Biophysical and biofunctional characterization of extracellular domains N3-N4-N5 and N4-N5 of N-cadherins

Arun Chhikara

MS14156

A dissertation submitted for the partial fulfilment of

BS-MS dual degree in Science

Indian Institute of Science Education and Research Mohali

April 2019

Certificate of Examination

This is to certify that the dissertation titled "**Biophysical and biofunctional characterization of extracellular domains N3-N4-N5 and N4-N5 of N-cadherin**" is submitted by **Mr. Arun Chhikara** (Reg. No. MS14156) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. S.B. Pandit Dr. Rachna Chaba Prof. P. Guptasarma

(Supervisor)

Dated: April 26, 2019

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Prof. Purnananda Guptasarma at the Indian Institute of Science Education and Research Mohali. This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Arun Chhikara
(Candidate)

Dated: April 26, 2019

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Prof. Purnananda Guptasarma
(Supervisor)

Acknowledgement

There are many people that have earned my gratitude for making my time worth in and outside the lab during this dissertation, without whom this dissertation would not have been possible.

First, I am indebted to my supervisor, Prof. Purnananda Guptasarma, not only because he allowed me to work in his lab but also, because of the things he has taught me during the time I have spent in IISER Mohali. Since my third year of BS-MS program when I took one of his courses on protein engineering, I became a big fan of his way of teaching and thinking, and his overall personality became a constant source of inspiration in my life. He continually and convincingly conveyed a spirit of adventure in regard to research although he had a busy schedule. I enjoyed all his talks, great stories and suggestions in the lab meetings profoundly. Under his supervision, I learnt a great deal from him, and therefore I express the deepest gratitude to my teacher.

Besides my advisor, I would like to thank my dissertation committee members Dr. Rachna Chaba who for the first time introduced me to the basic molecular biology when I joined IISER Mohali and Dr. S.B. Pandit who taught me bioinformatics and guided me with his invaluable advice and crucial remarks.

I also extend my gratitude to the senior most people of the lab Dr Rajendra Kumar (my mentor for cadherin work) and Dr Neeraj Dhaunta (my mentor for PfuRd work) for their great mentorship and guidance. Dr Prince Tiwari and Dr Pallavi for their cadherin idea on which I worked. About to become a doctor, Bhisham bhaiya is the coolest and funniest man I have ever met. I still remember the day I entered the lab as a summer intern, he explained to me nothing but the price of all the instrument in the lab, and I feel that the lab holds the jubilant environment mostly because of him and I am thankful for all the laughs, trips and aid I got from him. I want to thank always ready to help, calm and dedicated member of the lab Miss Arpita Mrigwani from cadherin team who taught me many new things during this journey but never admitted herself as my mentor no matter how many times Bhisham bhaiya made fun of it. Arpita Sarkar and Archit Gupta deserve special thanks for roasting, supporting and making moments joyful at the same time. I cannot forget to thank Angel D.

Samuel for the initial company and help with the experiments. Harman and Snehal, I am sorry cannot thank you because you were late and I know you both want my bench, but sorry again you are not getting it so soon. Thank you all for making my time exciting, fun and memorable.

I take pride in acknowledging my family for encouraging me in all of my pursuits and inspiring me to follow my dreams. Thank you, *Maa* and *Paa* for supporting me emotionally and financially.

List of Figures and Photographs

List of Figures and Photographs

List of Tables

Contents

Abstract

Cadherins are the principal cell surface proteins functioning as the biochemical anchor that bind cells together. Other than cell adhesion, cadherins also play a vital role in cell migration, proliferation, differentiation and overall organization of complex neural structures. Based upon the structural, and functional criteria classical cadherin subfamily is divided into two major subgroups: type I and type II. Prominently expressed in the nervous system, N-cadherin is a member of type I cadherin subgroup and is characterized by the presence of five extracellular (EC) domains, followed by a single transmembrane domain and conserved motif for catenin binding in the cytoplasm. Many studies published in the past and recent literature have suggested different structural and functional roles of extracellular domains. We intend to study the biophysical and biofunctional role of multiple combinations of EC domains. Here we describe the structural characterization of the constructs: N3-N4-N5 and N4-N5 of N-cadherin. Rest of the domain combinations are also being explored in parallel. By employing biochemistry, structural bioinformatics and biophysical studies, we aim to gain insights into the mechanical unfoldability, stretch-ability and refolding-ability of individual domains and linker regions, and their calcium-binding characteristics, to understand their modes of function.

Chapter 1

Introduction

1.1] Cell Adhesion Molecules (CAMs)

The ability of cells to hold themselves together is mediated by cell adhesion molecules (CAMs) which fall under the broad family of cell surface proteins. Every cell displays a variety of cell adhesion molecule on its surface and, of course, while adhesion is the primary task allotted to them by the cell, they do much more than that[1] [2]. Adhesion molecules are transmembrane proteins (mostly single transmembrane), and both their intra and extracellular parts possess ligand binding motifs, thus linking the extracellular matrix (ECM) and neighbouring cells to the intracellular cytoskeleton, making cadherins capable of transmitting biochemical and physical signals across the cell[3]. So far, many studies have uncovered the vital functions of cell adhesion molecules in cell migration, cell adhesion, proliferation, differentiation and overall organisation of complex neural structures[4] [5] [6] [1].

Depending on the function and type of interaction with calcium, CAMs can be categorized into two groups: (i) calcium binding-dependent and (ii) calcium binding-independent, CAMs. The cadherins and selectins are two examples of calcium binding-dependent CAMs whereas the immunoglobins and integrins function independent of calcium and hence they are calcium binding-independent CAMs[7].

Based on structure, CAMs are classified into four major classes[8] (Fig. 1.1):

a) Cadherins superfamily: Cadherins forms the largest class of CAMs and thus they substantially govern adhesion and downstream signalling among cells. As cadherins are the focus of this study, a dedicated section (2.1) is provided with a detailed description of their classification, structure and mode(s) of interaction.

b) Immunoglobulin superfamily (Ig-SF): Characterised by the presence of Ig-related folds and repeats, the Immunoglobulin superfamily, is the second major class of cell adhesive molecules, after the cadherins. Ig-SF members can participate in both homophilic and heterophilic interactions. Ig-SF members bind to integrins or the extracellular matrix during the heterophilic mode(s) of interaction[8].

c) Selectins: Calcium-binding dependent selectins are another well-studied class of adhesion molecule which undergo heterophilic interactions. Selectin heterophilic interactions requires a P selectin and its counter protein, PSGL-1, for adhesion. Mainly their interaction provides adhesion in the endothelium and blood cells, and they are not widely distributed like the cadherins and Ig-SF[8].

d) Integrins: Also, well-characterized is a fourth class of cell adhesion molecules which, unlike other classes, forms heterodimers without any requirement of calcium binding. By various α-β combinations, integrins predominantly bind to components of the extracellular matrix such as collagens, laminins, and fibronectin[8].

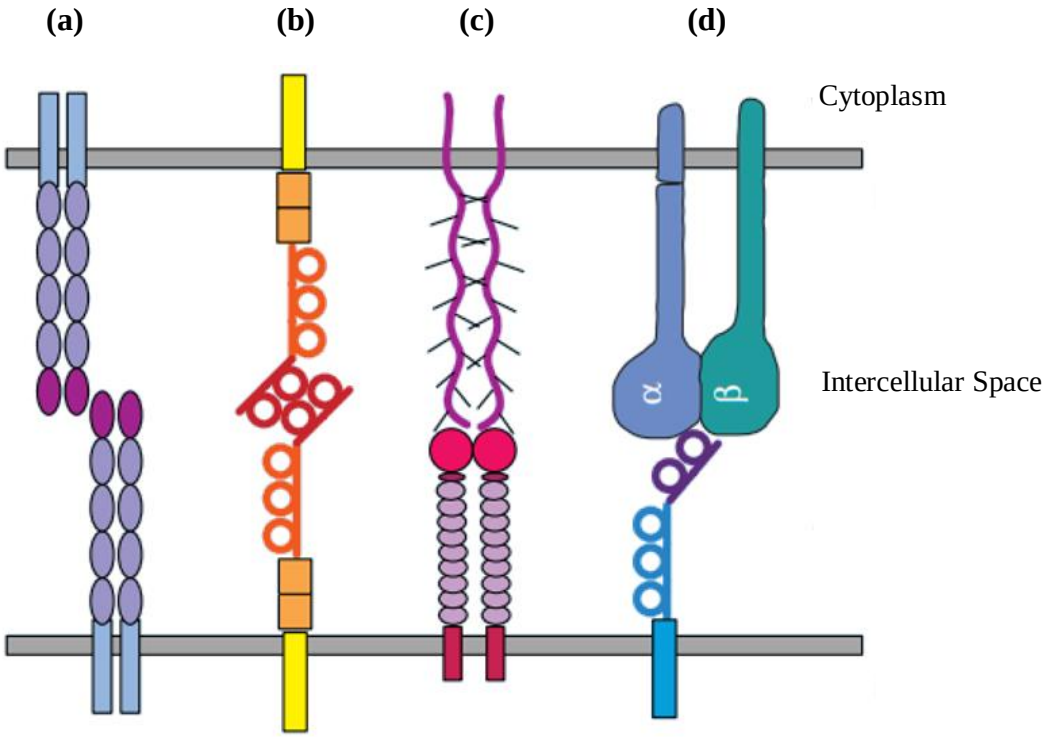


Fig. 1.1 Four major classes of CAMs (a) Cadherins (b) Immunoglobulin (c) Selectins (d) Integrins[8]

1.2] Cadherins

The Cadherins are the major cell-cell adhesion proteins present at cell-cell junctions and they function to bind cells together. The superfamily of cadherins encompasses a wide range of transmembrane proteins, and based on evolutionary, structural, and functional criteria, the superfamily is divided into four major subfamilies[9]: (a) classical cadherins, (b) protocadherins, (c) desmosomal cadherins and (d) atypical

cadherins. Of these subfamilies, classical cadherins have been most extensively studied using biophysical methods[9].

1.2.1] Classical Cadherins

The classical cadherins form a major subgroup of the cadherin superfamily and carry a folded cytoplasmic motif that directly interacts with p130 and β-catenin, and indirectly with α-catenin which, in turn, interacts with the actin cytoskeleton and mediates downstream signalling [7] (Fig. 1.2)

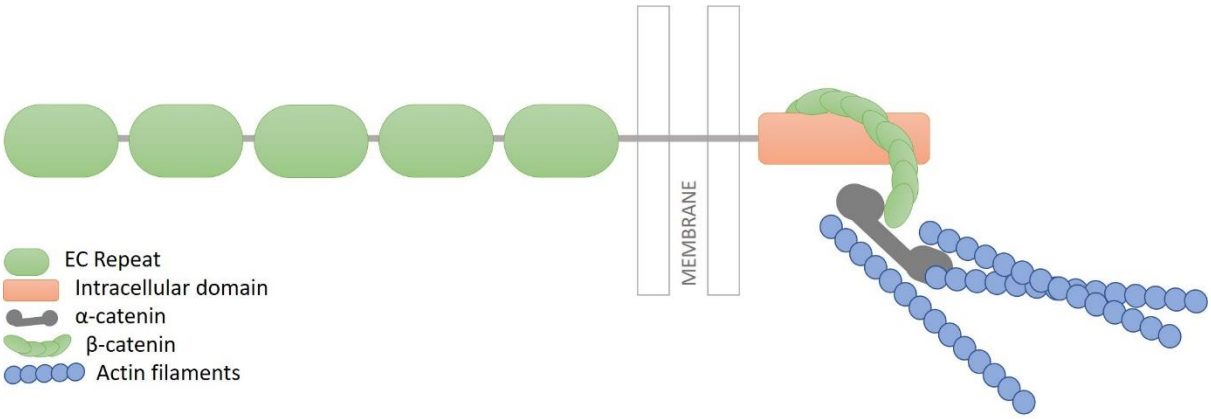


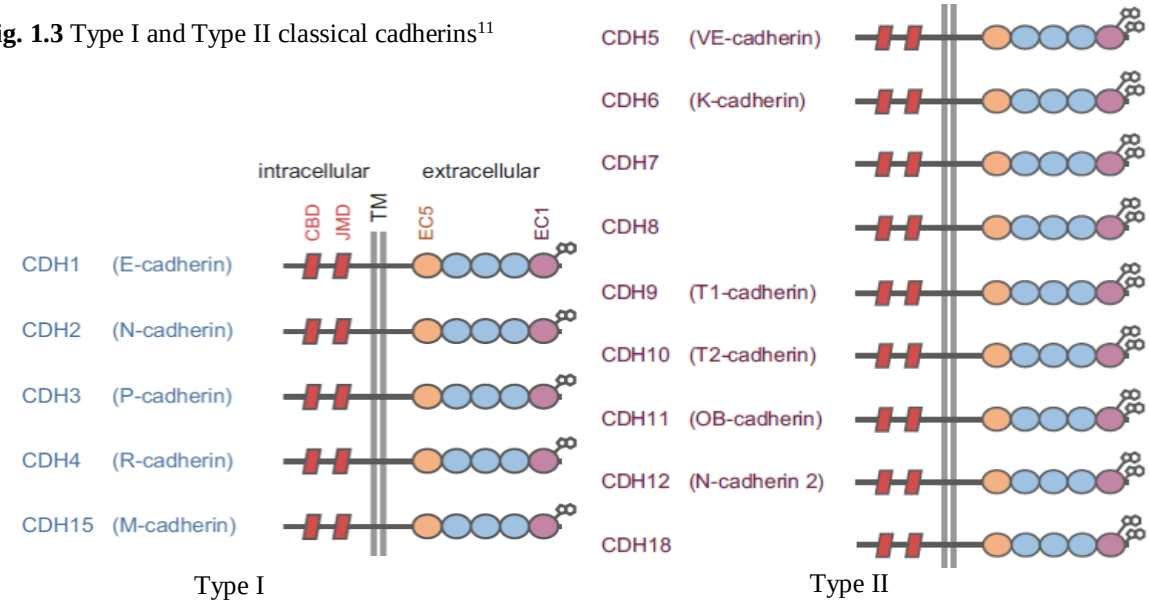
Fig. 1.2 classical cadherins

On the basis of their structure and conserved motifs, the classical cadherins are further classified into subclasses and primarily into type I, II, III and IV. Type I and Type II classical cadherins are characterised by the presence of five extracellular domains that participate in intercellular interaction and govern adhesion[10]. Only vertebrates express type I and type II cadherins, whereas the type III and type IV are found both in invertebrates and other vertebrates, barring mammals[11].

The type I cadherins includes CDH1 (E-cadherin, epithelial), CDH2 (N-cadherin, neuronal), CDH3 (P-cadherin, placental), CDH4 (R-cadherin, retinal) and CDH15 (M-cadherin, myotubule)[11] (Figure 1.3). All these members display a conserved HAV tripeptide motif and a conserved tryptophan in the N-terminal, EC1 extracellular domain, which has been shown to participate in interactions with other cadherins from across the cell-cell adhesion interface. The Type II cadherins, on the other hand, display two conserved tryptophan (fig 1.2) residues in the EC1 domain, and lack the HAV tripeptide motif[12]. So far 13 type II and 5 type I classical cadherins have been identified.

The type III/IV classical cadherin differ from type I/II as they show a variable number of EC repeats and they also possess a conserved PCCD motif called the primitive classical cadherin domain (PCCD) which lies between the cadherin repeats and the transmembrane helix[13].

