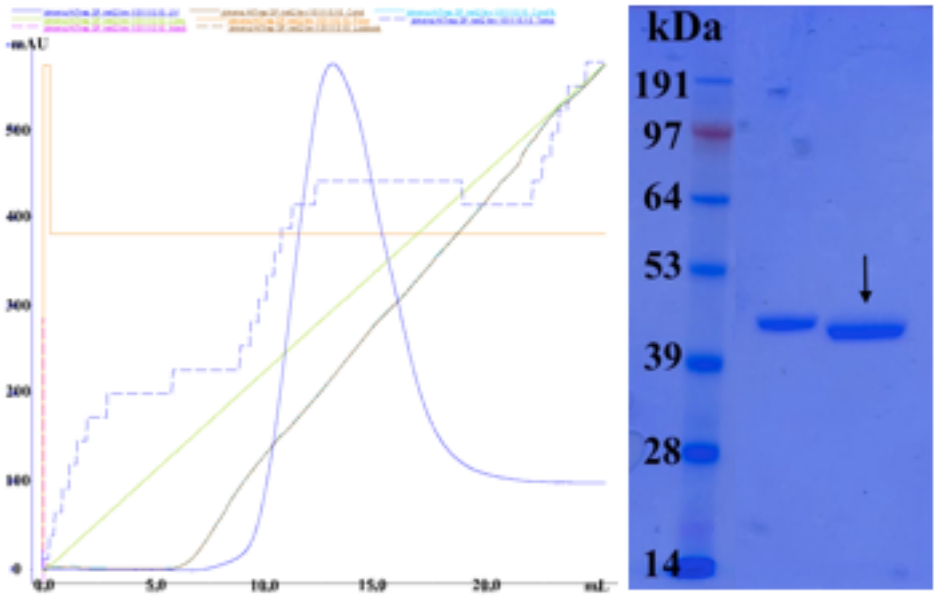


Figure 3.3 **Left: The HiTrap SP chromatogram of TEV protease treated hNeil2.** The blue line is the absorbance at 280nm. The peak at around 12-15mL is the hNeil2 protein. **Right: The SDS-PAGE result of TEV protease treated hNeil2 (1-324).** The band marked with an arrow is the TEV treated hNeil2 (1-324) sample while the left lane is the untreated control sample with hNeil2 (1-324).

3.2 Cross-linking of Neil proteins with DNA

Human Neil2 and abasic DNA cross-linking

The Neil enzymes are trifunctional DNA glycosylases and the cross-linking mechanism involves a Schiff base (see **Section 1.3 and 1.4**). This is a complex between the protein and the DNA that can be trapped by using the reducing agent NaBH4. Moreover, the amino acids in Neil2 that are involved could be Pro2, Lys50 and Lys 51. The full-length human Neil2 and the His-tag removed human Neil2 (1-324) were both subjected to cross-linking with DNA substrates. Unfortunately, the yield of the cross-linked Neil2-DNA was either very low or nearly none. Many factors can affect the cross-linking result. One of the possible reasons could be that the DNA substrates were not suitable for Neil2 binding. Some studies claim that Neil2 prefers bubble DNA substrates than normal linear DNA, duplex or single strand (Dou et al., 2003). The study also indicates that Neil1 has a higher affinity to duplex or single strand DNA than Neil2. So we decided to also try to cross-link Neil1 with the same DNA substrates as a control.

Human Neil1 and abasic DNA cross-linking

The abasic site, also known as AP (apurinic/apyrimidinic) site, is a damaged DNA base that contains neither a purine nor a pyrimidine. This site is also an intermediate in the base excision repair, and the AP site can be recognised by a DNA glycosylase (Boiteux & Guillet, 2004). The Neil proteins can thus bind to a DNA substrate containing an AP site which may improve the chances of obtaining crystals for structural studies. The AP site containing DNA substrate was created by adding UDG glycosylase to a DNA containing a uracil base. The uracil will be removed by the UDG enzyme and the AP site is formed.

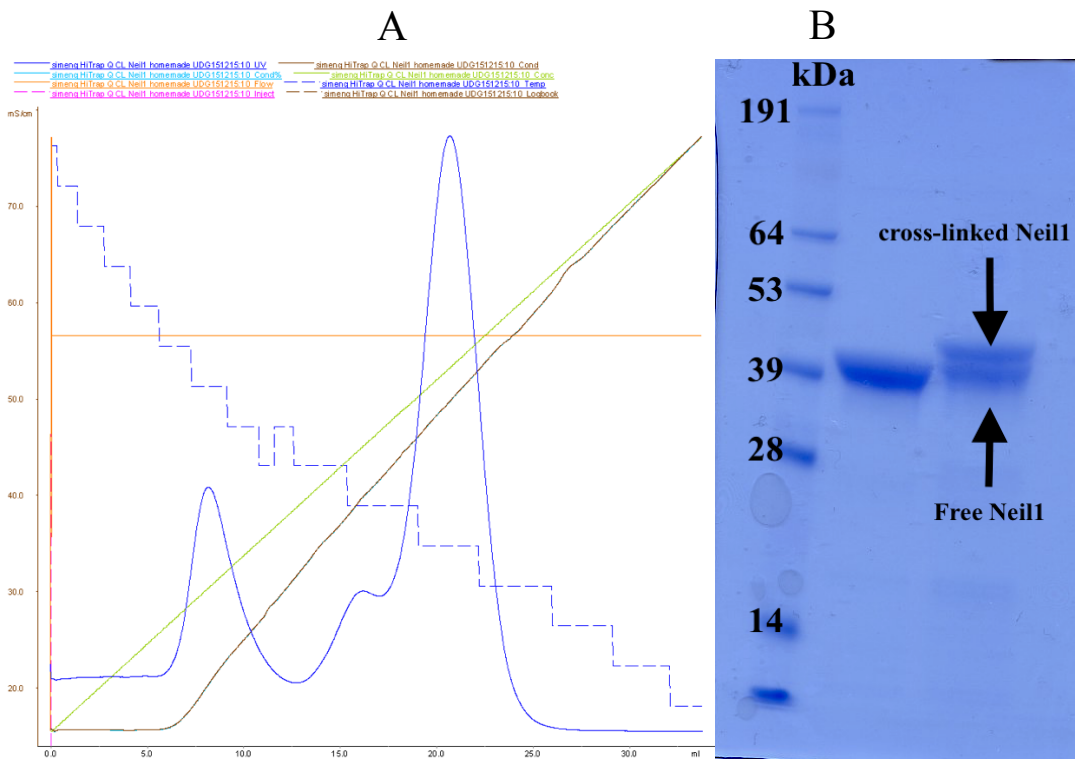


Figure 3.4 A. The cross-linking of hNeil1 with abasic site containing DNA. *The second lane is the untreated control sample while the third lane is the experimental sample. The third lane has two bands, and the lower one is free Neil1 that were not cross-linked.* ***B. The Hitrap chromatogram of the cross-linking Neil1-DNA complex.*** *The first peak at around 8mL is the cross-linked protein-DNA. The second peak at around 15-20mL is the unbounded DNA.*

3.3 Crystallisation of Neil proteins

A known protein structure plays a very important role in understanding the protein function. However, there is no available human Neil2 structure known yet, neither of the protein alone or in complex with DNA. The closest homologues structure is that of human Neil1, but there are important sequence differences between the two that can only be analysed by an experimental structure. In this study, many different conditions were tried for Neil2 protein screening, as well as several treatments of Neil2 with DNA. Despite numerous experiments, we did not obtain any Neil2 crystal in any of the screening trials. We also experienced that Neil2 is a very unstable protein. Even though no precipitation was observed during the purification procedure, there were still some aggregates in the solution based on DLS analysis (**Figure 3.5**). The results show that the average particle size in the protein solution corresponds to 1-3 megaDalton (mDa) while the molecule weight of hNeil2 is only 37kDa, which indicates that the Neil2 proteins formed large aggregates in the solution. Normally, monodisperse solutions of protein is more likely to succeed in crystallisation experiments (Meagher et al., 2015).

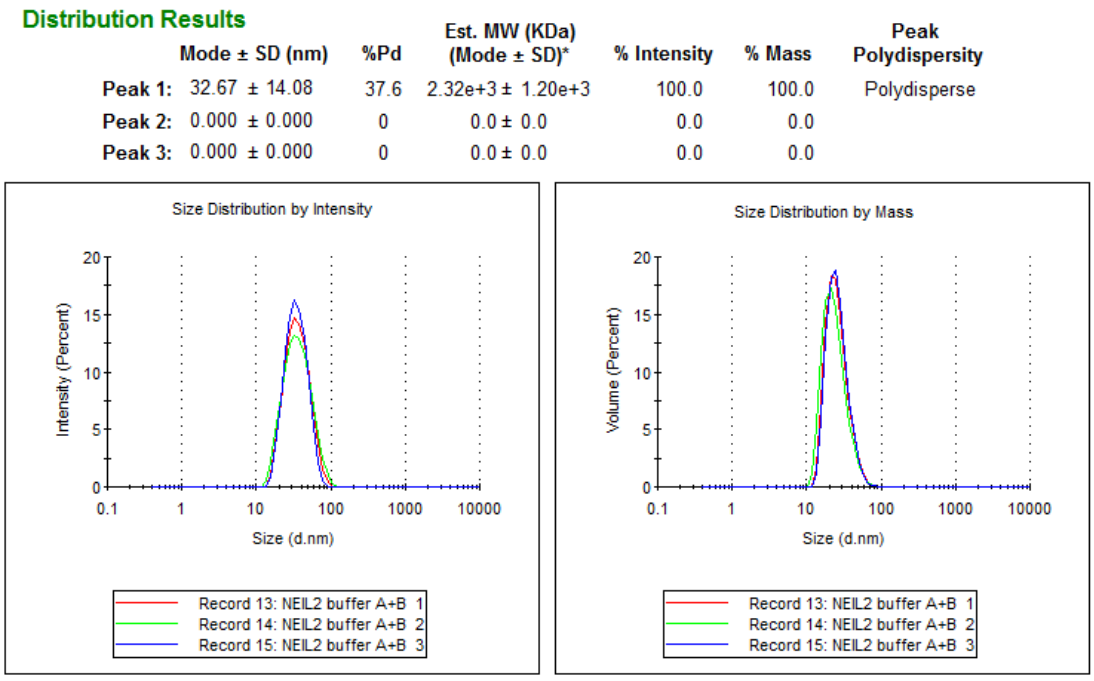


Figure 3.5 hNeil2 in Hitrap buffer analysed by DLS instrument. *The curve of 'size distribution by mass' indicates that the protein molecules in the buffer formed aggregates.*

We also screened the human Neil1 (1-337) cross-linked with DNA as a control trial. Some needle-shaped crystals were grown in the condition 7.7% (w/v) PEG4000, 0.02M Tris pH7.0 and 0.1M NaCl. The Neil1 with 10mg/mL concentration was cross-linked with AP: G, 12mer double strand DNA substrate. A matrix around the initial crystallisation condition was prepared in order to optimise crystal growth. However, no crystal was observed even after two weeks of incubation. Due to the time limitation and Neil1 is not the main study object, no more screening trails were tried with Neil1. The acquired crystals were sent to ESRF synchrotron to collect data. But since the crystal were too thin and small to get any diffraction image.