Fig. 1.3 Type I and Type II classical cadherins[11]



1.2.2] Neural-cadherin (N-cadherin)

Also known as CDH2, N-cadherin is a protein that is encoded by the *CDH2* gene on chromosome 18, in humans[14]. N-cadherin is expressed in multiple tissues and primarily functions to mediate cell-cell adhesion, but recent studies have reported their additional role in cell migration, cell proliferation, differentiation and overall organisation of complex neural structures[1] [4] [5] [6]. N-cadherin, originally named after its discovery from neural tissues, mediates synaptic strength and plays an essential role in learning and memory. Their presence and absence remodel synaptic architecture and thus the strength of a synaptic signal at a given time[15]. Later N-cadherin expression has been observed in other important tissues as well, including cardiac muscles. Recent studies report the vital role played by N-cadherin in cancer metastasis and development[16] [4]. Various types of cadherins are differentially expressed during the process of development, and specifically, N-cadherin has been shown to be a key adhesion molecule that is required for establishment of left-right asymmetry during gastrulation. Also, during early stages of development, N-cadherin promotes myoblast fusion and neurite outgrowth[17].

1.2.3] Structure of N-cadherin

Like other members of class I classical cadherin, N-cadherin is characterised by the presence of five extracellular domains called EC domains or EC repeats followed by a transmembrane region and a highly conserved cytoplasmic tail. The conserved cytoplasmic tail, among class I classical cadherins, possesses a β-catenin binding motif which, in turn, interacts with actin filaments, thus establishing a mechanical link between the fibrils of the ECM and the filaments of the cytoskeleton[11]. Each extracellular domain in N-cadherin has roughly about 110 residues which form an immunoglobulin-like fold[12] as shown in figure 1.4. Each EC domain is composed of seven β-strands arranged as two opposed β-sheets (BED and ACFG fig1.4) with carboxy- and amino-terminus at opposite ends[18]. The folding of each domain is independent of the other. The cadherins are calcium binding-dependent CAMs, and therefore the activity of N-cadherin is dependent on calcium binding.

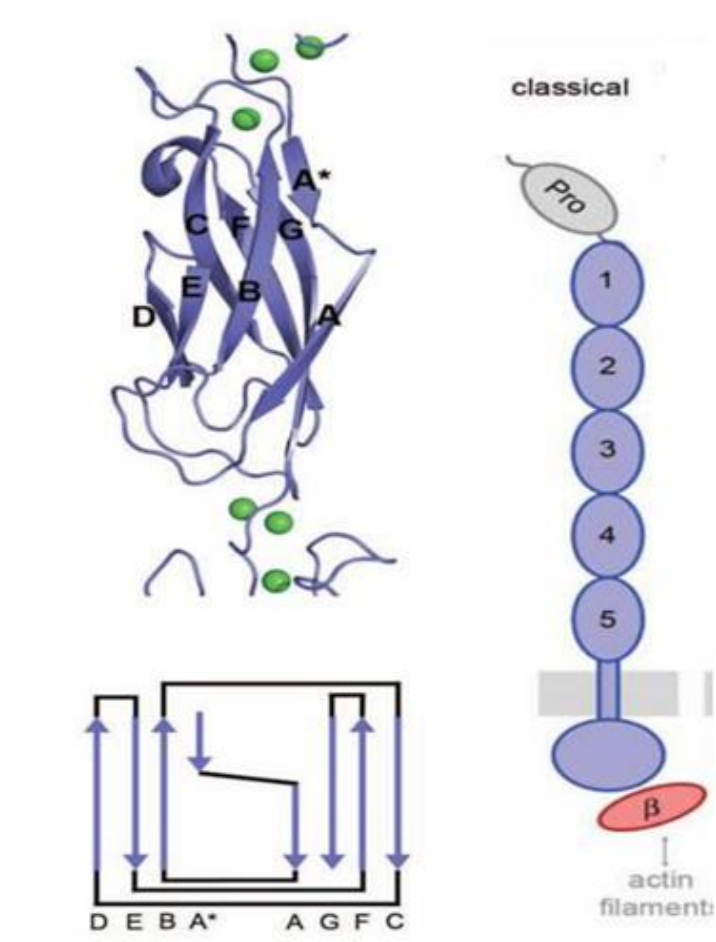


Fig. 1.4 Folding topology of N-cadherin a) EC domain of cadherin showing Ig like fold b) Immunoglobulin β-strand nomenclature c) Full length cadherin structure

1.2.4] Calcium and N-cadherin

The entire cadherin superfamily functions as adhesion entity only in the presence of calcium and so does N-cadherin. The crystal structure of mouse N-cadherin (fig.1.5)

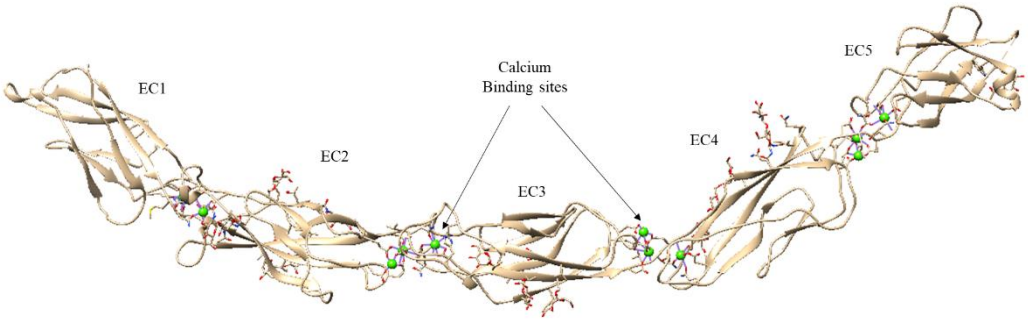


Fig. 1.5 Mouse N-cadherin (CDH2) crystal structure showing Ca^{2+} binding sites

shows the interdomain region where the calcium binds. Each interdomain region has three Ca^{2+} binding sites; therefore, a total of 12 Ca^{2+} ions can bind to four interdomain regions of one cadherin molecule. The calcium binding to cadherins leads to both structural and conformational changes thus making them elongated rigid structures with well-defined β-sheets orientations that confers function, whereas, in the absence of calcium, these domains remains a little disordered[19]. This suggests the importance of calcium in the functional role of cadherins.

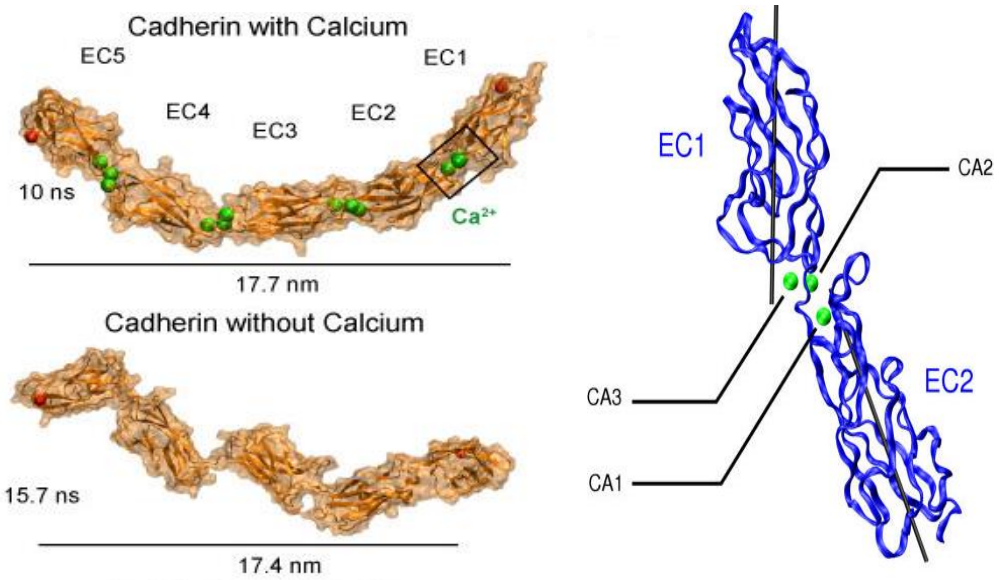


Fig. 1.6 Effect of Calcium on cadherin structure

1.2.5] Mechanism of cadherin interaction

It is well reported that cadherins undergo homophilic interaction and this homophilic interaction can be of two types.

1. The cadherin molecule on the cell surface can interact with another cell surface cadherin molecule from the same cell and thus form so-called *cis* interaction[9].
2. In another scenario, we have cadherin molecules from two different cells interacting and brings two cells in close proximity. This type of interaction is defined as *trans* interaction[9].

Various models have been proposed to explain the mechanism of interaction between cadherin molecules few of them are described below.

Domain swapping model

This model depicts that ectodomains from two opposing cell surfaces interact in a *trans* manner through their most membrane-distal cadherin domains (the EC1 domains). According to this model, swapping of conserved tryptophan during the event takes place, when the tryptophan from one EC1 domain goes and buries itself into the hydrophobic pocket of the partnering EC1 domain[18] and *vice versa* (fig.1.7).

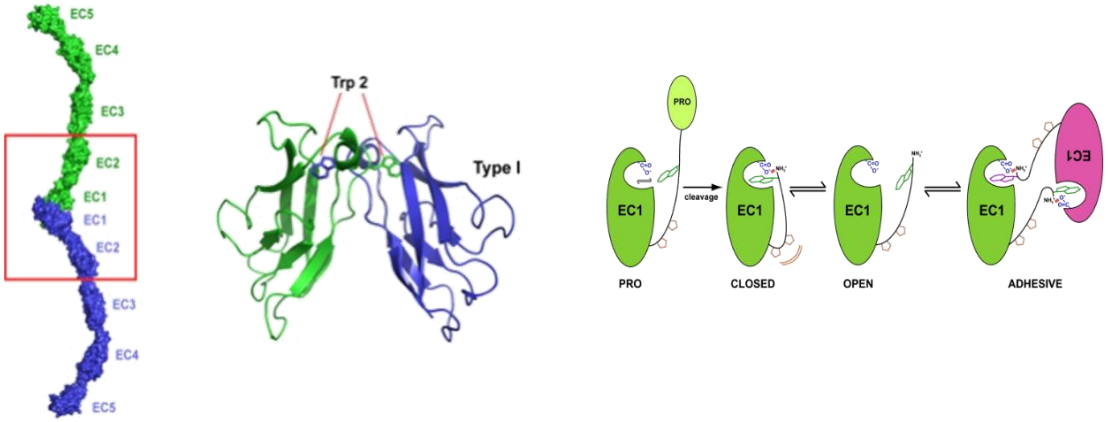


Fig. 1.7 Domain swapping model for *trans* interaction between EC1 domain.

Zipper model

This model says that initially, cadherins form *cis* dimers on the same cell's surface which then interacts with another *cis*-dimer of a different cell, thus making a trans-dimer through the interaction mediated by EC1 domain of cadherin[20] (fig.1.8a).

Sliding model

The sliding model is an extrapolation of the zipper model which states that the *cis* dimerized cadherins from the opposing cells are further involved in the *trans* interactions through all the five extracellular domains, such that EC1 slides all the way from EC1 to EC5 domain of opposite *cis* dimer, thereby pulling the cells together. By the end of sliding, EC1 interacts with EC5 *cis* dimer of another cell, EC2 interacts with EC4, EC3 with EC3 and so on[20] (fig 8.1b).

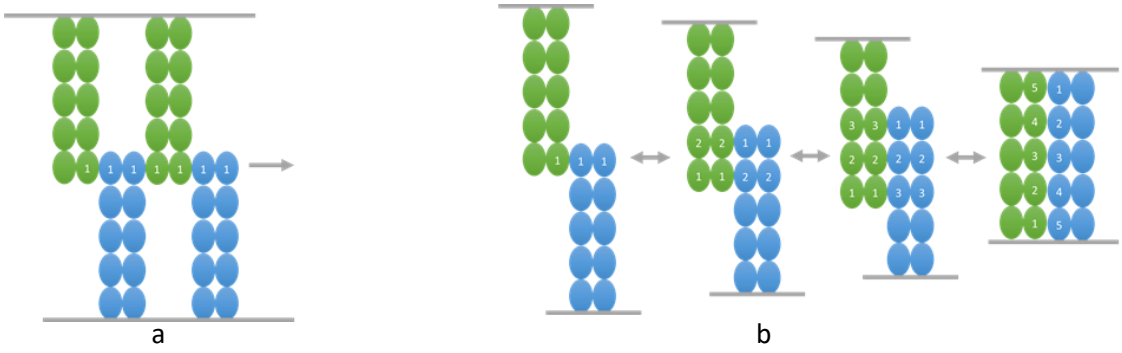


Fig. 1.8 a) Zipper model b) Sliding model

Our approach-

Recently few new models have also been proposed, but uncertainty still stands about the model that best explains the cadherin mode of interaction. Although it has been accepted by most that only the EC1 domain has a role in interaction, the role of other domains is still unclear. Therefore, by looking at the biophysical aspects of each domain, we decided to address the role of all the five domains of N-cadherin by producing and studying them at the molecular level in isolation and in combination.

Here we describe the structural characterization of the constructs: N3-N4-N5 and N4-N5 of N-cadherin. Rest of the domain combinations are also being explored in parallel. By employing biochemistry, structural bioinformatics and biophysical studies, we gained insights about the structure, aggregation and stability of these two and three domain constructs.

Chapter 2

Materials and Methods

2.1] Materials

2.1.1] Basic Components

Lysogeny broth for the growth of bacteria.

Component	Amount for 1 Litre (g)
NaCl	10g
Tryptone/Peptone	10g
Yeast Extract	5g

After mixing all three components in MQ water autoclaving is done in standard conditions (15 psi and 120°C).

Antibiotics

Antibiotic	Concentration of stock (1000X)
Ampicillin	100 mg/mL in water
Chloramphenicol	35mg/mL in methanol
Kanamycin	25mg/mL in water
Tetracycline	12.5mg/mL in 70% ethanol

Antibiotics are sterilised using a syringe driven filter (0.22 µm) and stored as aliquots at -20°C. The working concentration of antibiotics is 1X.

Primers:

Stock concentration	Working concentration
100µM	1µM

dNTPs

Stock concentration	Working concentration
100µM	1µM

Bacterial Strains:

Protein Expressed	Vector used	Strains used
N3-N4-N5	pET23a pQE30	*E. coli* XL1 Blue, *E. coli* BL21(DE3)
N4-N5	pET23a pQE30	*E. coli* XL1 Blue, *E. coli* BL21(DE3)

Plasmids Mini Prep:

Buffer	Composition
Buffer P1 (Resuspension Buffer)	50mM Tris-Cl, pH-8.0, 10mM EDTA, 100µg/ml, RNase A
Buffer P2 (Lysis Buffer)	200mM NaOH, 1% SDS (w/v)
Buffer P3 (Neutralization Buffer)	3.0 M potassium acetate, pH-5.5
Buffer PE (Column Binding/ Wash Buffer)	1.0 M NaCl, 50mM MOPS, pH 7.0, 70% ethanol
Buffer EB (Elution Buffer)	10mM tris, pH 8.0

2.1.2] Buffers and Solutions for recombinant DNA work:

Agarose gel (1%)

Agarose	1g
Deionized Water	100 mL

50X TAE Buffer

Tris-Cl	242g
Glacial acetic acid	57.1 mL
0.5M EDTA (pH 8.0)	100 mL

6X DNA gel loading buffer:

Bromophenol Blue	0.25%
Glycerol	30%

Ethidium bromide stock solution (1% w/v)

Ethidium bromide	0.1g
Deionized Water	10ml

TE Buffer:

Tris-Cl (pH 8.0)	10mM
EDTA	1 mM

2.1.3] Buffers and Solutions for SDS-PAGE:

Acrylamide:

Acrylamide	30 g
N,N'-Methylene - bisacrylamide	0.8 g
Deionized Water	100 mL

Ammonium persulfate (APS, 10%):

APS	100 mg
Deionized Water	1 mL

Lower Tris (4X), pH 8.8

Tris (MM-121)	18.17g (1.5M)
SDS (MM-288.37)	0.4g
Deionized water	100ml

Upper Tris (4X), pH 6.8

Tris (MM-121)	6.06g (0.5M)
SDS (MM-288.37)	0.4g
Deionized water	100ml

5X Sample loading buffer:

Tris-Cl (pH 6.8)	0.15 M
SDS	5%
Glycerol	25%
β-mercaptoethanol	12.5%
Bromophenol blue	0.06
Deionized water	10 mL

Gel Staining Solution:

Methanol	40%
Glacial acetic acid	10%
Coomassie Brilliant Blue R-250	0.1%
Deionized water	50%

Gel Destaining Solution:

Methanol	40%
Glacial acetic acid	10%
Deionized water	50%

SDS-PAGE COMPOSITION:

Resolving Gel

Percent	7%	8%	9%	10%	11%	12%	13%
Lower Tris	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Acrylamide	2.33	2.667	3	3.333	3.667	4	4.333
Water	5.066	4.733	4.4	4.067	3.733	3.4	3.067
APS	50	50	50	50	50	50	50
TEMED	10	10	10	10	10	10	10

Percent	14%	15%	16%	17%	18%
Lower Tris	2.5	2.5	2.5	2.5	2.5
Acrylamide	4.667	5	5.333	5.667	6
Water	2.733	2.4	2.067	1.733	1.4
APS	50	50	50	50	50
TEMED	10	10	10	10	10

Stacking Gel:

Percent	3%	5%
Upper Tris	0.5	0.5
Acrylamide	0.2	0.333
Water	1.3	1.167
APS	12.5	12.5
TEMED	5	5

2.1.4] Buffers and solutions for protein purification:

2.1.4.1] Native purification of 6X His-tagged proteins

Native Lysis Buffer

NaH_2PO_4, pH- 8.0 (MW-119.98)	0.05 M (6.9g in 1L)
NaCl (MW-58.44)	0.3 M (17.53g in 1L)
Imidazole (MW-68.077)	0.01 M (0.68g in 1L)

Native Wash Buffer

NaH_2PO_4, pH- 8.0 (MW-119.98)	0.05 M (6.9g in 1L)
NaCl (MW-58.44)	0.3 M (17.53g in 1L)
Imidazole (MW-68.077)	0.02 M (2 X 0.68g in 1L)

Native Elution Buffer

NaH_2PO_4, pH- 8.0 (MW-119.98)	0.05 M (6.9g in 1L)
NaCl (MM-58.44)	0.3 M (17.53g in 1L)
Imidazole (MM-68.077)	0.25 M (25 X 0.68g in 1L)

2.1.4.2] Denaturing Purification of 6X His-tagged proteins

Buffer B (Lysis buffer) pH = 8

Urea/ GdmCl	8 M
NaH_2PO_4	0.1 M
Tris-Cl	0.01 M

Buffer C (Wash buffer) pH = 6.3

Urea/ GdmCl	8 M
NaH_2PO_4	0.1 M
Tris-Cl	0.01 M

Buffer D (Wash buffer) pH = 5.9

Urea/ GdmCl	8 M
NaH_2PO_4	0.1 M
Tris-Cl	0.01 M

Buffer E (Elution buffer) pH = 4.5

Urea/ GdmCl	8 M
NaH_2PO_4	0.1 M
Tris-Cl	0.01 M

2.2] Methods

2.2.1] Cloning

The cloning of all possible combination of N-cadherin domains were done by previous lab members and therefore the construct: N3-N4-N5 and N4-N5 were already available to work on and were cloned into pET23a vector in our lab.

I was provided with two separate plasmids that had cloned human N-cadherin domain N3-N4-N5 and N4-N5 sequence.

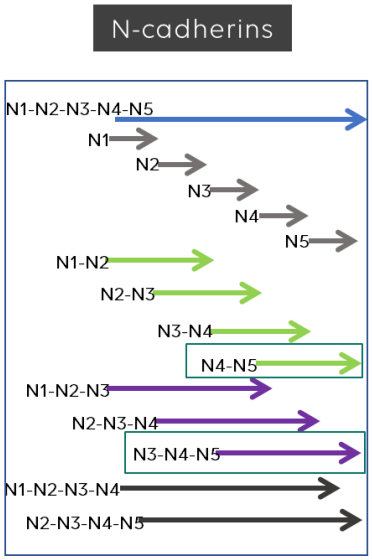


Fig. 2.1 N-cadherin cloning constructs.

2.2.2] Plasmid isolation

1. *E. coli* XL1 Blue cells containing the recombinant plasmids with the insert were grown in 10 mL LB media at 37 °C overnight with shaking. Antibiotics used (Tetracycline 10 µl + Ampicillin 10 µl).

2. The cells were pelleted down into 2 mL MCT by centrifugation 2 min at 8000 rpm. The supernatant was discarded.

3. Then the pellet was resuspended in 200 µl of resuspension buffer stored at 4°C.

4. Followed by addition of 200 µl of lysis buffer and gently mixing by inverting the tube 5-6 times until the solution became clear.

5. Neutralization step involves addition of 350 µl of neutralisation buffer to the MCT that lead to the formation of a white precipitate.

6. Centrifugation at 13000 rpm for 10 minutes is required to separate cell debris from plasmid.

7. The cleared lysate was then carefully transferred to the column without disturbing the pellet and centrifuged for 1 min at 13000 rpm. The flow through was discarded.

8. To wash the column, 500 µl of wash buffer-1 was added and then centrifuged at 13,000 rpm for 1min. The FT was discarded.

9. After this, 750 µl of wash buffer-2 was added and centrifuged at high speed for 1 min. The FT was discarded.

10. To let the residual ethanol completely elute, the column was then given a blank spin for 2 min at high speed. It was then kept for drying at room temperature.

11. The column was then transferred into a fresh 1.5ml MCT and the plasmid was eluted by adding 50 µl of autoclaved MQ water preheated at 50°C and spinning for 1 min at 13,400 rpm.

2.2.3] PCR (Polymerase Chain Reaction)

PCR was used in some of the experiments to amplify DNA segments of interest in large quantities from an initial minute template. Though conditions vary depending upon various factors including length of template or primers, but basically it involves three main steps: Denaturation, Annealing, Extension and the standard protocol for 50 µl reaction is as follows:

Protocol for PCR amplification:

Sr No.	PCR components	Amount in µl
1	10X thermopol buffer	5
2	dNTPs	0.5
3	Forward primer	0.5
4	Reverse primer	0.5
5	Template	1
6	Polymerase (Deep vent)	0.5
7	DW	42
		50

Table 2.1 Composition of PCR reaction.

Standard Program used for PCR reaction:

Sr No.	Steps involved	Time and Temp. for Deep vent
1	Initial Denaturation	98°C, 5 min
2	Denaturation	95°C, 30 sec
3	Annealing	55°C, 30 sec
4	Extension	30 sec/kb
5	Final Extension	10 min

Table 2.2 Steps in PCR cycle.

2.2.4] Restriction digestion

The restriction digestion reaction mixture was set up by adding the following components:

Sr No.	Components	Amount(µl)
1	10X FD Buffer	3 µl
2	DNA	1 µg
3	RE	2 µl
4	DW	x µl
	Total	30 µl

Table 2.3 Restriction digestion reaction.

The products of digestion reaction were run on 1% agarose gel for visualization and confirmation of digestion.

2.2.5] Agarose gel preparation

- 1X TAE buffer was prepared from by 50X TAE.
- To prepare 1% of agarose gel 1g of agarose was weighed and added to 100 ml of 1X TAE.
- The solution was then boiled for 2 min in a microwave.
- It was then allowed to cool down and 2 µl of EtBr was added to it.
- The gel apparatus was assembled and the solution was poured into it.
- The comb was removed when the gel solidified.

2.2.6] Gel extraction

Gel extraction was done mainly following the QIAGEN gel extraction kit.

- DNA fragment of desired size was excised out from the agarose gel using a clean and sharp scalpel.
- Gel slices were transferred into MCT and then weighed.
- Three volumes of QG buffer was added to 1 volume of gel.
- It was then incubated at 50°C for 10 min to completely dissolve the gel pieces.
- 1 gel weight equivalent volume of isopropanol was added to the sample and then mixed.

- A spin column was then taken and placed in the provided 2 ml collection tube.
- To bind the DNA to the spin column, sample mixture was carefully added to the centre of the column and centrifuged at 13,000 rpm for 1 min. The FT was discarded and column was placed back into the same tube.
- To wash the column, 750 µl of wash buffer PE was added to the centre of the column and it was then centrifuged for 1 min at 13,400 rpm. The FT was discarded.
- The column was then given a blank spin for 2 mins at high speed. It was then kept for drying at room temperature.
- Then the column was transferred to a fresh MCT and 30 µl of DW was added to the centre of the spin column.
- After allowing the column to soak the water for 1 min, it was then centrifuged at 13000 rpm to elute the DNA.

2.2.7] Ligation Reaction

Required molar (vector : insert, 1:1 and 1.3 depending on reaction to reaction) concentration of components was calculated using online ligation calculator. Modification and changes in standard ligase reaction is mentioned next to same experiment in result section.

Ligation reaction was set up by adding the following components-

Sr. No.	Components	(µl)
1	10X DNA ligase buffer	1
2	Digested DNA fragment	2
4	DNA ligase	1 (5 Unit)
5	DW	6
	Total	10

Table 2.4 Composition of ligation reaction.

2.2.8] Bacterial Transformation

The transformation of pET23a vector was done in BL21 DE3 cell.

- Competent cells were taken out of - 80 °C and thawed on ice for 15 mins.
- 2 µl of plasmid was added to 50 µl of competent cells. These were then incubated on ice for 30 mins.
- Heat shock was given for 90 sec at 42°C. This was followed by cold shock on ice for 2 min.
- 950 µl of LB media without any antibiotics was added and the MCT was then incubated at 37°C for 1 hour.
- The sample was then centrifuged at 2500 rpm for 5 mins. Most of the supernatant was discarded leaving behind around 150 µl of LB media to resuspend the pellet.
- The transformed cells were then plated on LB-agar plate with appropriate antibiotics. (Tetracycline/Ampicillin).
- The plate was then incubated overnight at 37°C.

2.2.9] Glycerol stock preparation:

500 µl of 1º culture was added to 500 µl of 50% glycerol in a 2 mL cryovial and gently mixed. This will give 25% glycerol to final glycerol stock. The cryovials then kept at -80°C for long term storage.

Glycerol stock in lab- 1) 60% glycerol

Take 1.5ml 1º culture (prepared in LB) and 0.5ml 60% glycerol in 2ml cryovial. Then store in -80ºC. This calculation will give 15% glycerol in final glycerol stock.

2.2.10] Native purification of the protein:

N3-N4-N5 and N4-N5 constructs have a high tendency to form inclusion bodies. So, different conditions were used during optimization of protein expression.

- Primary Culture in 10 ml LB media vial containing appropriate antibodies (Chloramphenicol/Ampicillin 10 µl) was inoculated using glycerol stock of the transformed bacterial cells. It was then incubated overnight at 37°C.
- Secondary culture: 10 ml of primary culture is transferred to 1L LB media containing appropriate antibiotics of $1/1000^{th}$ concentration of the stock (1ml). The flask was then kept at 37°C to let the cells grow.
- 500 µl of IPTG (conc. refer to materials) is added to this after the $O.D_{600}$ reaches between 0.6-0.7 and the culture was then kept at different range of conditions (including time, temperature, glucose concentration, sorbitol conc. etc.) in order to enhance protein solubility.
- After this, cells were harvested down at 8000 rpm for 7min at 4°C. The pellet was then resuspended in the native lysis buffer pH 8.0. (5ml of buffer per gram of pellet)
- Also 10 mg of lysozyme powder was added to the cells and it was then kept on ice for 30 mins.
- The solution was then sonicated for 15 min and then centrifuged at 12,000 rpm for 40 mins (4°C), to let the cell debris settle down.
- The 6X his tagged protein was then purified using Ni-NTA column.

    1. The column was first washed by passing one column length of water.
    2. The column was then equilibrated with native lysis buffer.
    3. The supernatant after sonication was then added to the column to allow the protein to bind with the Ni-NTA beads through their histidine tag.
    4. One column wash of native wash buffer (pH 8.0) was given to Ni-NTA bound proteins.
    5. Finally, the protein was eluted using native elution buffer and collected in1.5 ml MCTs.

- The eluted protein was then stored at 4°C

2.2.11] Denaturing purification

To study biophysical properties, we wanted a good amount of protein and as yield in native purification was quite low so we did denature purification. This method involves complete unfolding of protein under harsh denaturing conditions such as urea. Then refolding of protein is required after purification which can be achieved by several different methods mentioned in next section.

- Recombinant cells expressing protein of interest were grown and pelleted down. (section 2.2.10)
- Pelleted cells were resuspended in the native lysis buffer pH 8.0 containing 8M Urea.
- Sonication for 15 min was done and then centrifuged at 12,000 rpm for 40 mins (4°C), to let the cell debris settle down.
- The 6X his tagged protein was then purified using Ni-NTA column.

    1. The column was prepared for purification after passing two column length of water.
    2. Equilibration was done using one column volume of denaturing lysis buffer, 8M urea (pH 8.0).
    3. The supernatant after sonication was then added to the column to allow the protein to bind with the Ni-NTA beads through their histidine tag.
    4. Washing was done using one column volume of denaturing wash buffer (pH 6.3)
    5. To elute the protein, denaturing elution buffer (pH 4.5) was added to the column and protein was collected in 1.5 ml MCTs.

- The eluted protein was then stored at 4°C

2.2.12] Dialysis

In dialysis a semipermeable membrane is used to separate small molecules and protein based upon their size. A dialysis bag made of a semipermeable membrane and has small pores with cut-off of various sizes. The bag is filled with a concentrated solution containing proteins. Molecules that are small enough to pass through the pores of the membrane diffuse out of the bag into the buffer solution, or dialysate. Dialysis is sometimes used to change buffers. The molecules go from an area of high concentration to low concentration. When the level of concentration is equal between the bag and the buffer, there is no more net movement of molecules. The bag is taken out and inserted into another buffer, causing the concentration to be higher in the bag relative to the buffer. This causes more diffusion of molecules. This process is repeated several times to ensure that all or most of the unwanted small molecules are removed (usually done overnight). In general, dialysis is not a means of separating proteins, but is a method used to remove small molecules such as salts. At equilibrium, larger molecules (like protein) that are unable to pass through the membrane remain inside the dialysis bag while much of the small molecules have diffused out.

In case of N3-N4-N5 and N4-N5, the pore size of the membrane was 10KDa and the protein sample was in denaturing elution buffer. The protein samples were dialysed against different buffers buffer during optimization.

2.2.13] Step Dialysis

In case of step dialysis, the protein sample in denaturing elution buffer was dialysed against decreasing concentration of denaturant beginning from 8M urea to 1 M urea in some cases and 8M urea to 0M urea in some conditions.

2.2.14] On-column refolding

In case of on-column refolding, the bound protein to the Ni-NTA beads was treated with gradually decreasing concentration of the denaturant and increasing concentration of refolding buffer allowing protein to refold to its native fold and finally eluting in their appropriate buffers.

2.3] Instruments

2.3.1 Size-exclusion chromatography

Gel filtration chromatography or size-exclusion chromatography separates proteins, peptides, and oligonucleotides on the basis of size. Molecules move through a bed of porous beads, diffusing into the beads to greater or lesser degrees. Smaller molecules diffuse further into the pores of the beads and therefore move through the bed more slowly, while larger molecules enter less or not at all and thus move through the bed more quickly. Both molecular weight and three-dimensional shape contribute to the degree of retention. Gel Filtration Chromatography is used for analysis of molecular size, for separations of components in a mixture, or for salt removal or buffer exchange from a preparation of macromolecules.

For gel filtration of protein molecules ÄKTA pure chromatography system from GE Healthcare was used. ÄKTA pure is a reliable system where hardware and UNICORN™ system control software are designed to work together with columns and media to meet any purification challenge.