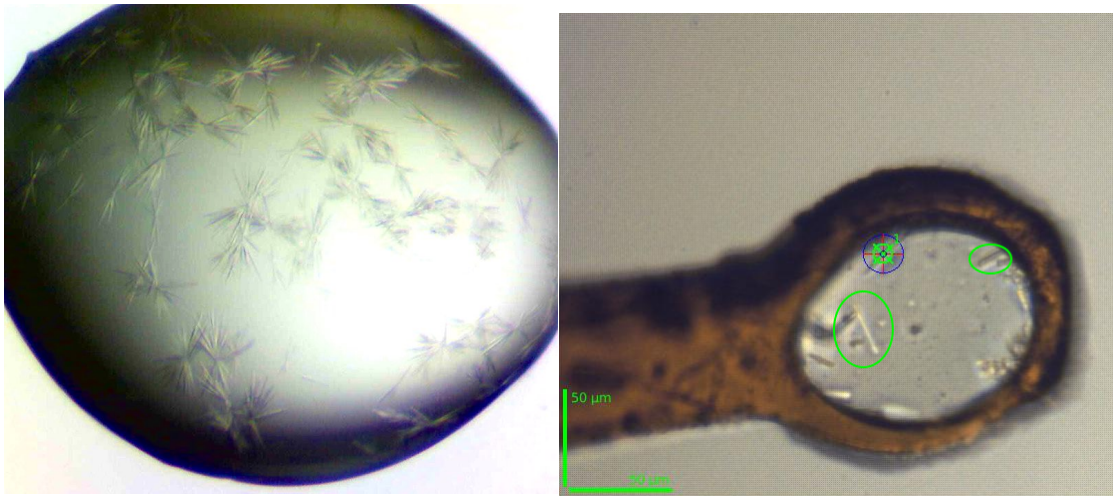


Figure 3.6 Left: Needle shaped crystals of hNeil1 (1-337) cross-linked with DNA. *The crystallisation drop was performed by sitting-drop method. The photo was taken by MotiConnect live camera.* ***Right: The Neil1 crystal mounted in a Litholoop.*** *In the green circle are the needle crystals of hNeil1 (1-337)*

3.4 Thermofluor shift assay for stability study of Neil2

Initially, we decided to use a thermofluor shift assay (TFA) to study the stability of Neil2, and the affinity of ligands with Neil2. Based on a previous screening of ligands for the related Neil1 homologue, a TFA experiment was run of the Neil2 with the 79 compounds. The TFA was following a protocol provided by Life Technology. However, the melting curves were abnormal for all the samples which giving no clear melting temperature (T_m) value. Thus, we investigated the T_m curves of the pure hNeil2 proteins. Different protein concentrations of full-length hNeil2, full-length mNeil2 and hNeil2 (1-324) were tried. We also explored the possible influence by DMSO or additives. Nevertheless, none of the Neil2 melting curves looked regular with a clear melting point. The curves have some common features and **Figure 3.7** shows one example of a Neil2 melting curve. Similar to the shown curve, all the tested samples have a rather high initial fluorescence. The principle of TFA is that the fluorescent dye binds to hydrophobic amino acids of a protein. As the temperature increases, the protein will start to unfold and result in strange binding of dyes as more hydrophobic amino acids are exposed, which gives higher fluorescence. The high initial signal illustrates that at the beginning of the experiment, Neil2 was already unfolded or the dye binds to much to the protein.

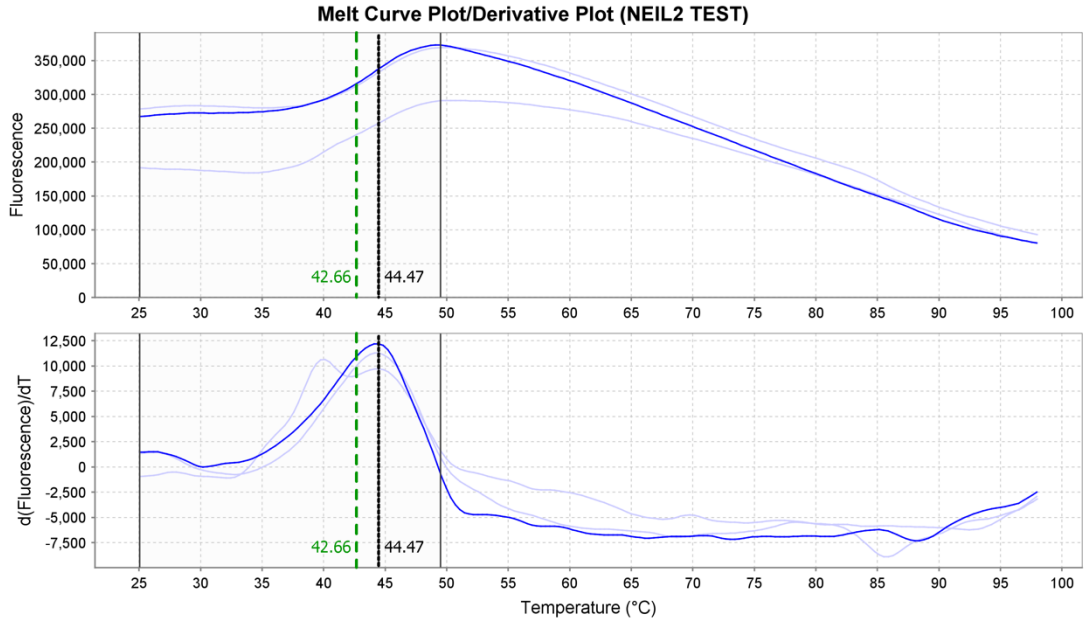


Figure 3.7 The TFA melting curve of full-length human Neil2. *The protein concentration is 5μM. The x-axis shows the temperature range while the y-axis is the fluorescence signal. The lower curve is the derivative plot of the upper one.*

The fluorescent dye used in this study was SYPRO Orange dye. In some cases, this dye interacts with hydrophobic pockets on the protein surface or also with some detergents, which leads to a high initial signal background (Boivin, Kozak, & Meijers, 2013). Besides such unspecific binding of the dye, the Neil2 aggregation might be another factor that can contribute to the unusual curve profile. Therefore, the samples used for TFA was analysed by DLS. The two figures below show the DLS analysis of full-length human Neil2 in the Hitrap buffer in the

presence of dye (**Figure 3.8**) and with 10% (v/v) DMSO (**Figure 3.9**). Compared to the figure of Neil2 in Hitrap buffer only (**Figure 3.5)**, it seems the presence of dye does not affect the protein much. But the size of the particles in solution was still quite large, which also means there are aggregation in the protein solution.

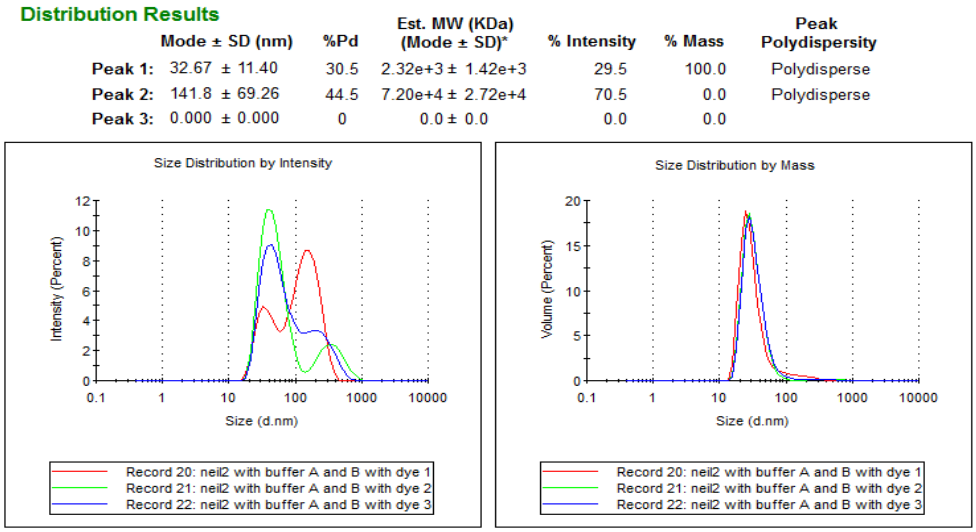


Figure 3.8: DLS profiles of full-length human Neil2 in Hitrap buffer with the SYPRO Orange dye. *The average particle size in the protein solution corresponds to 1-3 mDa.*

The experiments with 2%, 5%, 10% and 20% (v/v) of DMSO were performed to make sure the protein condition was acceptable in the presence of DMSO. The results suggested that if the DMSO concentration is not higher than 5% (v/v), the presence of DMSO in the protein solution is tolerated. However, when the DMSO concentration reached 10% (v/v), the Neil2 protein forms large aggregates. **Figure 3.9** indicates the presence of 10% (v/v) DMSO in the hNeil2 solution is not tolerated. There are only very large molecules (diameter higher than 500nm) in the solution.

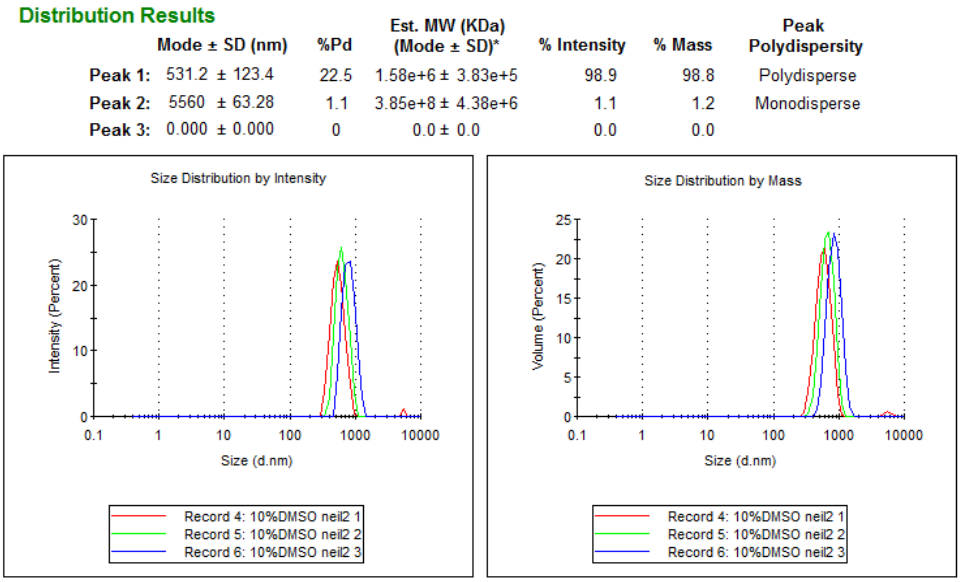


Figure 3.9: DLS profiles of full-length human Neil2 in Hitrap buffer with 10% (v/v) DMSO.

3.5 nanoDSF for inhibitor screening of Neil2

Since we were not able to determine ligand binding to Neil2 by the TFA method using the SYPRO Orange dye, we decided to test a selection of the ligand using a nano differential scanning fluorimetry (nanoDSF), which determines the melting temperature by analysing the fluorescence signal of the native protein. To simplify the description, the 44 ligands were sequentially numbered from ligand 1 to 44. For the information about all the ligands, see **Appendix IV.** All the figures in this section were adapted from the Prometheus NT.48 Report ("Prometheus NT.48 Report, 2016). The hNeil2 in the presence of 44 ligands were first pre-scanned once. **Figure 3.10** shows that the fluorescence intensities of most protein-ligands samples were in the allowed range except for the ligand 12 which has a high fluorescence. Since the basis of nanoDSF analysis of protein folding relies only on the properties of tryptophan in protein itself, the ligands or other additives should not have too high fluorescence background. The red dotted line in the **Figure3.10** is the upper limit. The raw fluorescence counts of the capillary with only hNeil2 was around 10,000. The disparate levels of the peaks with ligands are a result of the presence of the ligands.

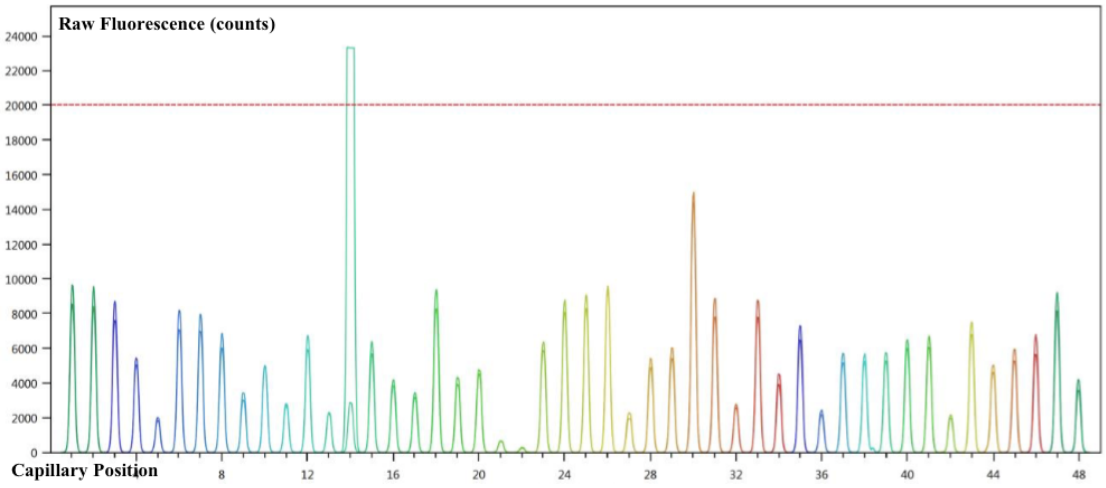


Figure 3.10 The capillary scan of hNeil2 with 44 ligands. *The first two and the second last capillaries (sample No.1, No.2 and No.47) contained only hNeil2 protein. The last capillary (sample No.48) contained lysozyme as a standard control. Capillaries (sample No.3-No.46) contained sequential ligands 1 to 44 mixed with hNeil2 protein. This scan displays the emission at both 330 nm and 350 nm. The outlier is ligand 12 (sample No.14). The red straight line is the maximum tolerated range of the fluorescence intensity.*

After ensuring that the ligands can be used for analysis, a full melting scan of Neil2 with 43 ligands was performed with a starting temperature of 20°C to final 95°C. The fluorescence was measured at regular intervals. The melting temperatures of all the ligands are listed in the **Appendix V First run**. We observed that some samples have two melting temperatures (T_m). The reason is not clear, but we have two hypotheses. One is that the hNeil2 structure can change conformation twice at two different temperatures. The other is related to the two domains in the hNeil2 structure. The melting temperature of these two domains may not be the same. However,

the two melting points did not appear in all of the samples, hence, a more extensive study is needed to find out the actual reason.

According to the initial melting scan of hNeil2 with the 43 ligands, 8 ligands were chosen to be analysed further with different concentrations. The eight ligands were ligand 3, 14, 15, 19, 20, 34, 39 and 43. It was a choice based on the largest difference of the melting temperature between the protein only sample and the protein-ligand samples, as well as a quantitative judgement of the distinct shapes of the different melting curves with ligands. The purpose of this run was to analyse the melting temperature of hNeil2 under the different concentrations of the ligands. The fluorescence was analysed at 330nm and the melting temperature was determined by the first derivative of the fluorescence counts. For an overview of the raw data for all eight ligands, see **Appendix V Second run**. Among those eight ligands, the sample with ligand 14 showed an interesting does-response trend of increasing hNeil2 melting temperature with the increase of ligand 14 concentration. The melting scan plot is shown in **Figure 3.11**. As the concentration of ligand 14 increases, so does the T_m value, which is an expected trend for a ligand binding to the protein.

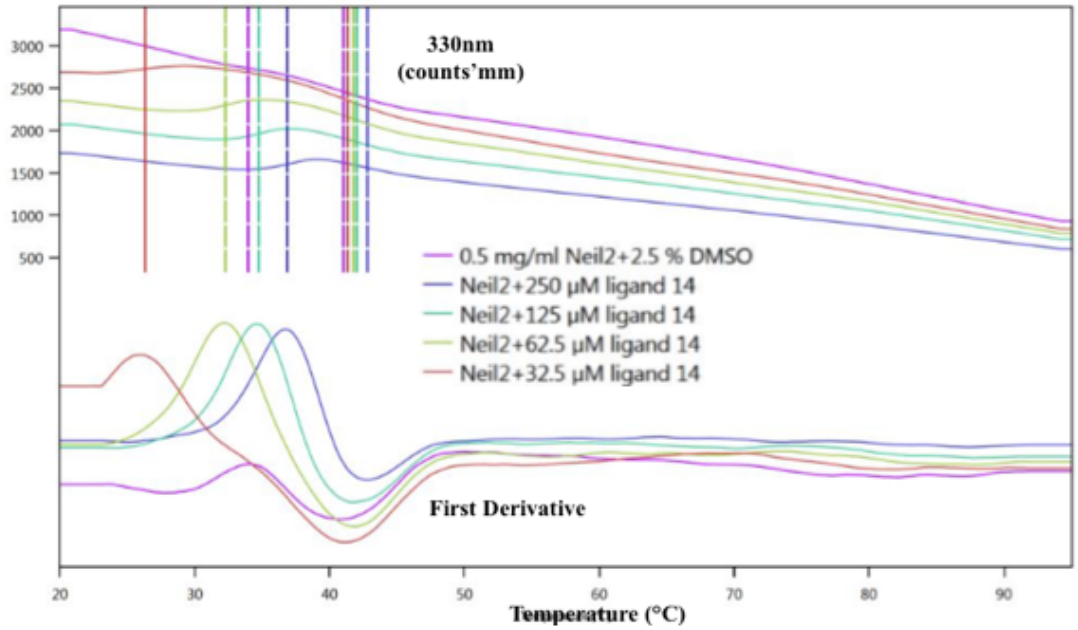


Figure 3.11 Melting scan and first derivate of hNeil2 with and without ligand 14. *The upper plot is the fluorescence counts at 330nm while the lower is the first derivatives of the upper curves. The different melting temperatures of hNeil2 can be read at the highest points of each peak.*

3.6 MST

Based on the results from nanoDSF experiment, the same eight ligands were chosen to be analysed with MicroScale Thermophoresis (MST) with full-length hNeil2. The Neil2 used for MST analysis was from the ion exchange chromatography purification.

3.6.1 Pre-scan of ligands and full-length hNeil2 on label-free MST

Similar to the nanoDSF, the label-free MST is also a dye-free, tryptophan based method. In order to make sure the target ligands do not show fluorescence that affects the output signal, a pre-scan of all the ligands without hNeil2 was performed. Except for ligand 3, the other ligands are all below the threshold. Unlike the dual-UV laser in a nanoDSF instrument, the label-free MST uses an infrared laser. Before starting a complete MST experiment, the optimal concentration of hNeil2 in PBS buffer was determined by a pre-test. **Figure 3.12** display the plot of this pre-test also with ligand 14. All the peaks in the plot are reasonable. The hNeil2 concentration gradient shows that the 2μM is too high for the label-free MST, 1μM is ok but still a little high. So we decide to use 0.5μM hNeil2 for the analysis.

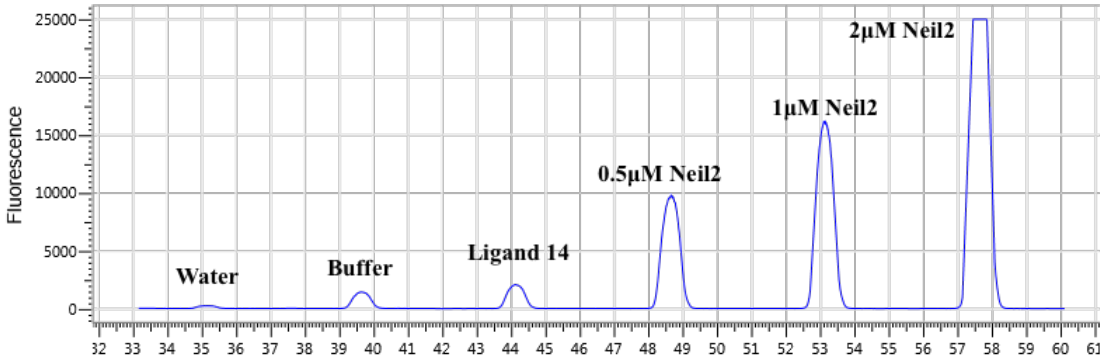


Figure 3.12 The pre-test of hNeil2 with different concentration on label-free MST. *The buffer used in this test was PBS buffer, the ligand 14 concentration was the highest concentration of a dilution series.*

A full label-free MST experiment was run with 500nM hNeil2 and a dilution series of ligand 14. The result was not as expected, the MST time traces were not overlap at all and some of the traces were crossed. Considering the label-free MST depends on the intrinsic tryptophan of the protein, so it mostly relies on the protein condition in the assay buffer, which may lead to the analysis results vary unpredictably and may lead to the unreasonable result. Therefore, no further test was performed on label-free MST.

3.6.2 Screening of ligands and full-length hNeil2 by NT.115 MST

The standard MST requires the target protein to be labeled. The labelling of Neil2 failed the first time PBS buffer and a size exclusion column to separate the protein and the free-dye. The labeled protein solution was supposed to be bluish but it was almost transparent. In the second attempt, a centrifuge filter was used to for the buffer exchange step instead of the buffer exchange column provided by NanoTemper Technology. The assay buffer was changed from

PBS to the one used for purification that has almost the same components as the Hitrap buffer except β-ME. This time, the labelling was successful. The degree of labelling (DOL) needs to be determined before applying the protein to MST analysis. To calculate the DOL, a UV spectrophotometer, NanoDrop, was employed to measure the absorbance at different wavelengths. The absorbance maximum of regular proteins is at 280nm while the maximum absorbance of the blue dye is at 650nm. The highest point of the two peaks in **Figure 3.13** are the absorbances at these two wavelength. The specific readings are listed in **Table 3.1**. There were four parallels in one measurement.

Table 3.1: The absorbances of labeled hNeil2 at 280nm and 650nm.