2.3.2 UV-Visible/ absorbance spectroscopy:

Absorption spectroscopy is one of the most widely used techniques employed for determining the concentrations of absorbing species (proteins) in solutions. It is a non-destructive technique which biologists and biochemists use to quantify the cellular components and characteristic parameters of functional molecules. This quantification is most relevant in the context of systems biology. For quantitative measurements of protein concentration in a sample we employed this technique. Instrument used for measurements was Cary 50BIO fluorimeter loaded with Cary Win UV program.

The spectra were recorded for 100 µl of protein sample under following conditions:

Pathlength of cuvette - 3mm, Wavelength Max: 400 nm, Wavelength Min: 250 nm, Speed: medium, Baseline correction: On

The absorbance at 280 nm and 310 nm were taken into account for calculation. The concentration was calculated using Beer-Lambert's Law:

$$A=\log_{10}(Io/I) = \epsilon CL$$

Where: Io = Incident intensity

I = Transmitted intensity

C = concentration

ϵ = extinction coefficient

L = Pathlength

2.3.3 Fluorescence Spectroscopy

For conformation studies of recombinant protein in presence of different solvents and chemicals fluorescence spectroscopy was done. The fluorescence signal from a protein sample is contributed by the three aromatic residues i.e. tryptophan, tyrosine and phenylalanine. In case of cadherin, we monitored the tryptophan (W) fluorescence. Typically, tryptophan has a wavelength of maximum absorption of 280 nm and an emission peak that is ranging from 300 to 400 nm depending on the polarity of the local environment.

The fluorescence spectra were recorded from Cary 50BIO fluorimeter operated by Cary Eclipse software. Following condition were followed:

Instrument was allowed to stabilize for few minutes before collection spectra.

100 µl of protein sample used in quartz cuvette of 3mm pathlength

Parameter used for Set up Emission mode:

Excitation (nm) - 280 nm
Start - 300 nm
End - 400 nm

2.3.4 Circular Dichroism (CD)

To study protein secondary structure CD spectroscopy was used. Secondary structure can be determined by CD spectroscopy in the "far-UV" spectral region (190-250 nm). At these wavelengths the chromophore is the peptide bond, and the signal arises when it is located in a regular, folded environment. Alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectrum.

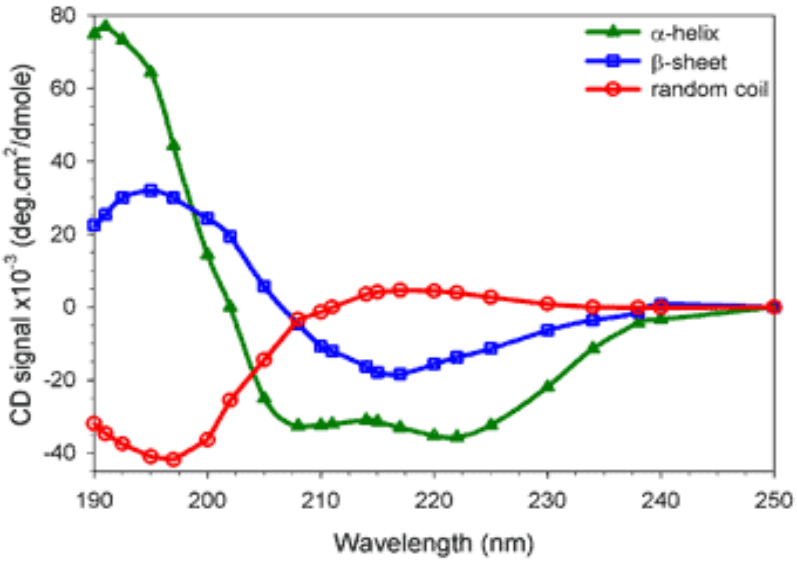


Fig 2.2 CD spectra of secondary structures of protein.

All the spectra were recorded by BioLogic CD spectrophotometer and the following parameters were used:

Cuvette Pathlength: 1mm.
Wavelength range: 200nm-250nm
Slit width: 2-5nm
Spectra: Reverse
Step: 1nm
Baseline: Subtracted

Before collection of spectra the protein sample was centrifuged at 13000 rpm for 5min to avoid any aggregates and 200 µl was added to the cuvette.

The instrument provides output in millidegrees called the epsilon which were converted to MRE by using the following formula:

$$MRE = \frac{\Theta * 100 * MRW}{1000 * C * L}$$

$$MRW = \frac{\text{Molecular weight (Da)}}{\text{No. of amino acids}}$$

Where: C – Concentration (mg/ml)
L – Pathlength of the cuvette (cm)

2.3.5 Differential Scanning Calorimetry (DSC)

Differential scanning calorimetry is a thermo analytical technique which can be used to measure the melting temperature, latent heat of melting, heat of fusion or heat capacity. There is a sample cell and a reference cell. It measures the difference in the amount of heat required to increase the temperature of the sample and reference cell as a function of temperature.

The principle underlying DSC is that when the sample undergoes a physical transformation such as phase transitions, more or less heat will need to flow to it than the reference to maintain both at the same temperature. The amount of heat that needs to flow to the sample depends on whether the process is exothermic or endothermic. If a solid sample melts to a liquid, it will require more heat flowing to the sample to increase its temperature at the same rate as the reference. This is due to the absorption of heat by the sample as it undergoes the endothermic phase transition from solid to liquid. Likewise, as the sample undergoes exothermic processes less heat is required to raise the sample temperature. By observing the difference in heat flow between the sample and reference, differential scanning calorimeters are able to measure the amount of heat absorbed or released during such transitions.

2.4] Programs

2.4.1] PyMOL

PyMOL is an open-source computer software for a molecular visualization released under the Python License. It is currently commercialized by Schrödinger, Inc. PyMOL can produce high-quality 3D images of small molecules and biological macromolecules, such as proteins.

For visualization of protein PDB files PyMOL 2.2 was used and images of docking structures were exported from PyMOL 2.2.

2.4.2] UCSF Chimera

UCSF Chimera (or simply Chimera) is an extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and movies can be created. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics (RBVI) at the University of California, San Francisco.

Chimera 1.12 was used for studying docking results and other docking characteristics. Images of protein structures were also extracted from this program.

2.4.3] ClusPro 2.0

ClusPro represents the first fully automated, web-based program for the computational docking of protein structures. Users may upload the coordinate files of two protein structures through ClusPro's web interface, for docking. The docking algorithms evaluate billions of putative complexes, retaining a preset number with favorable surface complementarities. A filtering method is then applied to this set of structures, selecting those with good electrostatic and desolvation free energies for further clustering. The program output is a short list of putative complexes ranked according to their clustering properties[21].

To gain insights into aggregation from cadherin point of view we did docking studies and most of the docking of cadherin domains was done on web-based ClusPro 2.0 Program.

2.4.4] HADDOCK

HADDOCK (High Ambiguity Driven protein-protein DOCKing) is an information-driven flexible docking approach for the modelling of biomolecular complexes. HADDOCK distinguishes itself from *ab-initio* docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modelling problems including protein-protein, protein-nucleic acids and protein-ligand complexes.

The default HADDOCK server is now version 2.2.

2.4.5] PRODIGY

PRODIGY (PROtein binDIng enerGY prediction) is a collection of web services focused on the prediction of binding affinity in biological complexes as well as the identification of biological interfaces from crystallographic one[22].

This server was used to predict the binding affinity in PROTEIN-PROTEIN complexes.

2.4.6] TM-align

TM-align is an algorithm to identify the best structural alignment between protein pairs. For two protein structures, TM-align generates optimized residue-to-residue alignment based on structural similarity using heuristic dynamic programming iterations. An optimal superposition of the two structures built on the detected alignment, as well as the TM-score value which scales the structural similarity, will be returned. TM-score has the value in (0,1], where 1 indicates a perfect match between two structures. Following strict statistics of structures in the PDB, scores below 0.2 correspond to randomly chosen unrelated proteins while those higher than 0.5 assume generally the same fold in SCOP. For both folded and misfolded models, TM-align can almost always find close structural analogs, with an average root mean square deviation, RMSD, of 3 Angstrom and 87% alignment coverage[23].

Here we have used TM-align for calculating TM values for different N- and E-cadherin domains to study structural similarity between the domains.

Chapter 3

Results

3.1] About the Constructs

N3-N4-N5:

MASEFTAMTFYGEVPENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAIQTDPNSNDGLVTVVKPIDFETNR
MFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNENPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYT
KLSDPANWLKIDPVNGQITTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVLPQEAETCETP
DPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQLNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKV
CQCDSNGDCTDVDRLEHHHHHH

Number of amino acids: 345

Molecular weight: 38128.83

Theoretical pI: 4.75

Ext. coefficient: 30160 M-1 cm-1, at 280 nm

N4-N5:

MASYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQITTIAVLDRESPNVKNNIY
NATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVLPQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWT
ITRLNGDFAQLNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDCTDVDRLEHHHHHH

Number of amino acids: 230

Molecular weight: 25623.85

Theoretical pI: 4.92

Ext. coefficient: 21680 M-1 cm-1, at 280 nm

3.2] Protein Purification

3.2.1] Native purification of N3-N4-N5 and N4-N5

The N3-N4-N5 and N4-N5 protein constructs were overexpressed in *E. coli* BL21 (DE3)pLysS* cells, and protein purification was done following a standard protocol (Fig. 3.1).

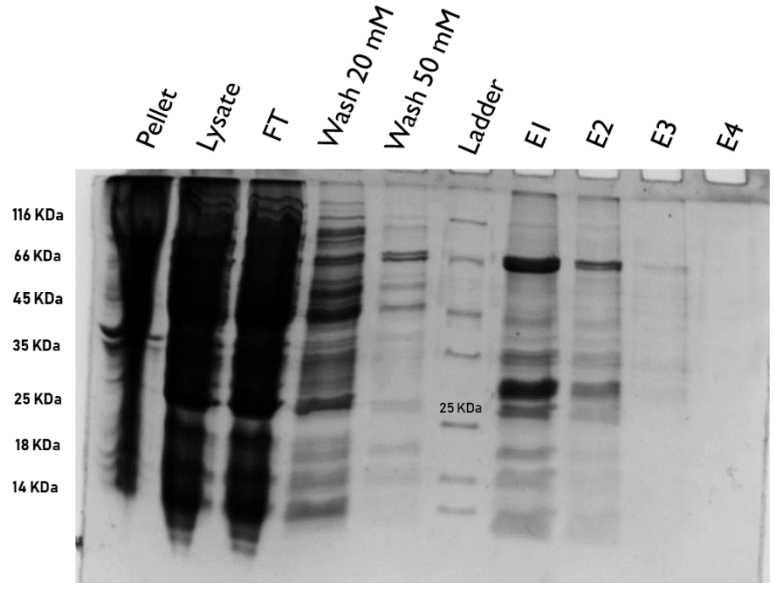
Primary culture
(in 10ml LB media + 10 µl Chloramphenicol and 10 µl Ampicillin to a final concentration of /1000th of stock concentration for both the antibiotics) incubate @ 37ºC for 12-14h)

Secondary culture
(1L LB media 1ml Chloramphenicol and 1 ml Ampicillin Amp + 10ml 1º culture incubated @ 37ºC)

IPTG Induction
(when O.D. reaches 0.6, 1ml IPTG added to 2ºculture and incubated @37º for 5h)

Harvesting and sonication followed by Ni-NTA purification

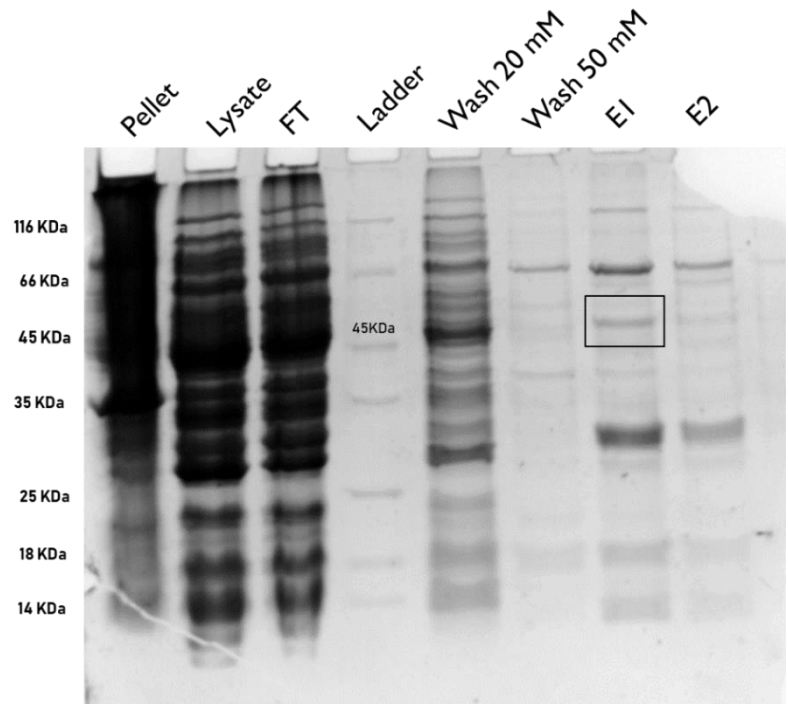
Fig. 3.1 Standard protein purification protocol.



E= Elution
FT= Flow through

Fig. 3.2 N4-N5 Protein profile in 13% SDS PAGE.

In N4-N5 protein purification, we saw a low yield of protein in addition to many contaminants, so optimization of the protocol was required.



E= Elution
N-3-4-5 expected size – 38 KDa

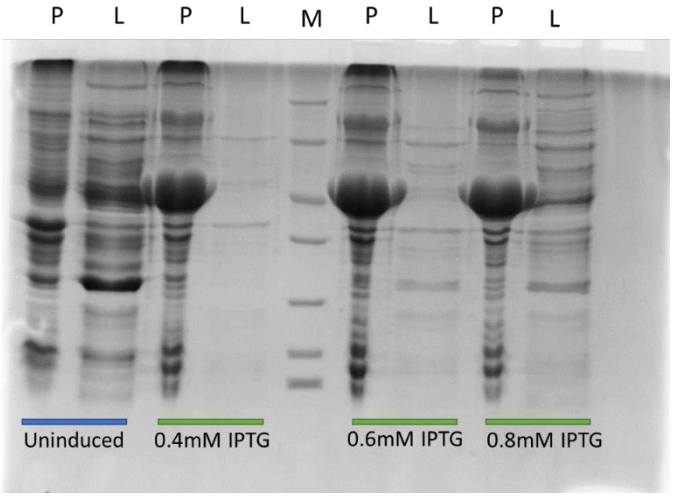
Fig. 3.3 N3-N-N5 Protein profile in 13% SDS PAGE.

Same as N4-N5, N3-N4-N5 also had low protein expression, and elution fraction had many contaminants under standard protein purification conditions.

Therefore, to improve the amount of protein in the soluble fraction, we optimized the protocol.

3.2.2] Optimization of Protein Expression

3.2.2.1] Optimization of IPTG for N3-N4-N5 purification.



P= Pellet
L= Lysate
M= Marker

Fig. 3.4 Protein expression at different IPTG concentration.

Protein expression was quite satisfactory even in 0.4mM IPTG induction but here we observed that the protein was aggregation prone and the majority of the fraction was obtained in the pellet as inclusion bodies.

3.2.2.2] Glucose and temperature optimization.