Well	280nm	650nm
A2	0,066	0,119
A3	0,064	0,115
B2	0,078	0,13
B3	0,074	0,125

The equation used for calculating hNeil2 concentration was:

$$c_{Prot} = \frac{A_{Prot}}{\varepsilon_{Prot} d} = \frac{A_{280} - A_{max} CF}{\varepsilon_{Prot} d}$$

And the equation for calculating DOL was:

$$DOL = \frac{c_{Dye}}{c_{Prot}} = \frac{A_{max}/\varepsilon_{max}}{A_{Prot}/\varepsilon_{Prot}} = \frac{A_{max}\varepsilon_{Prot}}{(A_{280} - A_{max} CF)\varepsilon_{max}}$$

A_{280} and A_{max} are the total absorbance at 280nm and 650nm. ε_{prot} and ε_{max} are the specific extinction coefficients of hNeil2 and the dye. $\varepsilon_{Neil2} = 53315 M^{-1}cm^{-1}$, $\varepsilon_{dye} = 250\,000 M^{-1}cm^{-1}$. CF is the correction factor of the dye which is 0.027. *d* is the path length of the spectrophotometer which is 0.05cm here. Calculation by the above equation indicated that 50% of the Neil2 was labeled.

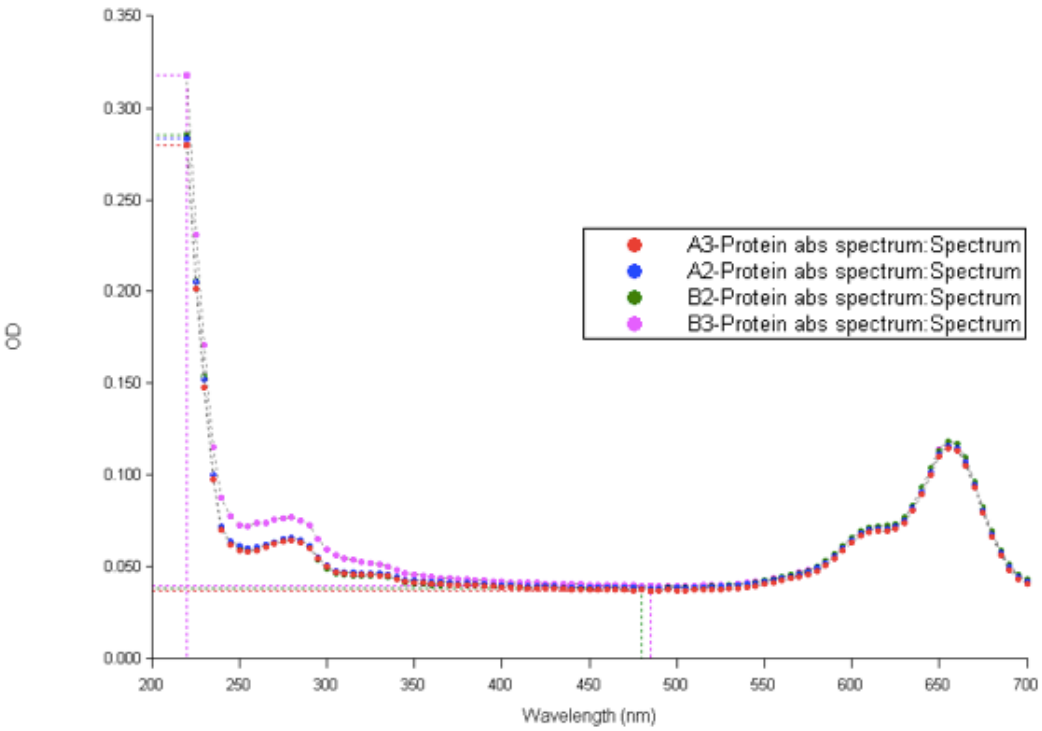


Figure 3.13: The absorbances of the labeled hNeil2 at 280nm and 650nm. *A full UV spectral scan from 220nm-700nm of four parallel hNeil2 samples.*

The standard MST is a very sensitive instrument which can analyse the protein sample at a very low concentration, typically 50nM. So after the labelling, the hNeil2 was diluted to 100nM by buffer H with 0.1% Pluronic F-127 detergent and mixed 1:1 with the 7 ligands in serial dilutions. One full dilution series (from 250µM to 7nM) was collected using standard capillaries. The results of the first experimental series with all the 7 ligands were divergent. The hNeil2 behaved well with the standard capillary since the capillary shape of all the experimental groups were symmetrical and overlapping. Except for ligand 20 and 34, the initial fluorescence intensities were all in the +-10% range which is a prerequisite for the MST analysis. The initial fluorescence intensities of ligand 20 and 34 varied in a large range (**Figure 3.14 Upper**). However, no significant binding was observed except for two ligands (ligand 14 and 43) that had binding potential. These were selected to be analysed once more with a higher ligand concentration.

Three experimental repeats for ligand 14 and 43 were collected and with the fluorescence signal was plotted against the ligand concentration to obtain a dose-response curve. The ligand binding affinity can be deduced from this curve. The dissociation constant (K_d), which is an equilibrium constant, can be estimated from the steepest point of the dose-response curve. In this case, the K_d value represents the concentration of the unbound ligand at which half of the protein molecules are associated with the ligand. The experiment series of ligand 14 and 43 was analysed by MST Analysis mode in the MO.Affinity Analysis programme. But the programme failed to give any K_d value since the difference of the fluorescence signal between the high and low ligand concentration was rather small. Normally, a dose-response curve analysed by MST Analysis should have more than 5 responses units amplitude between the bound and unbound

state. This suggests there was almost no binding of these two ligands with hNeil2.

In the case of ligand 20 and 34, a SD-test was performed to rule out any material loss caused by nonspecific adsorption to capillary walls or due to protein aggregation. The result of the SD-test is shown in **Figure 3.14 Lower**, the fluorescence intensities of samples with the highest and lowest ligand concentration are at the same level. The result indicates that the large difference of the initial fluorescence change was because of the presence of the ligand. Both of ligands 20 and 34 had the same SD-test result which suggested the ligands have the ability to bind hNeil2. Thus, these two ligands were also reanalysed with even higher concentrations.

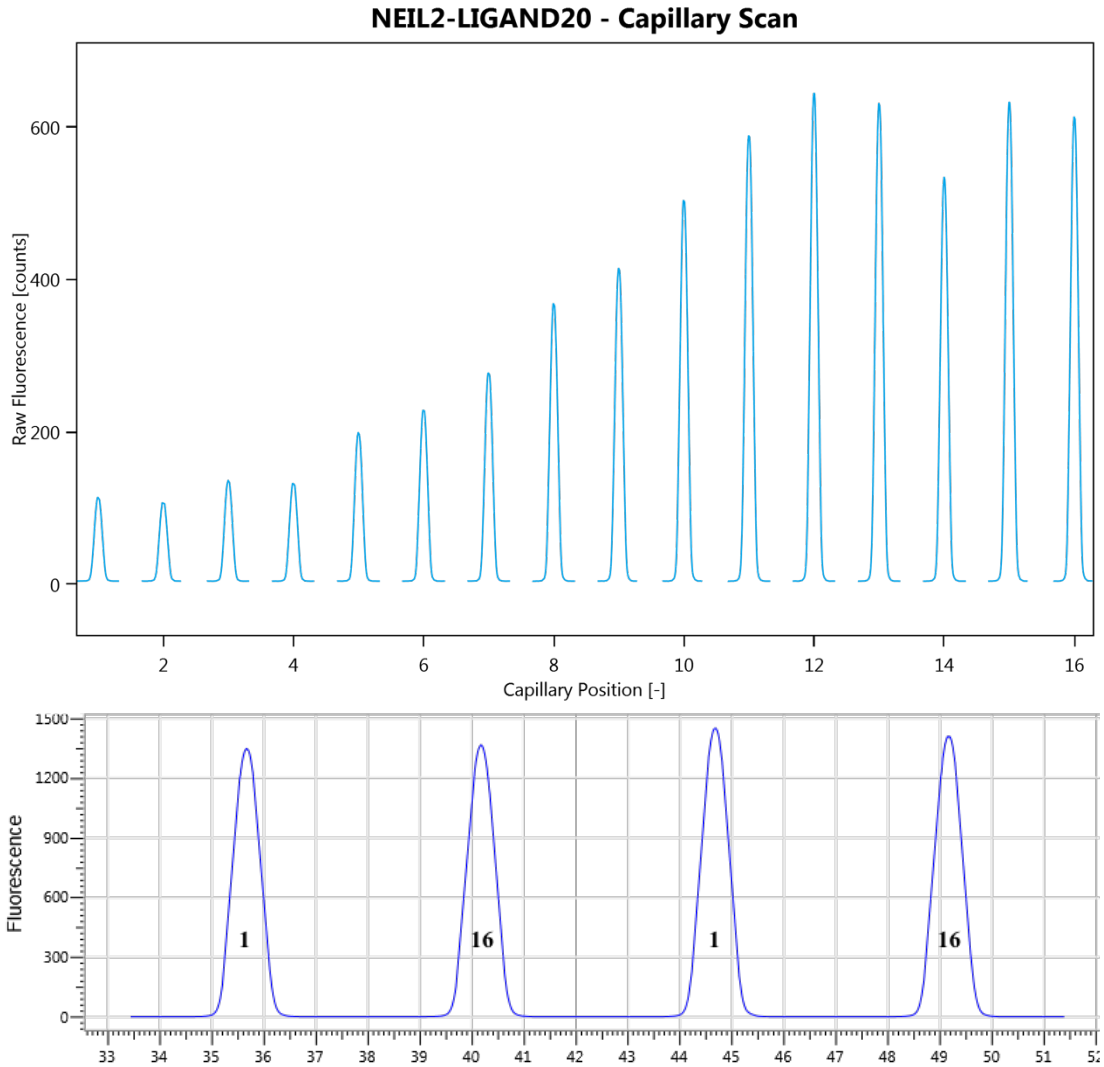


Figure 3.14 Upper: The capillary scan of hNeil2 titrated with ligand20. *The concentration of ligand 20 decreased sequentially from left to right (from 250μM to 7nM). The low fluorescence counts for the ligand concentration samples 1-10 might be because of material loss or the ligand-induced fluorescence change. This can be determined by a SD-test.* ***Lower: The SD-test of ligand 20 with two parallels.*** *The number '1' and '16' denotes the capillary number of the ligand dilution series. '1' was the highest concentration while '16' was the lowest. After the SD-test, all the peaks were almost at the same height which indicates the initial fluorescence change was due to binding of the ligand and not material loss due to protein precipitation.*

The experiment series of ligand 20 and 34 were analysed by the Initial Fluorescence Analysis mode in the MO.Affinity Analysis software since there were ligand-induced fluorescence changes. The dose-response curves of ligand 20 and 34 was shown in **Figure 3.15**. Based on the analysis by the software, the K_d values of ligand 20 and 34 are 13 +/- 2.5μM and 92 +/- 40μM, respectively. Obviously, ligand 20 is the best ligand among those have been tested on the MST instrument.

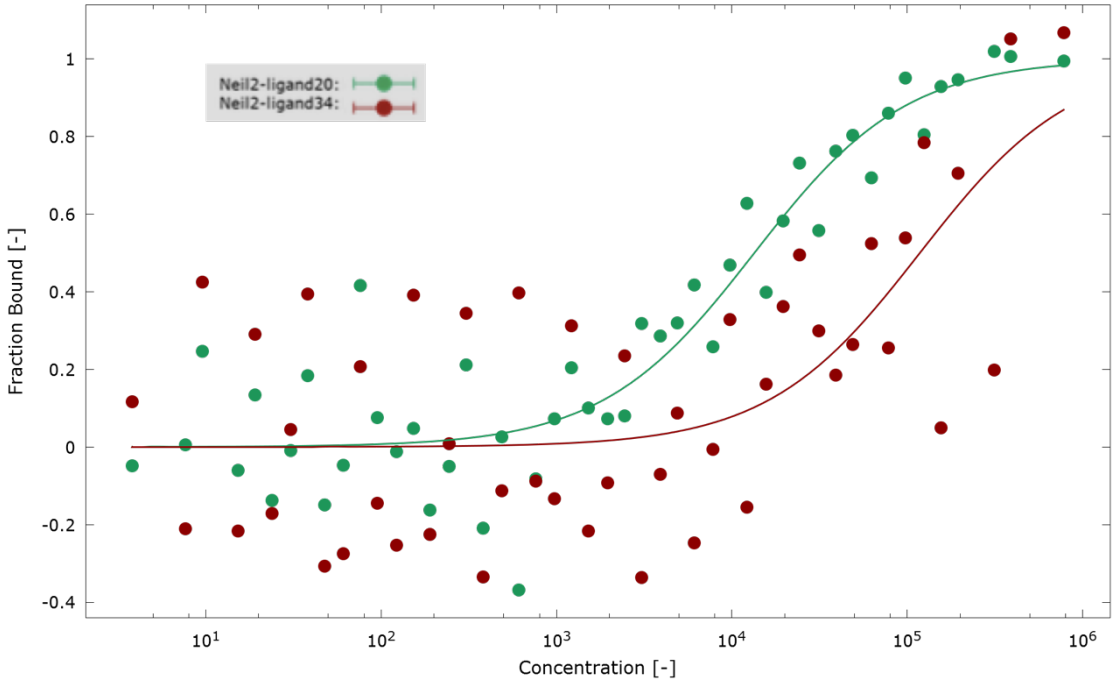


Figure 3.15 The binding curve at 20% MST power of ligand 20 and 34. *The plot contains all the three experiments. Since the ligand concentration ranges for the three experiments were different, these are regression curves based on every single concentration point. The binding affinity of the ligand and Neil2 increases as the increasing of ligand concentration.*

The Neil1 inhibitor (ligand 40) discovered by You (You, 2014) was also tested on MST. The capillary scan indicated the same situation as ligand 20 and 34. However, the SD-test result of ligand 40 still had a huge initial fluorescence difference between the highest and lowest concentrations (**Figure 3.16**). suggesting that hNeil2 protein aggregated in the presence of ligand 40. This conjecture was proven in the following experiment by the activity assay (**Section 3.7**).

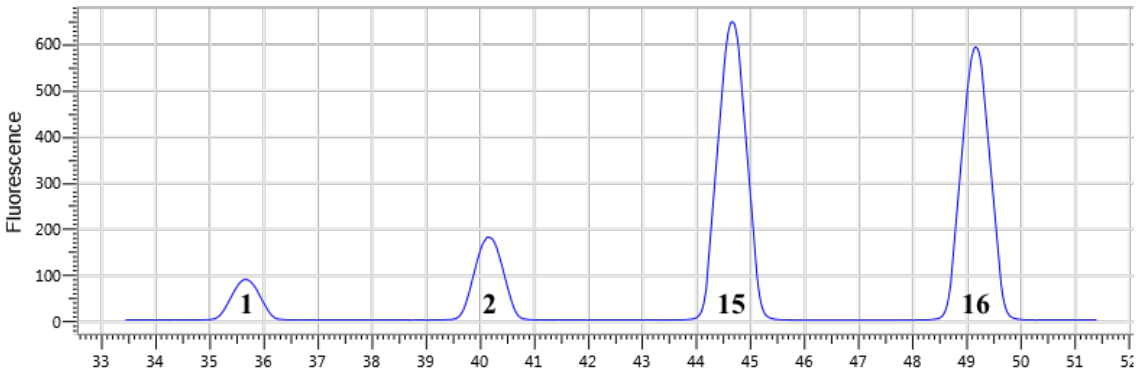


Figure 3.16: The SD-test result of ligand 40. *The numbers in the figure denote to the tube numbers in the sample series. The '1' was with the highest ligand concentration while '16' was the lowest. The huge fluorescence difference between the high and low ligand concentration sample was due to unspecific adsorption. The denaturation of Neil2 proteins by the SD-mix identified the fluorescence loss was probably because of the aggregation.*

Even though the result of ligand 40 SD-test was negative, a full MST experiment was still performed (**Figure 3.17**). The dose-response curve also proved the aggregation of Neil2. The high concentration of ligand 40 leaded to the hNeil2 protein aggregate, so the fluorescence was rather high at these points. As the ligand concentration decreasing, the Neil2 protein was stable in those samples, so the fluorescence signal was very low. This is the reason we can still get a fake binding curve like shown here.

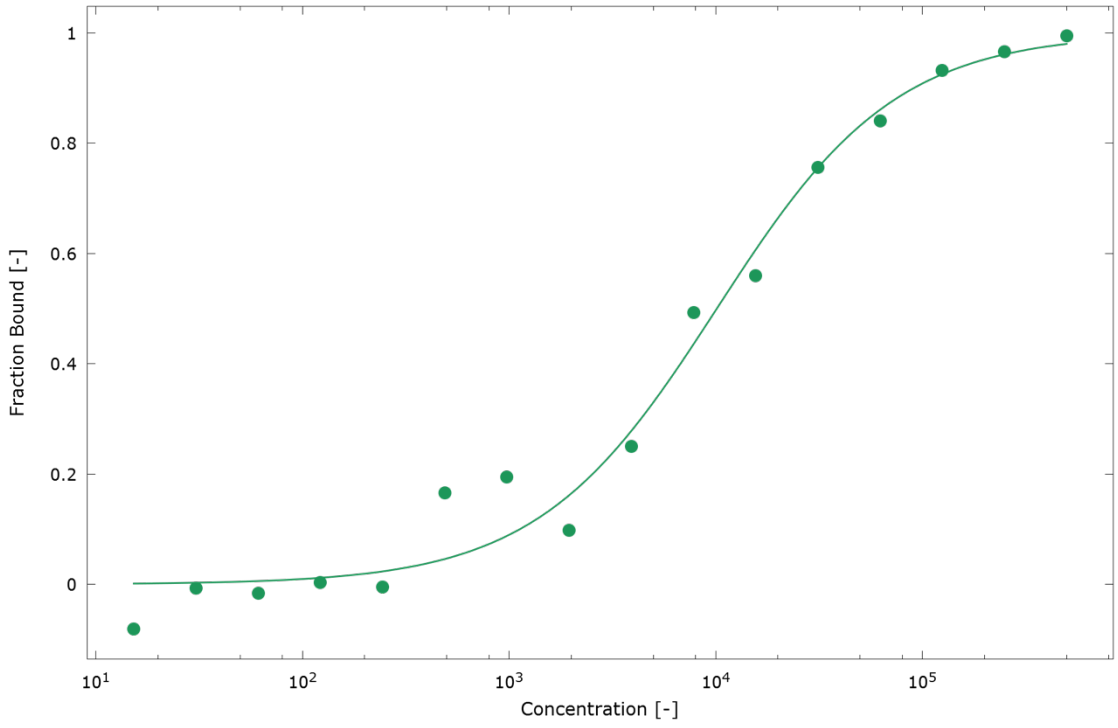


Figure 3.17 The 'binding' curve at 20% MST power of ligand 40.

3.7 Activity assay

Activity assay in this study refers to hNeil2 AP lyase activity assay. UDG was used to create an AP site in the DNA substrate. The DNA substrate was also with a TAMRA reportor, which fluorescence is quenched in a double-strand DNA structure. With the release of TAMRA containing DNA sequence into solution, the fluorescence is no longer quenched.

3.7.1 Expression and purification of full-length wild type and mutant hNeil2 proteins

In the study of hNeil2 by Hazra et al, the Neil2 protein was stable and active after purification by Superdex75 column in Buffer F (Hazra, Kow Yw Fau – Hatahet et al.). Therefore, in this hNeil2 activity assay, we planned to purify all the hNeil2 WT and mutants by use of two affinity chromatography columns as well as a size exclusion chromatography step. However, the yield of all the hNeil2 variants was very low. For mutant C (multiple mutant: P2A + K50A + K51A), almost no protein was left after purification by the ion exchange chromatography. The short induction time of four hours might be responsible for some of the low yield of protein. Thus, we were unable to further purify those proteins on the Superdex75 column. But the WT hNeil2 used for the inhibitors screening was expressed as usual and purified by all three steps.

3.7.2 AP lyase assay of WT and mutants hNeil2

Even though the structure of hNeil2 is not known yet, there are still some conjectures of the possible active site based on the structure of its homologue, hNeil1. In this study, four possible active site residues were selected to be tested using an activity assay. They were Pro2, Glu3, Lys50 and Lys51. All these amino acids were mutated to alanine. The Neil2 mutant variants used in this assay were either with these single mutants or some combinations. The kinetic analysis of the assay is shown in **Figure 3.18**. *E.coli* EndoIV was used as the positive control and completed the reaction within the first ten minutes. The activity of WT hNeil2 shows well an increasing trend, while the activity of all the mutants is at the same level as the background which means the mutants had either extremely low or no activity. The result reveals that all the mutated sites are important for the AP lyase activity of hNeil2. A mutation at any of these sites can lead to loss of Neil2 AP lyase activity.