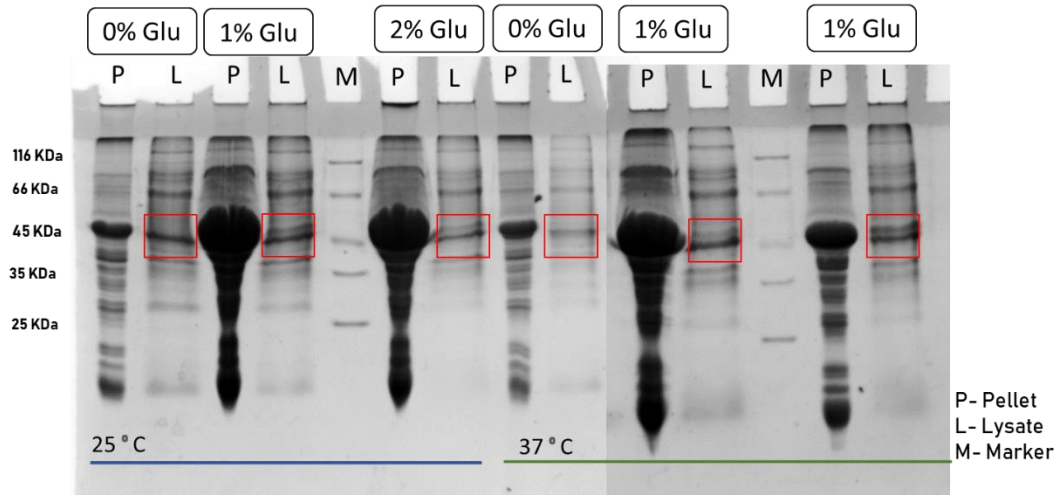


Fig. 3.5 Glucose and temperature optimization in N3-N4-N5 purification.

3.2.2.3] Glucose, Sorbitol and Rosetta expression.

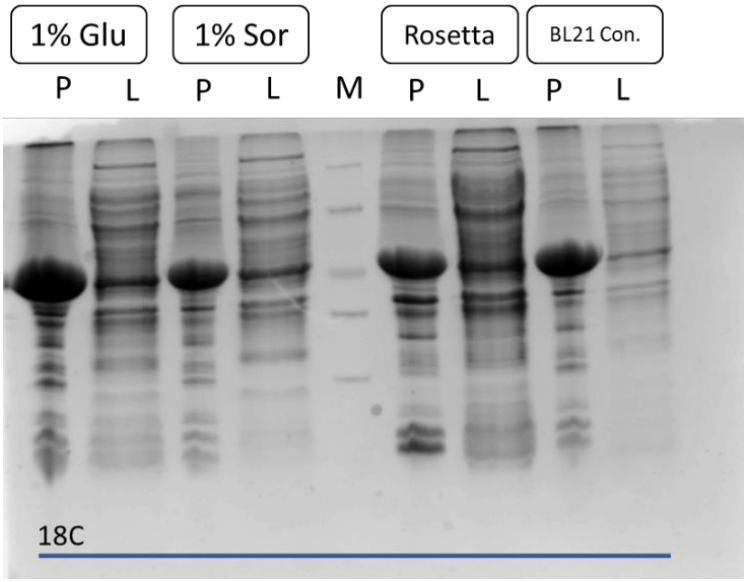


Fig. 3.6 Protein expression in presence of glucose, sorbitol and after retransformation in Rosetta (DE3) strain.

3.2.2.3] N-3-4-5 Protein Purification in optimized conditions using Ni-NTA

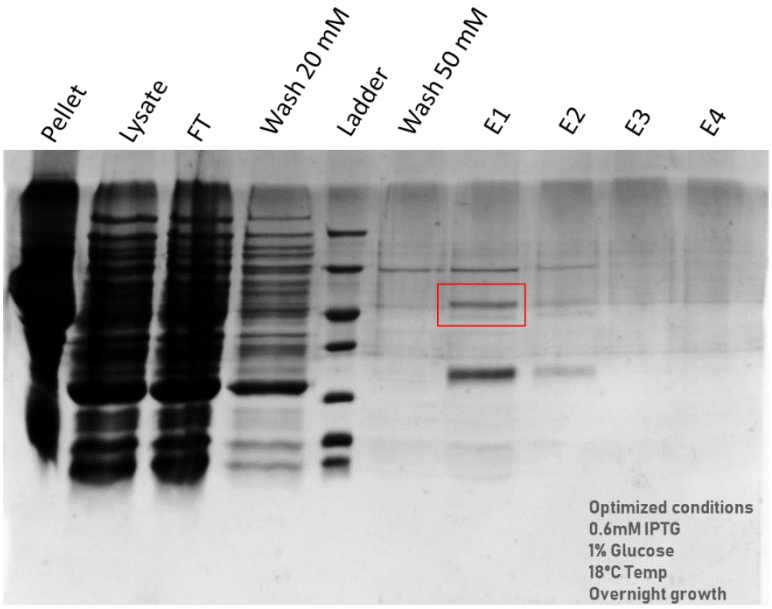


Fig. 3.7 Protein expression in optimized conditions.

Even after standardization and optimization of the parameters, the protein yield was not that satisfactory to work with although we were able to get rid of most of the contaminations. As we found both of these constructs were highly prone to aggregation, we decided to recover the protein from the pellet. So, denaturing purification of these constructs was done in urea.

3.2.3] Denaturing purification of N3-N4-N5.

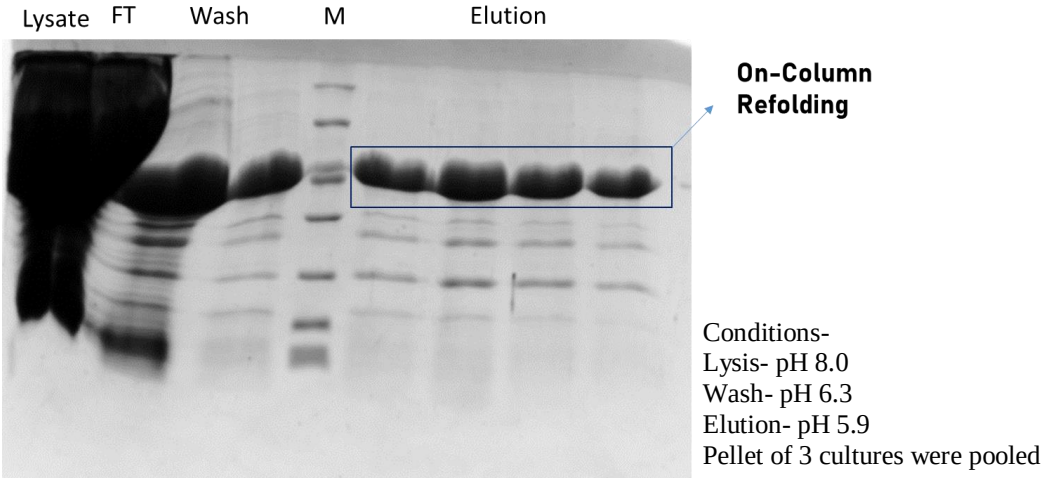


Fig. 3.8 Denaturing purification of N3-N4-N5 in 8M urea.

N-3-4-5 On-Column (Ni-NTA) Refolding

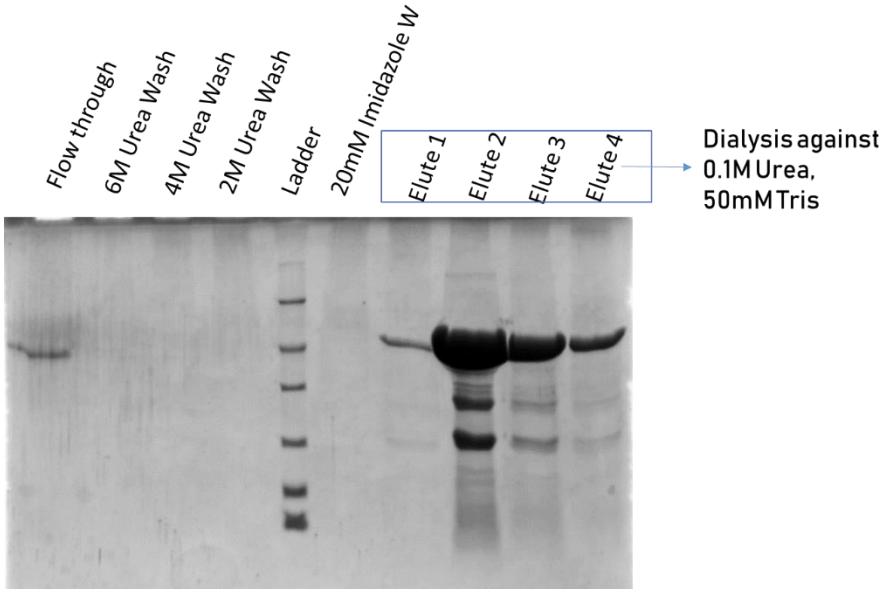


Fig. 3.9 On-Column (Ni-NTA) Refolding.

3.2.4] Confirmation of the constructs through mass spectrometry.

Peptide mass fingerprinting was done to confirm the identity of band observed on SDS PAGE, the bands of natively purified protein were carefully excised out from the gel (fig 3.10). The molecular weight of N3-N4-N5 is 38.12 KDa. However, we see some anomalous mobility on the gel, and the band of interest appears near 45 kDa. To resolve the dispute, we confirmed the construct through peptide mass fingerprinting.

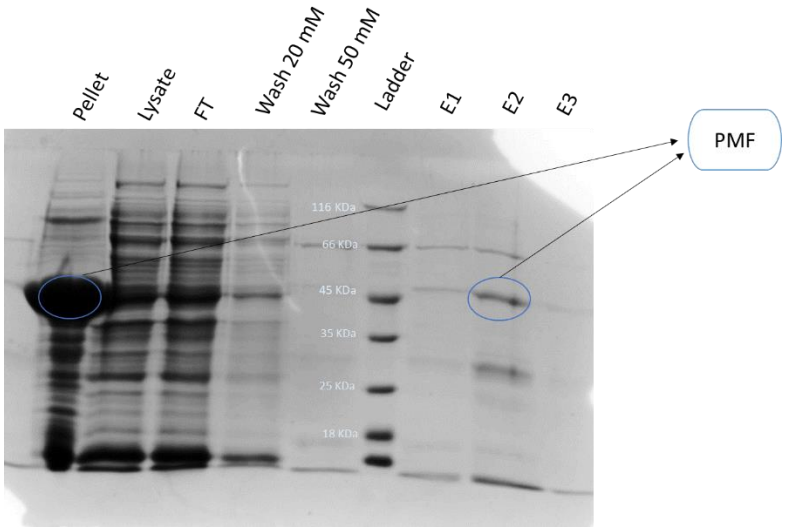


Fig. 3.10] Bands excised out of gel for PMF.

Results: Red arrows representing matched peaks with the *in-silico* digests.

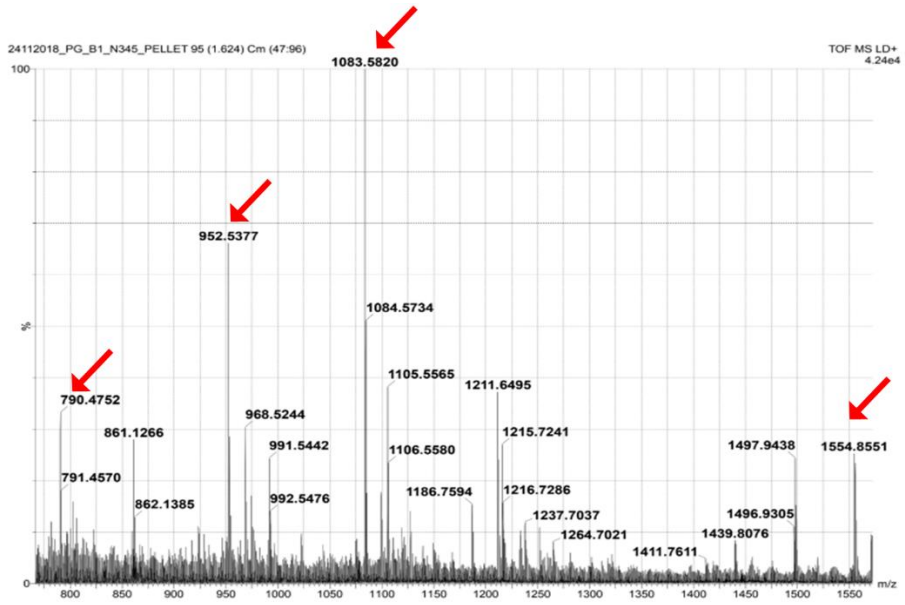


Fig. 3.11 Confirmation of N3-N4-N5 construct from mass spectrometry.

To study the structural properties of N3-N4-N5 construct, the eluted protein (both native and denatured-purified) was initially dialysed against 50 mM Tris buffer but it turned out that N3-N4-N5 forms soluble aggregates in Tris buffer. So, different buffer conditions were tried to enhance protein solubility - some of the experimental results indicating N3-N4-N5 as soluble aggregates are mentioned in the next section.

3.3] Structural Studies

3.3.1] Gel filtration

After optimization of native protein purification protocol, we ended up with three bands (arnA, slyD (known from previous studies) and N3-N4-N5) in elution fraction (figure 3.7). To separate our protein of interest from contaminations, we did gel filtration on Superdex 200 increase 10/300GL Column and the buffer used was 50 mM Tris.

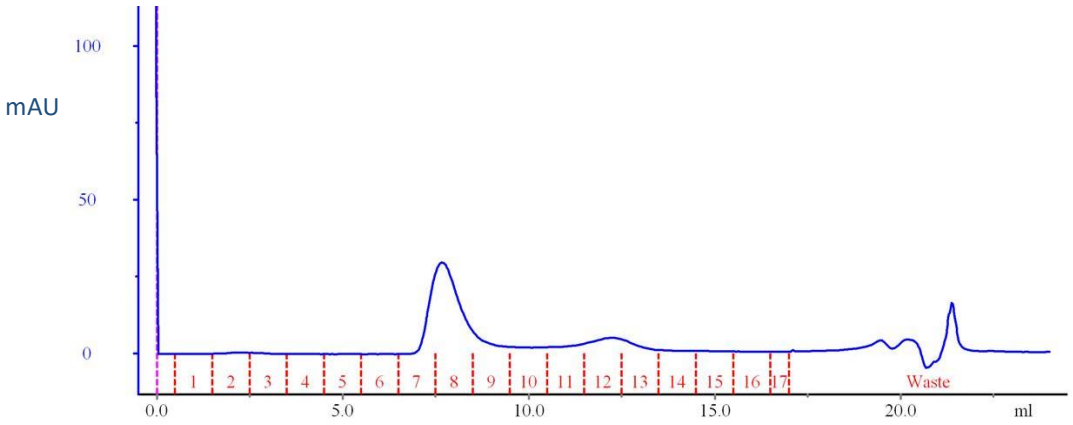


Fig. 3.12 Gel filtration curve of native purified N3-N4-N5 protein in 50mM Tris.

The gel filtration curve indicates N3-N4-N5 exists as soluble aggregates as most of the protein eluted at void volume, i.e. 8 ml. This suggests that the protein is not stable in 50 mM Tris buffer. Therefore, a low amount of urea was kept in the buffer to prevent aggregation of the protein. This time we used protein which was purified from denaturing purification followed by refolding in 50 mM tris.