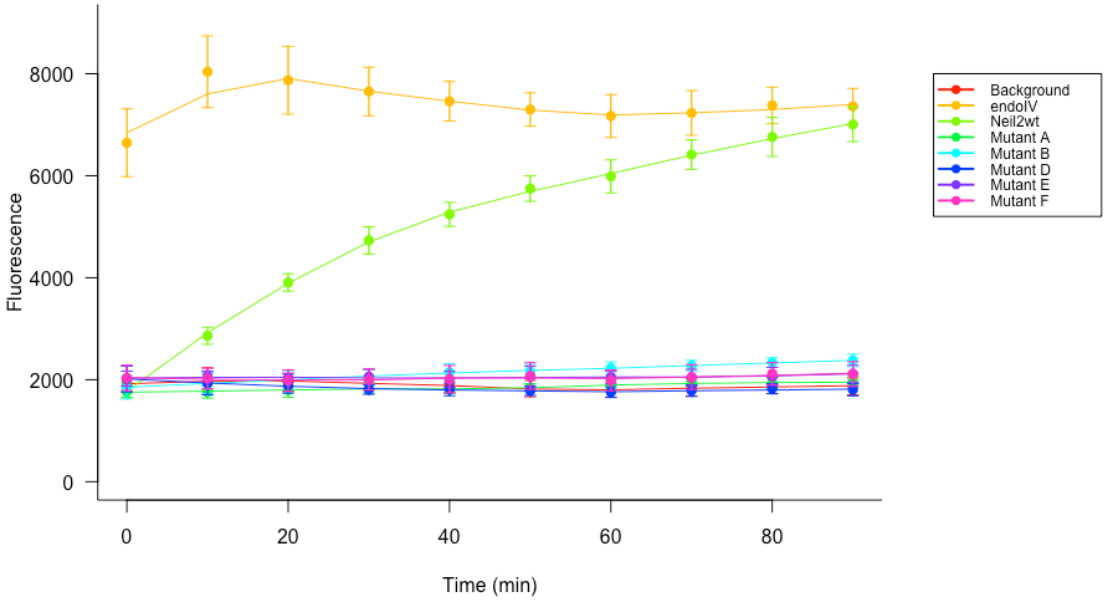


Figure 3.18: Kinetic analysis of AP lyase assay of WT hNeil2 and 5 mutants. *The protein concentration was 500nM and the DNA concentration was 100nM in each reaction. Error bars are the standard deviation of five measurements from each sample. Figure was plotted in R programme.*

3.7.3 AP lyase assay of WT hNeil2 with ligands

In this activity assay for inhibitor screening of WT hNeil2, five ligands were chosen based on the combined results from nanoDSF and MST. A protein ligand is usually a small molecule that can bind to a specific site on a protein. It can inhibit or stimulate the activity of the protein. The assay was performed as for the study of hNeil2 mutants, except for the addition of ligands and 5% (v/v) DMSO. The kinetic analysis of the assay is shown in **Figure 3.19**. Ligand 43 showed a significant stimulation at 50μM, the last few reading points are even higher than the EndoIV control. To verify if ligand 43 is really stimulating the hNeil2 activity, more experiments are needed. Ligand14 is certainly not an inhibitor based on this experiment, since the sample with ligand14 has overlapping curve with the free hNeil2 protein. Except for ligands 14 and 43, the three remaining ligands show significant inhibition. However, in the case of ligand 40 is not that simple. It seems that hNeil2 lost activity totally with the presence of ligand40, which means this ligand is either a strong inhibitor or causes the protein to aggregate. Combined with the result of MST, ligand 40 probably made the hNeil2 aggregating. Finally, the most promising ligands 20 and 34 were chosen to be reanalysed with different ligand concentrations, so that a dose-response curve could be made and the IC50 value of these two ligands could be determined.

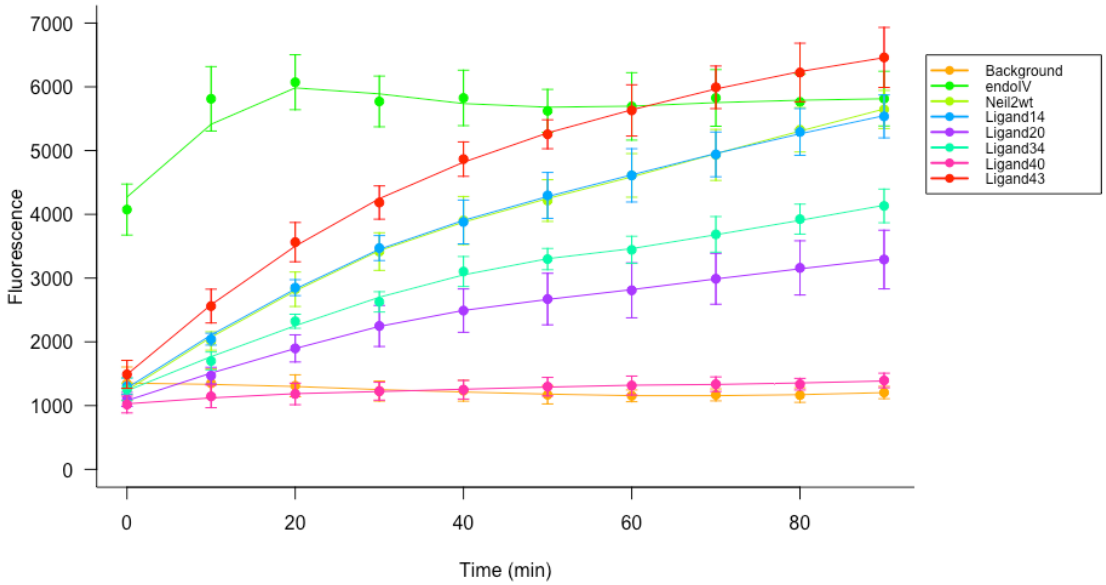


Figure 3.19 Kinetic analysis of AP lyase activity of hNeil2 WT with 5 ligands. *The concentration of WT hNeil2 was 500nM and the concentration of ligands was 50μM in each reaction. Error bars are the standard deviation of seven measurements from each sample. E.coli EndoIV was the positive control. Figure was plotted in R programme.*

Two serial-dilutions of ligands 20 and 34 with WT hNeil2 enzyme were prepared. The enzyme concentration in each reaction was 500nM while the ligand concentrations ranged from 400μM to 3.125μM. The assay was performed as for the study of hNeil2 with 5 ligands, but with a real-time read out of the fluorescence signal after 30 minutes. The readout data were plotted against the ligand concentrations (**Figure 3.20 A**). The dose-response curves of both ligands 20 and 34 show a gradual decrease as the ligand concentration increase. The IC_{50} value represents the concentration of the inhibitor that reduces the protein activity by 50%. **Figure 3.20 B** shows a curve fitting of the dose-response curve, which the y-axis is the enzyme activity and x-axis is the ligand concentration. The IC_{50} value is 80μM according to the figure.

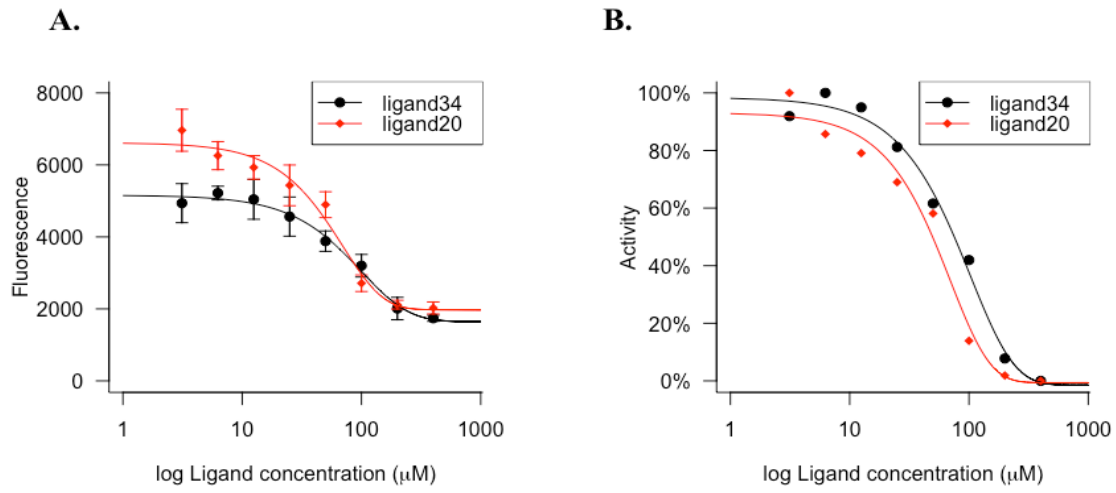


Figure 3.20 A. The dose-response curves of ligands 20 and 34. *Error bars are the standard deviation of seven measurements from each sample.* ***B. The fitting curves of ligands 20 and 34.***

4. Summary and future work

This study presents new and interesting results of human DNA glycosylases Neil2. Those results may lay the foundation for future studies of the biological characterisation of Neil2 protein.

The crystallisation screening for Neil2 in this study indicates the protein is very difficult to crystallise, which has been crystallised the protein alone or in complex with DNA. Comparing the sequences of Neil1 and Neil2, there is a long region in the middle part of Neil2 which is not present in Neil1. This region maybe the reason that Neil2 is harder to crystallise than Neil1. A truncated Neil2 protein with removal of that region can be tried in the future crystallisation. Other strategies include trying other truncations of Neil2, other types of DNA ligands, like oxidized bases in combination with the inactive mutants, e.g. Lys50Ala and Lys51Ala or using DNA with different sizes.

In this thesis, we show that the Pro2, Glu3, Lys50 and Lys51 are all active site residues for the AP lyase activity of Neil2. Since Neil2 is a trifunctional enzyme, it can cleave the DNA backbone by both β and δ elimination. The mutagenesis study in this thesis indicates the β elimination happens before the δ elimination. If the amino acid, which is responsible for β elimination is mutated (Lys50 or Lys51), the enzyme loses its AP lyase activity totally. A future extension of the fluorescent-based assay presented this thesis should include use of a DNA substrate with an oxidized residue (e.g. 5-hydroxy cytosine) to also investigate the role of these residues in the glycosylase activity.

Two small molecule inhibitors of human Neil2 have been found based on the docking results of Neil1, which means the active sites are similar between these two homologues. The K_d values for the two ligands 20 and 34, are 13±2.5μM and 92±40μM, respectively. The IC_{50} value is 80μM for both inhibitors. Only 44 compounds from the Neil1 docking library were tested in this study, so a large library of compounds should also be screened for the Neil2 protein. The present study showed that an ordinary TFA assay does not work for Neil2 because of unspecific binding of the dye. Alternative ways to screen compound libraries using Neil2 include a plate-based activity assay, as presented in this thesis, Surface Plasmon Resonance or nanoDSF, although the latter two are not as high-throughput as the TFA assay. At the moment, fragment based screening using X-ray crystallography is also out of reach due to lack of crystallization conditions for Neil2. Fragment screening by NMR is an alternative, albeit rather slow.

References

Ames, B. N., et al. (1993). "Oxidants, antioxidants, and the degenerative diseases of aging." Proceedings of the National Academy of Sciences of the United States of America 90(17): 7915-7922

Bandaru, Viswanath, Sunkara, Sirisha, Wallace, Susan S., & Bond, Jeffrey P. (2002). A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. *DNA Repair, 1*(7), 517-529. doi: http://dx.doi.org/10.1016/S1568-7864(02)00036-8

Bhakat, Kishor K, Hazra, Tapas K, & Mitra, Sankar. (2004). Acetylation of the human DNA glycosylase NEIL2 and inhibition of its activity. *Nucleic acids research, 32*(10), 3033-3039.

Biological Chemistry. "DNA damage: 8-oxo-guanine." Retrieved May 20, 2016, from http://biologicalchemistry.tumblr.com/post/92485581945/dna-damage-8-oxoguanine-in-its-most-general.

Boiteux, Serge, & Guillet, Marie. (2004). Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA Repair, 3*(1), 1-12. doi: http://dx.doi.org/10.1016/j.dnarep.2003.10.002

Boivin, Stephane, Kozak, Sandra, & Meijers, Rob. (2013). Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification, 91*(2), 192-206. doi: http://dx.doi.org/10.1016/j.pep.2013.08.002

Bornhorst, J A, & Falke, J J. (2000). Purification of proteins using polyhistidine affinity tags. *Methods in enzymology, 326*, 245-254.