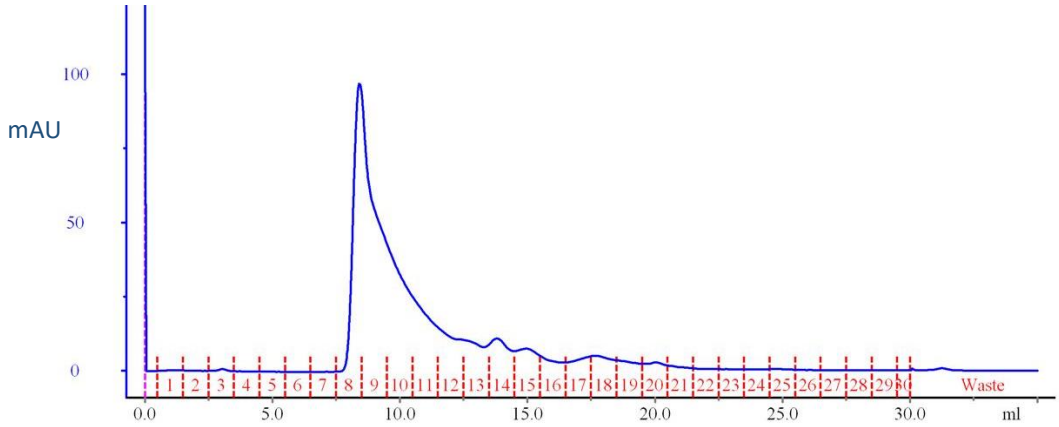


Fig. 3.13 Gel filtration curve of column refolded N3-N4-N5 protein in 50mM Tris, 0.1M Urea.

N3-N4-N5 forms soluble aggregates even in the presence of urea. Figure 3.13 shows that the protein starts eluting early in the range of large aggregates and because the protein population consisted of multiple conformations of partially unfolded species in the solution, we see elution till 12ml; however, most of the protein eluted at void volume.

Our next objective was to see whether the presence of calcium can have some effects which we can observe on the gel filtration curve. So, we checked the effect of increasing concentration of calcium on N3-N4-N5 using gel filtration by equilibrating the column with different molar concentration of calcium containing buffer.

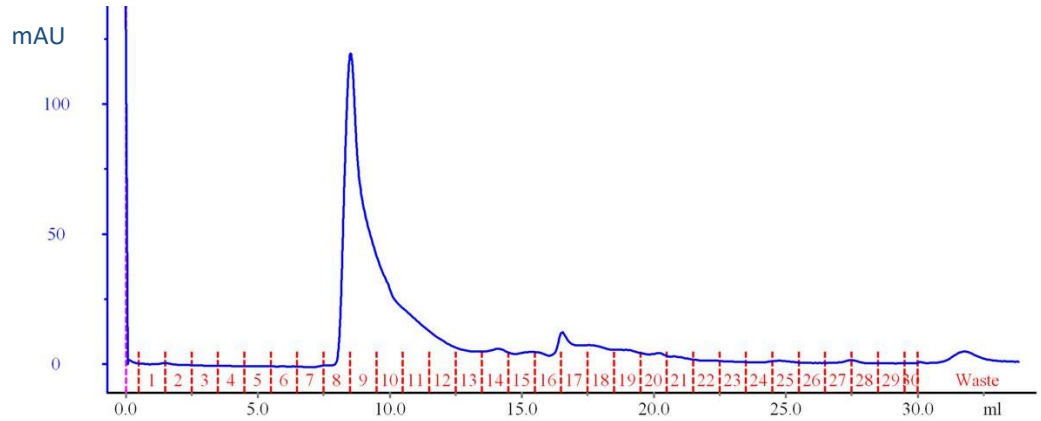


Fig. 3.14 Gel filtration curve of column refolded N3-N4-N5 protein in presence of 5 mM.

The presence of calcium had no effect on the chromatogram. From 1mM calcium to 10mM calcium, no change was observed.

Recently we adopted a new buffer (5 mM Arginine and 50 mM Tris) for cadherin storage. Initially, we started with 200 mM arginine concentration in the buffer but arginine showed hindrance during structural studies of these constructs. After a series of observations at different concentrations, the spectroscopic studies were feasible in 5 mM arginine concentration.

3.3.2] CD spectra

To gain insights into the secondary structures of these constructs we collected CD spectra.

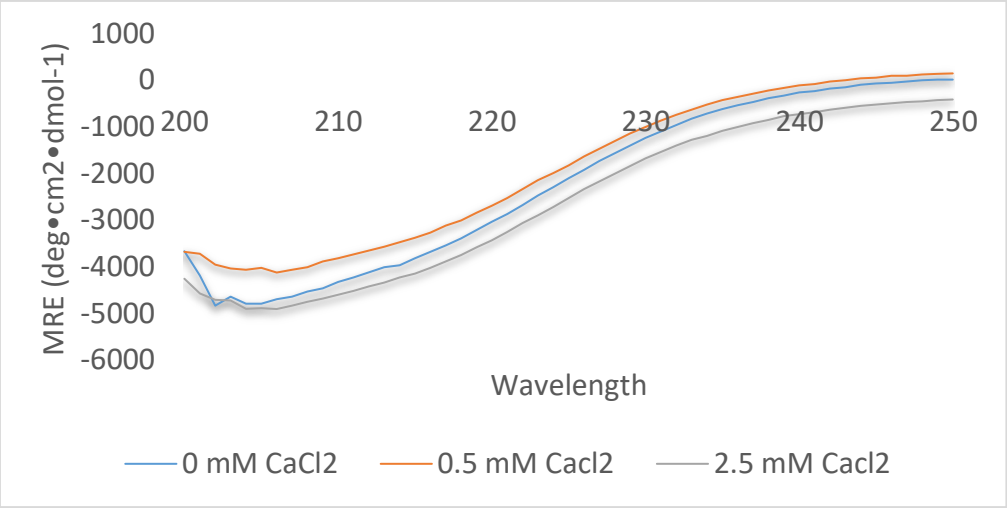


Fig. 3.15 CD spectra of refolded N3-N4-N5 protein with and without calcium.

The CD spectrum was taken in 50 mM tris buffer (pH.7.4) containing 0.1 mM Urea. The CD shows folded β-sheet along with signal from random coil. Also, we did not find any change in the secondary structure when the domain construct was incubated with calcium.

While we were looking for better buffer conditions, some CD spectra were collected to validate our previous results.

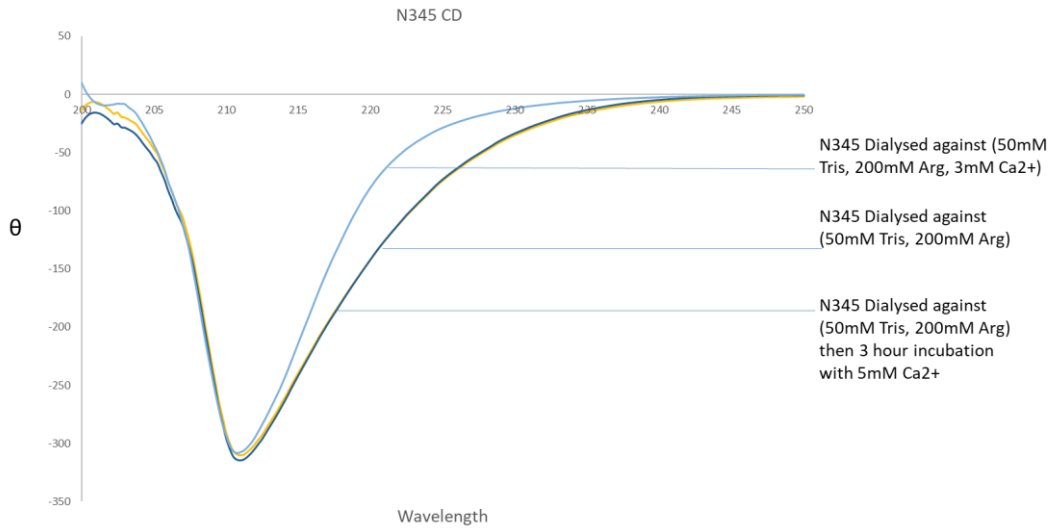


Fig. 3.16 CD spectra of N3-N4-N5 in 200mM arginine buffer.

The CD spectra of N3-N4-N5 in 200 mM arginine buffer in the first glance appears to be a perfect beta sheet, but it is not. The dip is around 210 nm, and the value of millidegree is more than 300, which is unusual for a protein. The spectrum might not belong solely to a protein, instead it might be a mixed effect of arginine and protein. Hence. the amount of arginine in buffer was reduced to 5mM after a series of test and CD spectra was taken in new conditions which matched our previous results.

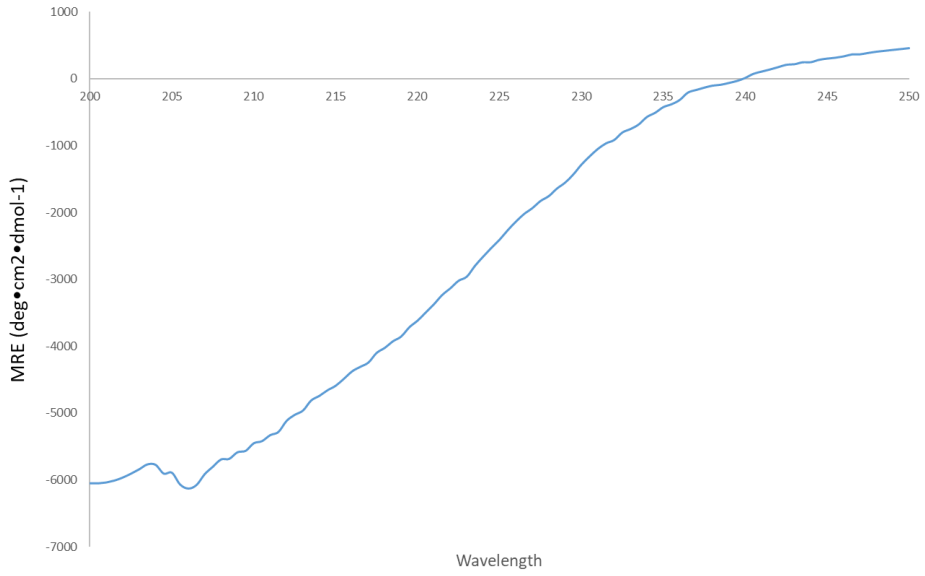


Fig. 3.17 CD spectra of N3-N4-N5 in 5mM arginine, 50mM tris buffer.

3.3.2] Fluorescence spectra

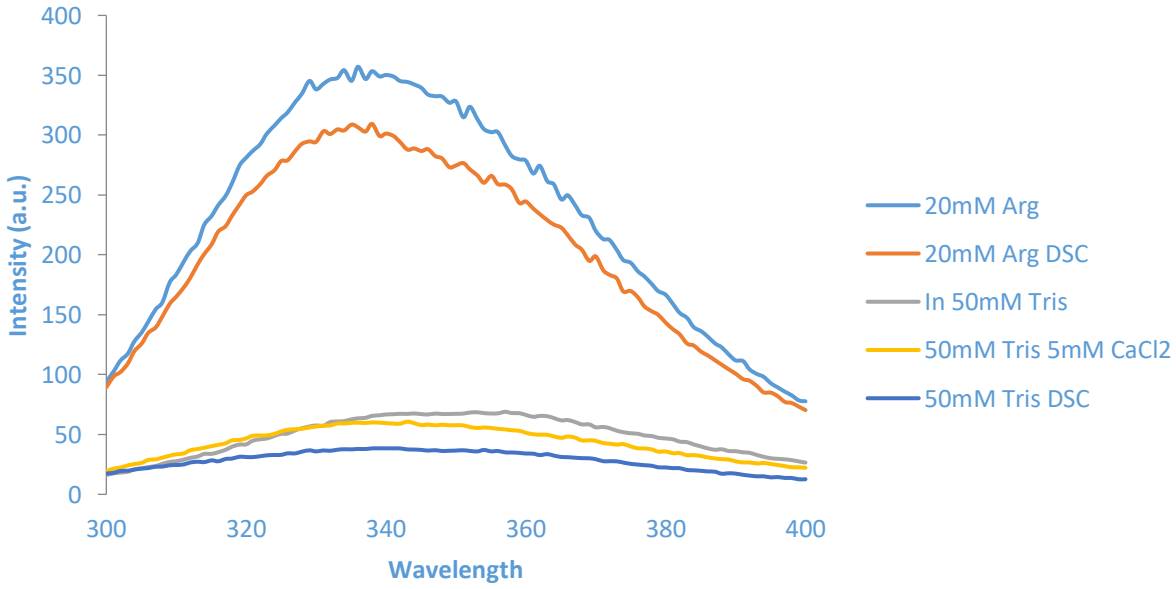


Fig. 3.18] Fluorescence spectra of N3-N4-N5 in 50mM Tris+100mM Urea buffer in different conditions.

Three interpretations can be made from the fluorescence spectra (figure 3.18)

First, when 50 mM tris and 0.1 mM urea buffer was used for fluorescence studies, we saw the effect of calcium binding which we did not see in gel filtration curve. A peak shift from 357 nm (in 50 mM Tris, fig 3.18) to 335 nm (50 mM Tris, 5 mM $CaCl_2$, fig 3.18) was observed, suggesting burial of tryptophan.

Second, The DSC curves (for both buffers, 200 mM Arginine and 0.1 mM urea) in the figure 3.18 indicates that the fluorescence decreases when the protein is heated and cooled in a DSC scan. (DSC plot- The protein solution after DSC scan was spun, and the supernatant was used for fluorescence studies).

Third, presence of arginine in buffer greatly enhanced the fluorescence intensity of N3-N4-N5 along with a blue shift in peak from 357 to 337nm. The effect can be explained by Arginine induced folding of N3-N4-N5 domain.

3.4] Computational Studies

3.4.1] N-cadherin single domain constructs

3.4.1.1] Human N-cadherin vs Mouse N-cadherin

The sequence of the N-cadherin constructs designed in our lab to study biophysical characteristics is of human N-cadherin. Crystal structure of human N-cadherin is not known whereas the crystal structure of mouse N-cadherin is known and freely available on the RCSB PDB database. So, for docking studies, we have made single and fused domain constructs of mouse N-cadherin because their sequence shares high sequence identity and similarity.

Below are the alignment results of human and mouse N-cadherin from EMBOSS Water Pairwise Sequence Alignment.

Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 553
Identity: 544/553 (98.4%)
Similarity: 548/553 (99.1%)
Gaps: 0/553 (0.0%)
Score: 2868.0

Initially, we tried docking of predicted human N-cadherin domains, but in some cases, the structural topology of predicted structures deviates so significantly that it completely loses the integrity of Ig folds. TMalign score of <0.5 was obtained when alignment with their homologous mouse N-cadherin domain was done which could not be possible given 99.1% sequence similarity. So, we decided to begin docking studies using mouse N-cadherin rather than relying on predicted human N-cadherin structures.

3.4.1.2] Mouse N-cadherin constructs

Domain 1 (N1) (N-terminus)

DWVIPPINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDRE
QIARFHLRAHAVDINGNQVENPIDIVINVIDMNDNRPEFLH

Domain 2 (N2)

EFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTPSPNMFTINNETGDIITVA
AGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVNDNPPEFTA

Domain 3 (N3)

EFTAMTFYGEVPENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAIQTDPNSNDGLVTVVKP
IDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNENPYFAP

Domain 4 (N4)

YFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQITTIAVLDRESP
NVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL

Domain 5 (N5) (C-terminus)

APQVLPQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQLNLKIKF
LEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDCTDVDR

3.4.1.3] Domain alignment and comparison

	N1	N2	N3	N4	N5
N1	1	0.73(2.81)	0.69(2.86)	0.72(2.51)	0.57(3.69)
N2		1	0.75(2.02)	0.74(2.67)	0.63(3.63)
N3			1	0.71(2.41)	0.63(3.04)
N4				1	0.68(2.54)
N5					1

Table 3.1] TM score and RMSD value of aligned N-cadherin domains.

Direction: TM-score(RMSD)

TM-scores are normalized by short chain.

3.4.2] Crystal structure and calcium binding residues in N-cadherin

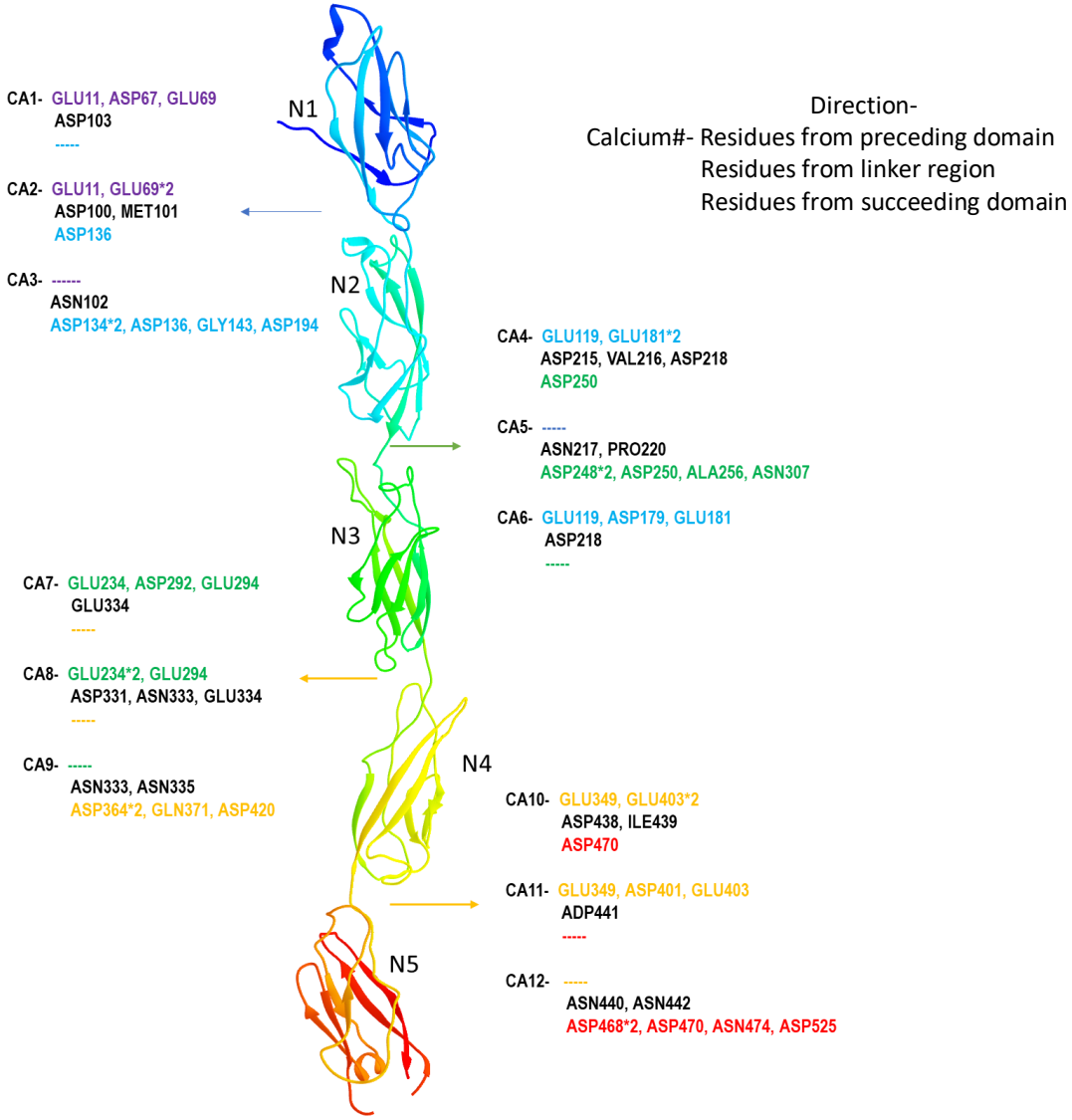


Fig 3.19 Calcium binding residues in N-cadherin.

3.4.3] Docking of single domains

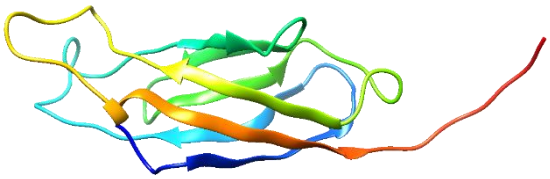
We did docking of single domain in all possible 15 combinations on ClusPro, and binding energies of docked structures were obtained from PRODIGY (PROtein binDIng enerGY prediction) as follow:

N1	N1N1	N1N2	N1N3	N1N4	N1N5
N2	N2N1	N2N2	N2N3	N2N4	N2N5
N3	N3N1	N3N2	N3N3	N3N4	N3N5
N4	N4N1	N3N2	N4N3	N4N4	N4N5
N5	N5N1	N5N2	N5N3	N4N4	N5N5

Table 3.2] Docking combination of single constructs.

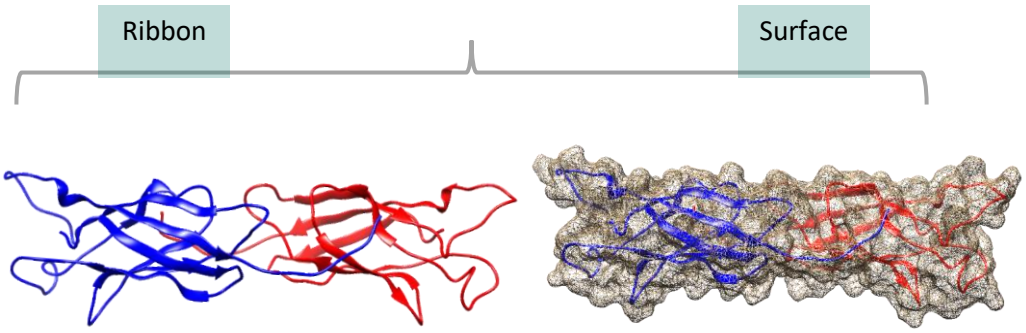
Top ten docking structures ranked by cluster size were taken in the account and analysed.

So, a total of 150 models were obtained from 15 possible combinations of single constructs but here we would like to focus our studies on N3 as N3 domain was observed to show highest tendency of aggregation experimentally among all five domains.

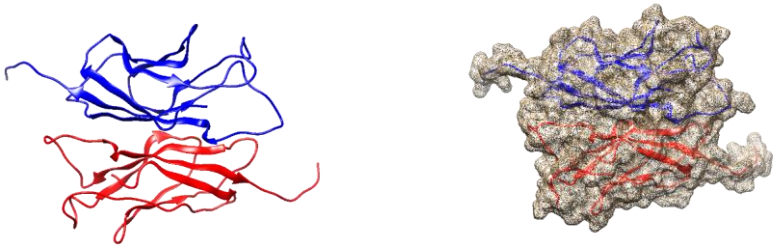


N3 Domain of N-cadherin

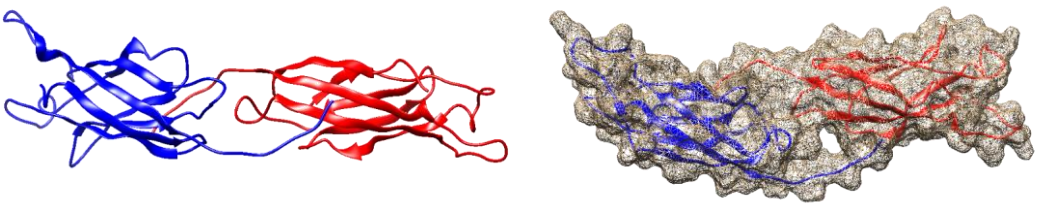
N3 docked with N3



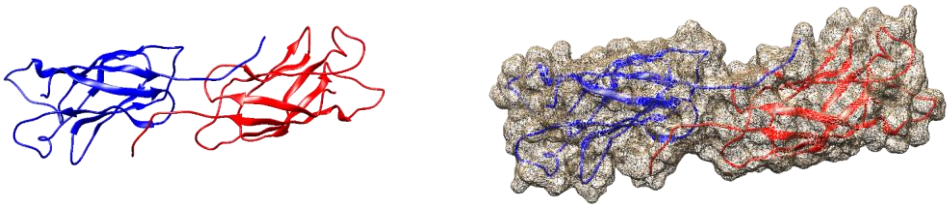
model.000.00, 73 clusters, N3N3(0)



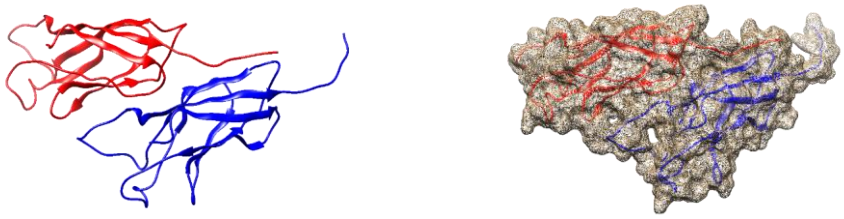
model.000.01, 71 clusters, N3N3(1)



model.000.02, 49 clusters, N3N3(2)



model.000.03, 45 clusters, N3N3(3)



model.000.04, 35 clusters, N3N3(4)

Fig 3.20 Top five N3N3 docked models and their cluster size.

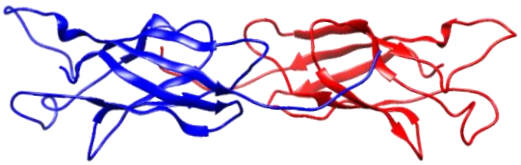
Protein-protein complex	ΔG (kcal mol-1)	Kd (M) at 25.0 °C
model.000.00	-10.8	1.30E-08
model.000.01	-11.2	5.90E-09
model.000.02	-8.6	5.20E-07
model.000.03	-11.6	2.90E-09
model.000.04	-9.3	1.60E-07
model.000.05	-12.1	1.40E-09
model.000.06	-10.4	2.40E-08
model.000.07	-8.4	6.70E-07
model.000.08	-12	1.70E-09
model.000.09	-8.5	5.70E-07

Table 3.3] Protein binding energies of docked N3N3 models.

Among the top five models, model.000.00, model.000.02, and model.000.03 roughly depicts same mode of interaction and if again clustered forms the largest cluster of 177 among all possible modes. So, we took model.000.00 for further docking process to examine N3-N3 assembly.

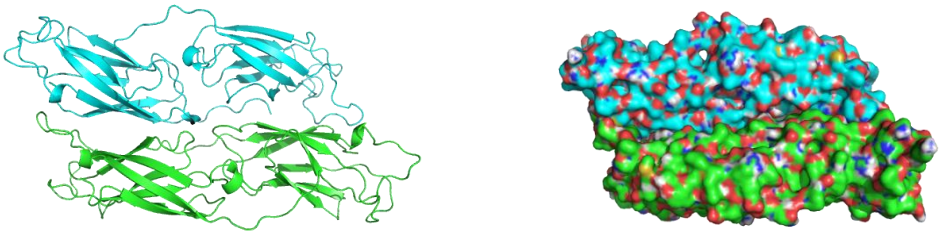
3.4.4] Docking studies with N3 to examine N3-N3 assembly

3.4.4.1] N3N3(0) docked with N3N3(0) to form N3N3.N3N3

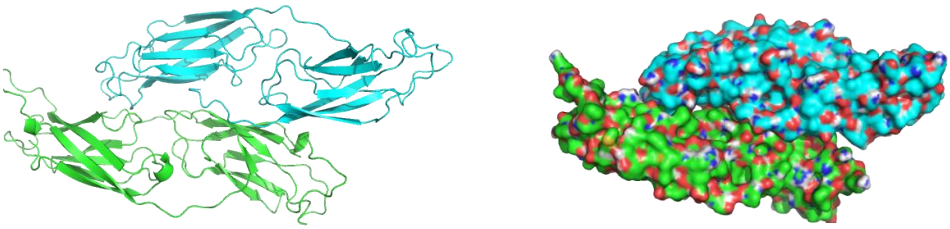


model.000.00 of N3N3

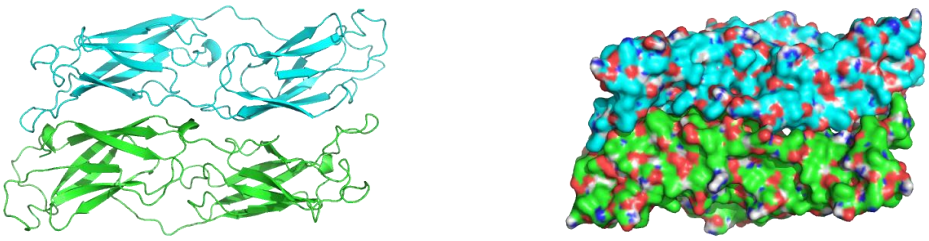
model.000.00 of N3N3
docked to itself



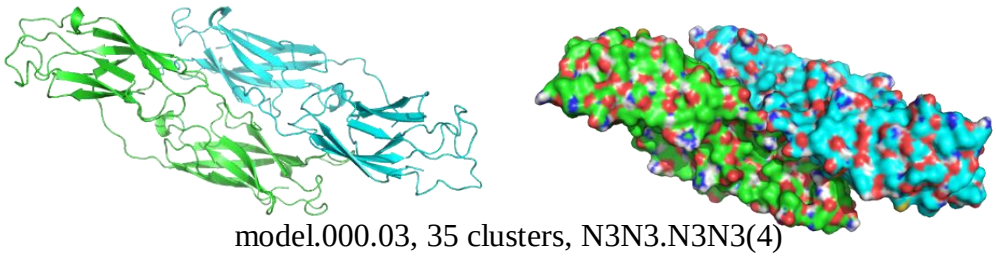
model.000.00, 40 clusters, N3N3.N3N3(0)



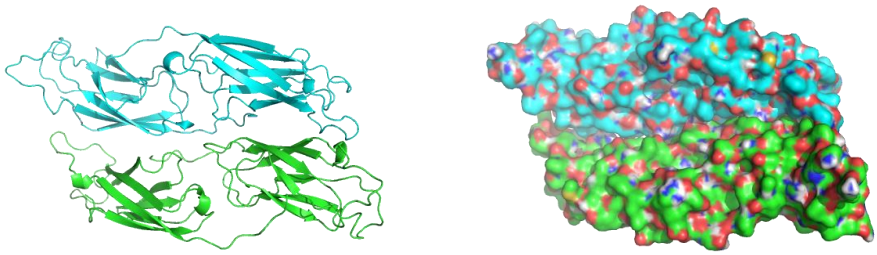
model.000.01, 40 clusters, N3N3.N3N3(1)



model.000.02, 39 clusters, N3N3.N3N3(2)



model.000.03, 35 clusters, N3N3.N3N3(4)

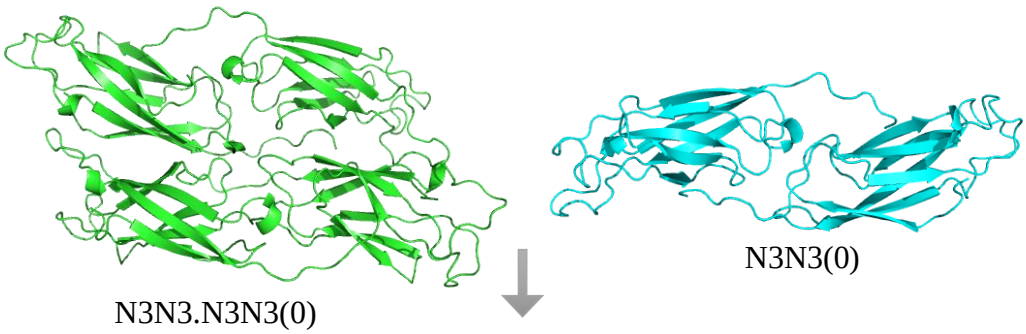


model.000.04, 34 clusters, N3N3.N3N3(5)

Fig 3.21 Top five N3N3.N3N3 docked models and their cluster size.