Chayen, N. E., & Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods, 5*(2), 147-153. doi: 10.1038/nmeth.f.203

Clancy, Suzanne. (2008). DNA damage & repair: mechanisms for maintaining DNA integrity. *Nature Education, 1(1):103*.

Cooke, Marcus S, Evans, Mark D, Dizdaroglu, Miral, & Lunec, Joseph. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal, 17*(10), 1195-1214.

Das, Aditi, Rajagopalan, Lavanya, Mathura, Venkatarajan S, Rigby, Samuel J, Mitra, Sankar, & Hazra, Tapas K. (2004). Identification of a zinc finger domain in the human NEIL2 (Nei-like-2) protein. *Journal of Biological Chemistry, 279*(45), 47132-47138.

De Bont, Rinne, & Van Larebeke, Nik. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis, 19*(3), 169-185.

Donley, Nathan, Jaruga, Pawel, Coskun, Erdem, Dizdaroglu, Miral, McCullough, Amanda K, & Lloyd, R Stephen. (2015). Small Molecule Inhibitors of 8-Oxoguanine DNA Glycosylase-1 (OGG1). *ACS chemical biology, 10*(10), 2334-2343.

Dou, Hong, Mitra, Sankar, & Hazra, Tapas K. (2003). Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *The Journal of biological chemistry, 278*, 49679-49684. doi: 10.1074/jbc.M308658200

Doublié, Sylvie, Bandaru, Viswanath, Bond, Jeffrey P, & Wallace, Susan S. (2004). The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proceedings of the National Academy of Sciences of the United States of America, 101*(28), 10284-10289.

EPSIM 3D/JF Santarelli, Synchrotron Soleil. (2005). General diagram of Synchrotron Soleil. Lastet ned fra https://commons.wikimedia.org/wiki/File:Schéma_de_principe_du_synchrotron.jpg

Franka J. Rang, Johannes Boonstra. (2014). Causes and Consequences of Age-Related Changes in DNA Methylation: A Role for ROS? *biology, 3(2)*, 403-425. doi: 10.3390/biology3020403

GeneTex. DNA repair mechanisms *New sampler kits for studying DNA repairs*.

Grin, I. R., & Zharkov, D. O. (2011). Eukaryotic endonuclease VIII-Like proteins: New components of the base excision DNA repair system. *Biochemistry (Moscow), 76*(1), 80-93. doi: 10.1134/s000629791101010x

Hans E. Krokana, Pål Sætroma, Per Arne Aasa, Henrik Sahlin Pettersena, Bodil Kavlia, Geir Slupphaug. (2014). Error-free versus mutagenic processing of genomic uracil-Relevance to cancer. *DNA Repair*. doi: 10.1016/j.dnarep.2014.03.028

Hazra, Tapas K, Kow, Yoke W, Hatahet, Zafar, Imhoff, Barry, Boldogh, Istvan, Mokkapati, Sanath K, Mitra, Sankar, & Izumi, Tadahide. (2002). Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *Journal of Biological Chemistry, 277*(34), 30417-30420.

Hegde, Muralidhar L., Hazra, Tapas K., & Mitra, Sankar. (2008). Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res, 18*(1), 27-47.

Hoeijmakers, Jan HJ. (2001). Genome maintenance mechanisms for preventing cancer. *nature, 411*(6835), 366-374.

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica, 16*(9), 907-916. doi: doi:10.1107/S0365110X63002437

Huhn, D, Bolck, Hella A, & Sartori, Alessandro A. (2013). Targeting DNA double-strand break signalling and repair: recent advances in cancer therapy. *Swiss Med Wkly, 143*, w13837.

J.K. Horton, M. Watson, D.F. Stefanick, D.T. Shaughnessy, J.A. Taylor, S.H. Wilson. (2008). XRCC1 and DNA polymerase beta in cellular protection against cytotoxic DNA single-strand breaks. *Cell Res., 18*, 48-63.

Jerabek-Willemsen, Moran, André, Timon, Wanner, Randy, Roth, Heide Marie, Duhr, Stefan, Baaske, Philipp, & Breitsprecher, Dennis. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure, 1077*, 101-113. doi: 10.1016/j.molstruc.2014.03.009

Jordan, Elfrieda. (2015). Crystal Diffraction pattern Electron density map Atomic structure model X-ray Phasing Model building Macromolecular crystallography. Lastet ned fra http://slideplayer.com/slide/9238918/

Krebs, Jocelyn E, Lewin, Benjamin, Goldstein, Elliott S, & Kilpatrick, Stephen T. (2013). *Lewin's genes XI*: Jones & Bartlett Publishers.

Krokan, Hans E, & Bjørås, Magnar. (2013). Base excision repair. *Cold Spring Harbor perspectives in biology, 5*(4), a012583.

Krokeide, Silje Z, Laerdahl, Jon K, Salah, Medya, Luna, Luisa, Cederkvist, F Henning, Fleming, Aaron M, Burrows, Cynthia J, Dalhus, Bjørn, & Bjørås, Magnar. (2013). Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. *DNA repair, 12*(12), 1159-1164.

Kuzminov, Andrei. (1999). Recombinational Repair of DNA Damage inEscherichia coli and Bacteriophage λ. *Microbiology and molecular biology reviews, 63*(4), 751-813.

Laboratory, Argonne National. (2015). Fluorescence-based Thermal Shift Assay curve. I Thermal Shift Assay.png (Red.).

Le May, N., Egly, J. M., & Coin, F. (2010). True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair. *Journal of nucleic acids*, *2010*.

Li, Xuan, & Heyer, Wolf-Dietrich. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res, 18*(1), 99-113.

Lindahl, Tomas. (1974). An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proceedings of the National Academy of Sciences, 71*(9), 3649-3653.

Lord, C. J., & Ashworth, A. (2012). The DNA damage response and cancer therapy. *Nature*, *481*(7381), 287-294.

Martin, L., Schwarz, S., & Breitsprecher, D. Analyzing Thermal Unfolding of Proteins: The Prometheus NT.48. Application Note NT-PR-001

Meagher, Aaron J, Whyte, David, Banhart, John, Hutzler, Stefan, Weaire, Denis, & García-Moreno, Francisco. (2015). Slow crystallisation of a monodisperse foam stabilised against coarsening. *Soft matter, 11*(23), 4710-4716.

Mittal, M. et al. 2014). "Reactive Oxygen Species in Inflammation and Tissue Injury." Antioxidants & Redox Signaling 20(7): 1126-1167.

Moore, J Kent, & Haber, James E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and cellular biology, 16*(5), 2164-2173.

Morland, Ingrid, Luna, Luisa, Gustad, Ellen, Seeberg, Erling, & Bjørås, Magnar. (2005). Product inhibition and magnesium modulate the dual reaction mode of hOgg1. *DNA repair, 4*(3), 381-387.

NanoTemper. (2016). Prometheus NT.48 Report. Department of Biosciences, Centre for Immune Regulation/Rikshospitalet

Prakash, Aishwarya, Eckenroth, Brian E., Averill, April M., Imamura, Kayo, Wallace, Susan S., & Doublié, Sylvie. (2013). Structural investigation of a viral ortholog of human NEIL2/3 DNA glycosylases. *DNA Repair, 12*(12), 1062-1071. doi: http://dx.doi.org/10.1016/j.dnarep.2013.09.004

. Prometheus NT.48 Report. (2016). Oslo, Norway: Department of Biosciences, Centre for Immune Regulation/ Rigshospitalet.

Rhodes, Gale. (2006). Chapter 3 - Protein Crystals. I *Crystallography Made Crystal Clear (Third Edition)* (s. 31-47). Burlington: Academic Press.

Sancar, Aziz, Lindsey-Boltz, Laura A, Ünsal-Kaçmaz, Keziban, & Linn, Stuart. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry, 73*(1), 39-85.

ScanTech. (2014). DNA Damage Responses. I DNA Damage Responses (Red.), *DNA Damage in the Cell and Mechanisms of Cancer: A Quick Overview Part II*.

Smyth, M. S., & Martin, J. H. J. (2000). x Ray crystallography. *Molecular Pathology, 53*(1), 8-14.

Szostak, Jack W., Orr-Weaver, Terry L., Rothstein, Rodney J., & Stahl, Franklin W. (1983). The double-strand-break repair model for recombination. *Cell, 33*(1), 25-35. doi: http://dx.doi.org/10.1016/0092-8674(83)90331-8

T. W. Green, P. G. M. Wuts. (1999). Protective Groups in Organic Synthesis. Lastet ned fra http://www.organic-chemistry.org/protectivegroups

Takao, Masashi, Oohata, Yoshitsugu, Kitadokoro, Kengo, Kobayashi, Kumiko, Iwai, Shigenori, Yasui, Akira, Yonei, Shuji, & Zhang, Qiu-Mei. (2009). Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in Escherichia coli mutant. *Genes to Cells, 14*(2), 261-270.

Tell, G., & Wilson, D.M. (2010). "Targeting DNA repair proteins for cancer treatment.". Cellular and molecular life sciences: CMLS **67**(21):3569-3572

Tentori, Lucio, Orlando, Laura, Lacal, Pedro Miguel, Benincasa, Elena, Faraoni, Isabella, Bonmassar, Enzo, D'Atri, Stefania, & Graziani, Grazia. (1997). Inhibition of O6-alkylguanine DNA-alkyltransferase or poly (ADP-ribose) polymerase increases susceptibility of leukemic cells to apoptosis induced by temozolomide. *Molecular pharmacology, 52*(2), 249-258.

UniProt. http://www.uniprot.org/.

Vinson, Paige. (2006). Automation of Protein Crystallography - Successful Identification of Structure Dependent on Producing High Quality Crystals. *GEN, Vol.26*. http://www.genengnews.com/gen-articles/automation-of-protein-crystallography/1855/

Wallace, S.S., et al. (2012). "Base Excision Repair and Cancer." Cancer letters **327**(1-2): 78-89

Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr Purif, 80*(2), 283-293. doi: 10.1016/j.pep.2011.08.005

Weterings, Eric, & Chen, David J. (2008). The endless tale of non-homologous end-joining. *Cell Res, 18*(1), 114-124.

Wood, Richard D, Mitchell, Michael, Sgouros, John, & Lindahl, Tomas. (2001). Human DNA repair genes. *Science, 291*(5507), 1284-1289.

Yanan Liu, Nan Liu, Xinhua Ma, Xiaoli Li, Jia Ma, Ya Li, Zhijiang Zhou and Zhixian Gao. (2015). Highly specific detection of thrombin using an aptamer-based suspension array and the interaction analysis via microscale thermophoresis. *Analyst, 140*, 2762-2770. doi: 10.1039/C5AN00081E

You, Panpan. (2014). *Human Neil DNA glycosylases-crystal screening, biochemical characterization and inhibitor screening.* (Master), University of Oslo.

Zharkov, Dmitry O, Rieger, Robert A, Iden, Charles R, & Grollman, Arthur P. (1997). NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg) protein. *Journal of Biological Chemistry, 272*(8), 5335-5341.