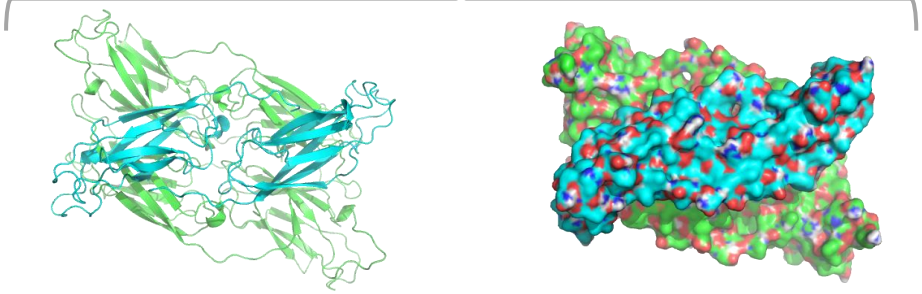
Among the top five models' model.000.00, model.000.02, and model.000.04 depicts same mode of interaction and if again clustered, forms the largest cluster of 113 among all possible modes. So, we took model.000.00 of N3N3.N3N3 for further docking process to examine N3-N3 assembly.

3.4.4.2] N3N3.N3N3(0) docked with N3N3(0) to form N3N3.N3N3.N3N3

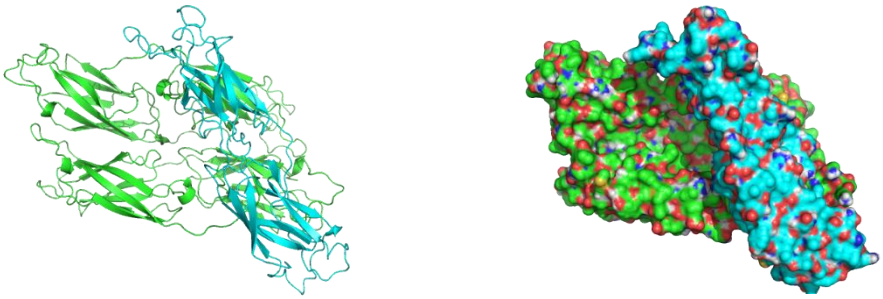


N3N3.N3N3(0)

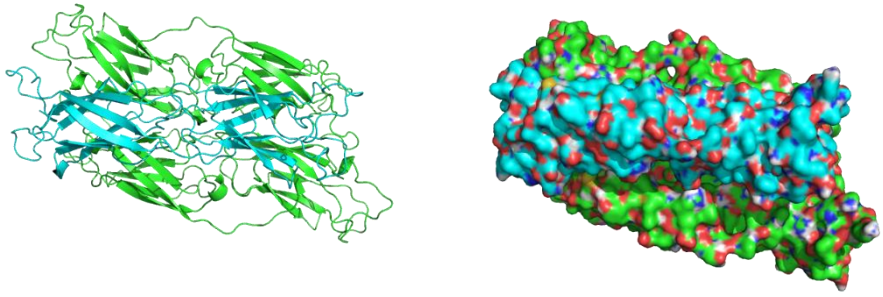
N3N3(0)



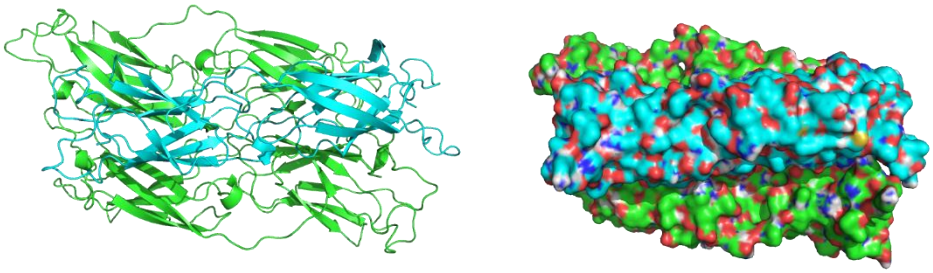
model.000.00, 30 clusters, N3N3.N3N3.N3N3(0)



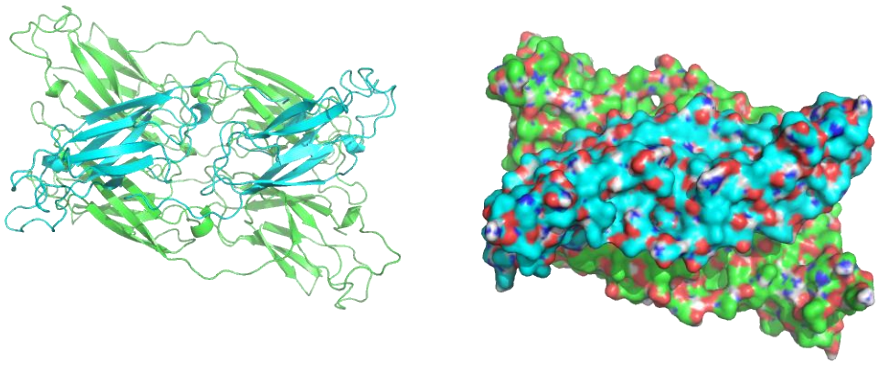
model.000.01, 29 clusters, N3N3.N3N3.N3N3(1)



model.000.02, 28 clusters, N3N3.N3N3.N3N3(2)



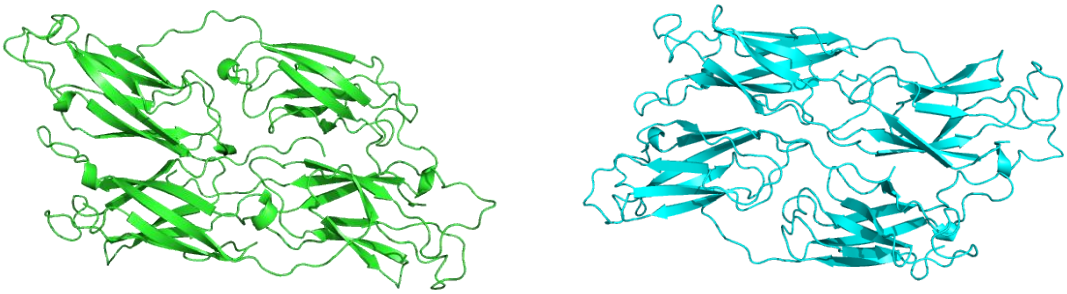
model.000.03, 24 clusters, N3N3.N3N3.N3N3(3)



model.000.04, 23 clusters, N3N3.N3N3.N3N3(4)

Fig 3.22 Top five N3N3.N3N3.N3N3 docked models and their cluster size.

3.4.4.2] N3N3.N3N3(0) docked with N3N3.N3N3(0) to form $N3N3^4$



N3N3.N3N3(0) N3N3.N3N3(0)

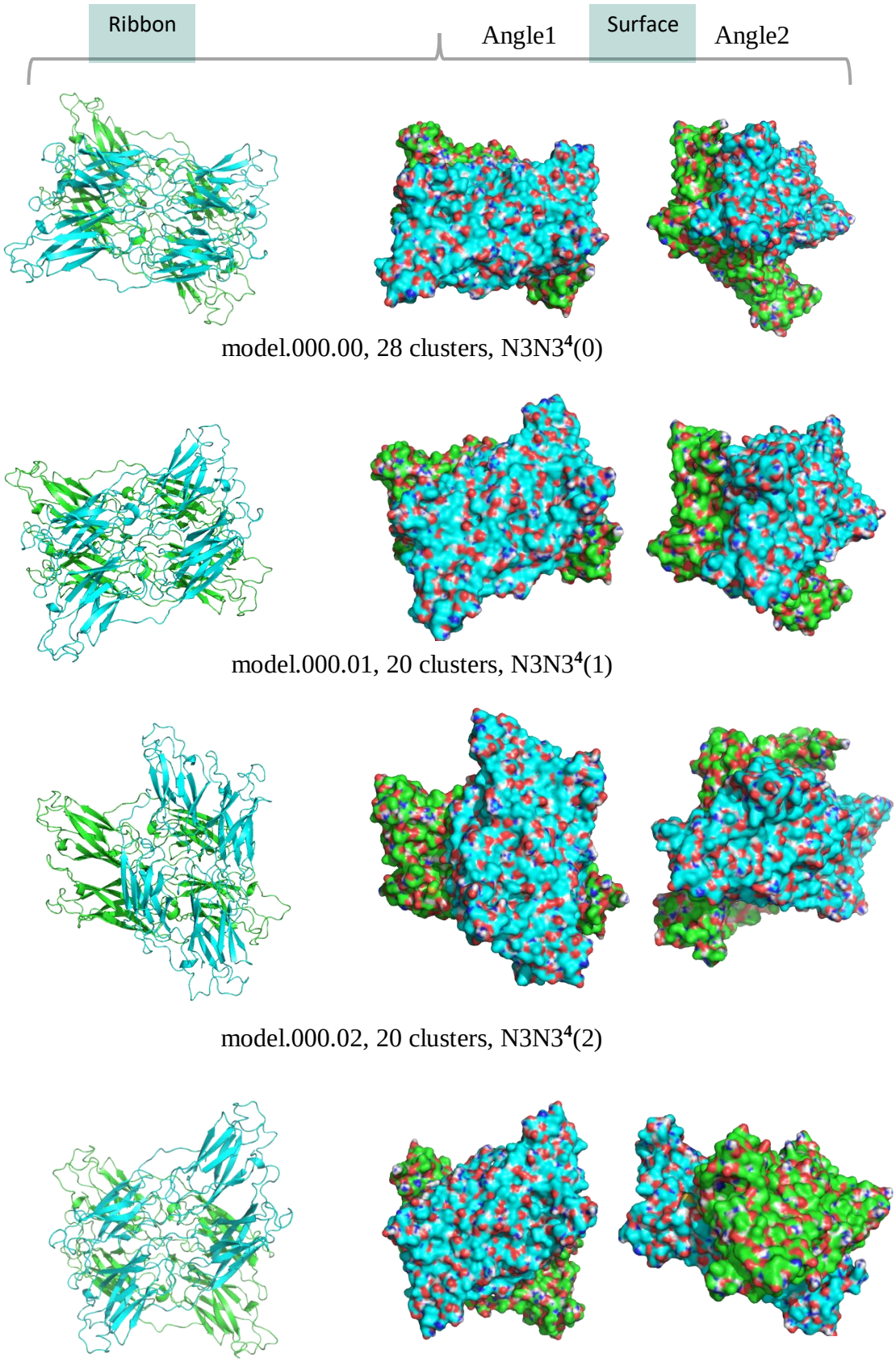


Fig 3.23] Top four N3N3[4] docked models and their cluster size.

Discussion: -

In order to keep EC1 and EC2 of N-cadherin free for interactions of a *cis* or *trans* nature, one possibility is that EC3 and EC4 display a tendency to bind to EC3 and EC4, respectively, from another molecule, either in *cis* or *trans* fashion. In fact, where there is a high density of N-cadherin molecules on membranes, such as in synapses, it is possible that molecules that cluster together on the same membrane are joined at-the-waist by EC3-EC3 interactions (or N3-N3 interactions) or even by EC4-EC4 interactions. Such a possibility is compatible with our observations that N3 and N4 of N-cadherin display a tendency to form aggregates, both in experiments with real protein constructs in solutions and in docking experiments involving computation of docking (especially for N3). In the experimental parts of this thesis, we have demonstrated that constructs containing N3 and N4 have a tendency to aggregate. Some insight into the manner and mode of aggregation is provided by the docking experiments in which we show that N3 monomers dock to dimerize, and N3 dimers dock to tetramerize, and that tetramers develop surfaces that display modes of interactions with other tetramers, all the way up to octamers, hexadecamers and beyond. It is difficult to state whether these experimental and computational results simulate natural situations at cell-cell junctions of neurons, but we advance the proposal that the aggregation-prone character of the N3 (EC3) and N4 (EC4) domains could play a role in *cis* and *trans* interactions of N-cadherins at cell-cell junctions.

References: -

[1] C. D. Buckley, G. Ed Rainger, P. F. Bradfield, G. B. Nash, and D. L. Simmons, "Cell adhesion: More than just glue (review)," *Molecular Membrane Biology*. 1998.

[2] U. Cavallaro and E. Dejana, "Adhesion molecule signalling: Not always a sticky business," *Nature Reviews Molecular Cell Biology*. 2011.

[3] D. E. Leckband and J. de Rooij, "Cadherin Adhesion and Mechanotransduction," *Annu. Rev. Cell Dev. Biol.*, 2014.

[4] M. J. Wheelock, M. Maeda, Y. Shintani, K. R. Johnson, and Y. Fukumoto, "Cadherin switching," *J. Cell Sci.*, vol. 121, no. 6, pp. 727–735, 2008.

[5] J. M. Halbleib and W. J. Nelson, "Cadherins in development: Cell adhesion, sorting, and tissue morphogenesis," *Genes and Development*. 2006.

[6] S. Hirano and M. Takeichi, "Cadherins in Brain Morphogenesis and Wiring," *Physiol. Rev.*, 2012.

[7] P. Tiwari, A. Mrigwani, H. Kaur, P. Kaila, R. Kumar, and P. Guptasarma, "Structural-mechanical and biochemical functions of classical cadherins at cellular junctions: A review and some hypotheses," in *Advances in Experimental Medicine and Biology*, 2018.

[8] R. O. Hynes, "Cell adhesion: Old and new questions," *Trends in Biochemical Sciences*. 1999.

[9] A. V. Priest, O. Shafraz, and S. Sivasankar, "Biophysical basis of cadherin mediated cell-cell adhesion," *Experimental Cell Research*. 2017.

[10] L. Shapiro *et al.*, "Structural basis of cell-cell adhesion by cadherins," *Nature*, 1995.

[11] S. T. Suzuki and S. Hirano, *The Cadherin superfamily: Key regulators of animal development and physiology*. 2016.

[12] L. Shapiro and W. I. Weis, "Structure and biochemistry of cadherins and catenins.," *Cold Spring Harbor perspectives in biology*. 2009.

[13] H. Oda and S. Tsukita, “Nonchordate classic cadherins have a structurally and functionally unique domain that is absent from chordate classic cadherins,” *Dev. Biol.*, 1999.

[14] F. S. Walsh *et al.*, “N-Cadherin Gene Maps to Human Chromosome 18 and Is Not Linked to the E-Cadherin Gene,” *J. Neurochem.*, 1990.

[15] T. V. P. Bliss and G. L. Collingridge, “A synaptic model of memory: Long-term potentiation in the hippocampus,” *Nature*. 1993.

[16] I. Ramis-Conde, M. A. J. Chaplain, A. R. A. Anderson, and D. Drasdo, “Multi-scale modelling of cancer cell intravasation: The role of cadherins in metastasis,” *Phys. Biol.*, 2009.

[17] M. I. García-Castro, E. Vielmetter, and M. Bronner-Fraser, “N-cadherin, a cell adhesion molecule involved in establishment of embryonic left-right asymmetry,” *Science (80-.).*, 2000.

[18] J. Brasch, O. J. Harrison, B. Honig, and L. Shapiro, “Thinking outside the cell: How cadherins drive adhesion,” *Trends in Cell Biology*. 2012.

[19] T. D. Perez and W. J. Nelson, “Cadherin Adhesion: Mechanisms and Molecular Interactions,” 2010.

[20] B. Zhu, S. Chappuis-Flament, E. Wong, I. E. Jensen, B. M. Gumbiner, and D. Leckband, “Functional analysis of the structural basis of homophilic cadherin adhesion,” *Biophys. J.*, 2003.

[21] D. Kozakov *et al.*, “The ClusPro web server for protein-protein docking,” *Nat. Protoc.*, 2017.

[22] L. C. Xue, J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin, and A. Vangone, “PRODIGY: A web server for predicting the binding affinity of protein-protein complexes,” *Bioinformatics*, 2016.

[23] Y. Zhang and J. Skolnick, “TM-align: A protein structure alignment algorithm based on the TM-score,” *Nucleic Acids Res.*, 2005.