Appendix I Recipes of buffers

Buffer	Contents
3x protein cracking buffer	6% (w/v) SDS,15% (w/v) β-ME, 30% (w/v) glycerol and 180mM Tris pH6.8
Buffer A	300mM NaCl, 10mM imidazole, 50mM Tris pH7.5 and 10mM β-ME
Buffer B	300mM NaCl, 50mM imidazole, 50mM Tris pH7.5 and 10mM β-ME
Buffer C	300mM NaCl, 300mM imidazole, 50mM Tris pH7.5 and 10mM β-ME
Buffer D	100mM NaCl, 50mM MES pH6.0 and 10mM β-ME
Buffer E	2M NaCl, 50mM MES pH6.0 and 10mM β-ME
Buffer F	200mM NaCl, 25mM Tris pH7.5, 10% glycerol and 10mM β-ME
Buffer G	150mM NaCl, 25mM Tris pH7.5 and 1mM EDTA pH8.0
Buffer H	700mM NaCl, 50mM MES pH6.0
5xRB	250mM MOPS pH7.5, 25% glycerol,5mM EDTA and 5mM DTT

Appendix II Nucleotide Sequence

The uracil containing DNA substrates for cross-linking

AP: G, 10mer
5’- CTGT(dU)ATCTG -3’
5’- CAGATGACAG -3’

AP: G, 11mer
5’- GCTAC(dU)GATCG -3’
5’- CGATCGGTAGC -3’

AP: G, 12mer
5’- CGGACT(dU)ACGGG -3’
5’- CCCGTGAGTCCG -3’

The THF site containing DNA substrates for AP-site analogue creation

THF: G, 11mer
5’- GCTAC(THF)GATCG -3’
5’- CGATCGGTAGC -3’

THF: A, 13mer
5’- CATGTA(THF)ACTCGT -3’
5’- ACGAGTATACATG -3’

THF: G, 13mer, a
5’- CATGTA(THF)ACTCGT -3’
5’- ACGAGTGTACATG -3’

THF: G, 13mer, b
5’- CGGACT(THF)ACGGGC -3’
5’- GCCCGTGAGTCCG -3’

THF: A, 15mer
5’- TCATGTA(THF)ACTCGTG -3’
5’- CACGAGTATACATGA -3’

THF: G, 15mer, a
5’- TCATGTA(THF)ACTCGTG -3’
5’- CACGAGTGTACATGA -3’

THF: G, 15mer, b
5’- CCGCTAC(THF)GATCGCC -3’
5’- GGCGATCGGTAGCGG -3’

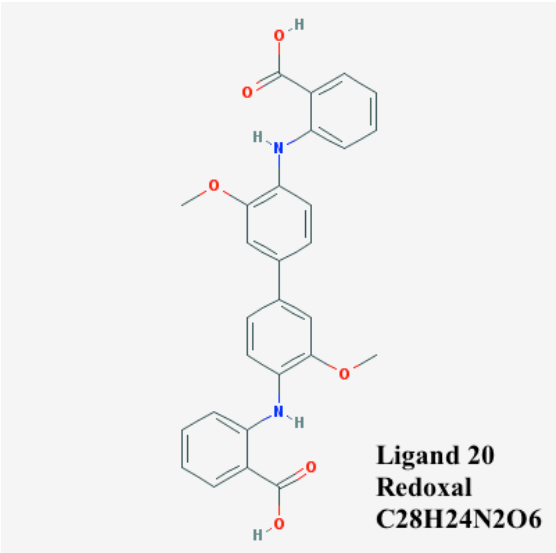
Appendix III Information of crystallisation kits

Name of kits	Purchased company
JCSG+	Molecular Dimensions
Wizard cryo	
PGA	
MembGold	
Index	Hampton Research
Natrix	

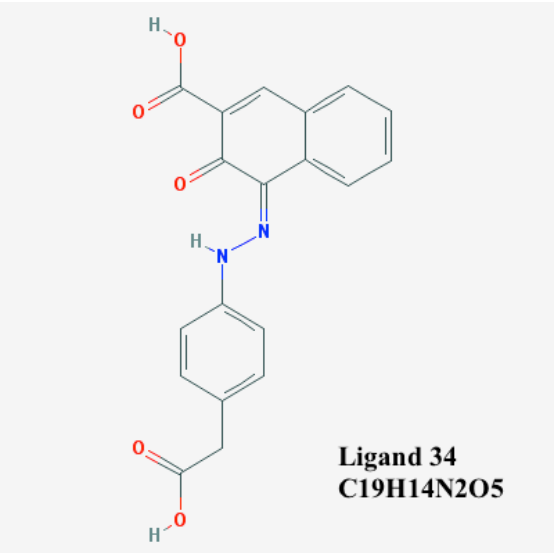
Appendix IV Ligands information

The 44 compounds chosen from the 79 compounds library. The compounds are all from NCI diversity set IV library. Their NSC number are shown in the table.

Ligand NO.	NSC NO.	Ligand NO.	NSC NO.	Ligand NO.	NSC NO.
1	3001	16	60034	31	133071
2	5157	17	60303	32	134199
3	5426	18	60339	33	143491
4	11150	19	60423	34	156563
5	11437	20	73735	35	156565
6	14142	21	81750	36	216623
7	19803	22	91357	37	241624
8	25678	23	91382	38	275266
9	36586	24	91397	39	319994
10	37187	25	91529	40	345674
11	40275	26	95916	41	362639
12	43344	27	106570	42	372499
13	45536	28	111210	43	601359
14	50650	29	126226	44	638080
15	50688	30	128606		



Ligand 20
Redoxal
C28H24N2O6



Ligand 34
C19H14N2O5

Appendix V Experimental data of nanoDSF

First run – full-length hNeil2 with 44 ligands

Capillary position	Content	Tm-1	Tm-2	Tm-3
1	hNeil2	33.7°C	44.2°C	
2	hNeil2	33.8°C	45.5°C	
3	hNeil2+ligand 1	32.8°C	44.2°C	
4	hNeil2+ligand 2	37.4°C		
5	hNeil2+ligand 3	38.2°C		
6	hNeil2+ligand 4	30°C	42.6°C	
7	hNeil2+ligand 5	38°C		
8	hNeil2+ligand 6	32.1°C	44.3°C	
9	hNeil2+ligand 7	37.9°C	45.9°C	
10	hNeil2+ligand 8	37.2°C	46.8°C	
11	hNeil2+ligand 9	36.6°C		
12	hNeil2+ligand 10	38.5°C		
13	hNeil2+ligand 11	34.5°C		
14	hNeil2+ligand 12	Too high fluorescence		
15	hNeil2+ligand 13	37.1°C	57.2°C	
16	hNeil2+ligand 14	39°C		
17	hNeil2+ligand 15	36.9°C	50°C	
18	hNeil2+ligand 16	31.8°C	44.7°C	
19	hNeil2+ligand 17	36.3°C		
20	hNeil2+ligand 18	31.8°C	44°C	
21	hNeil2+ligand 19	32.2°C		
22	hNeil2+ligand 20	38.1°C		
23	hNeil2+ligand 21	31.4°C	44.6°C	
24	hNeil2+ligand 22	37.6°C	46.7°C	
25	hNeil2+ligand 23	33.5°C	43.8°C	
26	hNeil2+ligand 24	31.4°C	45.3°C	
27	hNeil2+ligand 25	37.6°C	45.2°C	
28	hNeil2+ligand 26	37.2°C	44.8°C	
29	hNeil2+ligand 27	31.5°C	44.6°C	
30	hNeil2+ligand 28	38.2°C	45.3°C	
31	hNeil2+ligand 29	33.2°C	43.6°C	
32	hNeil2+ligand 30	35.7°C		
33	hNeil2+ligand 31	32.1°C	43.6°C	
34	hNeil2+ligand 32	36.8°C	44.8°C	
35	hNeil2+ligand 33	33°C	43.7°C	

36	hNeil2+ligand 34	36.3°C		
37	hNeil2+ligand 35	36.3°C		
38	hNeil2+ligand 36	36.4°C	45.8°C	
39	hNeil2+ligand 37	35.7°C	45.8°C	
40	hNeil2+ligand 38	44.9°C		
41	hNeil2+ligand 39	37.3°C	46.8°C	62.9°C
42	hNeil2+ligand 40	36.1°C		
43	hNeil2+ligand 41	32.4°C	43.6°C	
44	hNeil2+ligand 42	36.8°C	46°C	
45	hNeil2+ligand 43	36.6°C		
46	hNeil2+ligand 44	31.7°C	44.3°C	
47	hNeil2	32.9°C	44.5°C	
48	Lysozyme	71°C		

Second run – full-length hNeil2 with 8 ligands

Capillary position	Content	Tm-1	Tm-2
1	1.5mg/mL hNeil2	38.2°C	41.4°C
2	1.5mg/mL hNeil2 +2.5%DMSO	37.5°C	40.7°C
3	1mg/mL hNeil2	35.8°C	41.3°C
4	1mg/mL hNeil2 +2.5%DMSO	35.1°C	40.5°C
5	0.5mg/mL hNeil2	33.3°C	41.4°C
6	0.5mg/mL hNeil2 +2.5%DMSO	33.9°C	41°C
7	hNeil2 + 250μM ligand 3	41.4°C	
8	hNeil2 + 125μM ligand 3	40.2°C	
9	hNeil2 + 62.5μM ligand 3	40.7°C	
10	hNeil2 + 31.25μM ligand 3	40.9°C	
11	hNeil2 + 250μM ligand 14	36.9°C	42.8°C
12	hNeil2 + 125μM ligand 14	34.7°C	42°C
13	hNeil2 + 62.5μM ligand 14	32.2°C	41.8°C
14	hNeil2 + 31.5μM ligand 14	26°C	41.4°C
15	hNeil2 + 250μM ligand 15	39.1°C	55.3°C
16	hNeil2 + 125μM ligand 15	39.3°C	52.4°C
17	hNeil2 + 62.5μM ligand 15	39.2°C	50.2°C
18	hNeil2 + 31.25μM ligand 15	40.2°C	48.5°C
19	hNeil2 + 250μM ligand 19	36°C	
20	hNeil2 + 125μM ligand 19	37.8°C	
21	hNeil2 + 62.5μM ligand 19	38.9°C	

22	hNeil2 + 31.25μM ligand 19	39.7°C	
23	hNeil2 + 250μM ligand 20	40.2°C	
24	hNeil2 + 125μM ligand 20	39.3°C	
25	hNeil2 + 62.5μM ligand 20	39.2°C	
26	hNeil2 + 31.25μM ligand 20	39.7°C	
27	hNeil2 + 250μM ligand 34	39.8°C	
28	hNeil2 + 125μM ligand 34	40.1°C	
29	hNeil2 + 62.5μM ligand 34	40.7°C	
30	hNeil2 + 31.25μM ligand 34	40.8°C	
31	hNeil2 + 250μM ligand 39	40.3°C	
32	hNeil2 + 125μM ligand 39	40.6°C	
33	hNeil2 + 62.5μM ligand 39	40.8°C	
34	hNeil2 + 31.25μM ligand 39	40.6°C	
35	hNeil2 + 250μM ligand 43	33°C	40.8°C
36	hNeil2 + 125μM ligand 43	34°C	41.3°C
37	hNeil2 + 62.5μM ligand 43	33.6°C	41.8°C
38	hNeil2 + 31.25μM ligand 43	33.9°C	41.9°C