Studies of the structural-biochemical behaviour of individual (and combined) extracellular domains of human E- and N-cadherins

A Thesis

Submitted
By

Prince Tiwari

PH12123

For the award of the degree of

Doctor of Philosophy

Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) Mohali
Sector-81, Mohali-140306, Punjab, India

Dedicated to my family

Declaration

The work presented in this thesis entitled "Studies of the structural-biochemical behaviour of individual (and combined) extracellular domains of human E- and N-cadherins" has been carried out by me under the supervision of Prof. Purnananda Guptasarma at the Indian Institute of Science Education and Research (IISER) Mohali. This work has not been submitted in part or full for a degree, a diploma, or a fellowship to any 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 bona fide record of original work done by me and all sources listed within have been detailed in the references.

Date: Prince Tiwari

Place: (Candidate)

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

Date: Prof. Purnananda Guptasarma

Place: (Supervisor)

Table of Contents

Acknowledgements

Life is a journey, not a destination. To find the journey's end in every step of the road, one needs support and help at different stages which provide confidence and belief towards successful completion of a goal. I am grateful that I have been blessed with numerous such people who contributed directly or indirectly towards shaping this thesis. I want to thank all, without whom I would not have been reached this far.

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Purnananda Guptasarma; Sir, you have been a tremendous mentor to me. I would like to thank, for encouraging my research and for allowing me to grow as a research scholar. I would especially like to thank for your valuable feedback, motivation and the freedom that you provided me to do my experiments. Sir, your advice and quality discussions on both science as well as on life has been priceless. You have not only guided me in the entire period of PhD but genuinely enlightened the excitement of science inside me. Your keen interest towards proteins and scientific discussions, especially in the area of protein folding, stimulated me to think dipper and develop a sound understanding of the field. Sir, along with being a guide, you have also been a true friend and philosopher; in short, you are the best teacher one can have ever.

Besides my supervisor, I would like to thank my thesis committee members: Dr Kausik Chattopadhyay and Dr Mahak Sharma, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I would like to thank Prof. N. Sathyamurthy, former Director, IISER Mohali, for giving me the opportunity to work at this premier research institute. I would also like to thank Prof. Debi Prasad Sarkar, the Director, IISER Mohali to provide the opportunity to use the excellent infrastructure for carrying out my research work. I express my immense gratitude to all faculties, non-teaching staff members and students of the Department of Biological Sciences, IISER Mohali for providing me access to departmental facilities and their help and support.

I would like to thank Ministry of Human Resource and Development, Govt. of India and Centre for Protein Science Design and Engineering (CPSDE) for providing financial support and Department of Biotechnology, Govt. of India for providing international travel grant.

I would like to thank Prof. Manni Luthra Guptasarma for allowing me to work with her research group and providing me with an opportunity to learn cell culture techniques. I would also like to thank Dr Maryada Sharma, Dr Brij, Dr Manish, Dr Priti and a special thanks to Dr Anil Tiwari for his motivation and scientific discussions.

I am very fortunate to have a bunch of scientifically incredible lab members. Their active interest in my work, useful suggestions and constructive criticism improved the quality of my work. I thank Dr Satya Prakash, Dr Neeraj Dhaunta, Dr Javed, Dr Sukhdeep, Dr Prerna Sharma, Dr Arpana Jha, Dr Kanika Arora and Dr Nitin, Bhisham, Arpita Sarkar, Archit and also, to MS students Anu, Prachi, Vaishnavi, Shahnaz and Santosh. I also thank the 'cadherin group' members; Dr Rajendra Kumar, Arpita Mrigwani and Harpreet. A special thanks to Pallavi; thank you for your continual support and warm humour – I am lucky to have a lab mate and friend like you.

My time at IISER Mohali was made enjoyable in large part due to the many friends and batchmates that became a part of my life. I thank Prashant, Rohan, Devashish, Shivani, Manisha, Anwar, Shiv, Soumitra, Krishna, Shashank, Poonam for creating such a vibrant environment. I also thank Aastha and Bhupinder for long sessions of discussions and for a great time. I would especially like to thank the DAVV gang; Saikat, Saurabh, Ravinder, Sudeep, Sushil, Jai Shankar, Kirti, Ramakant, Sharad, Garima, Anjali, Varsha, Sunita, and Shweta being great friends since M.Sc. days and making the journey memorable.

Finally, but not least, I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life, I am thankful to my parent, Mr Sachchida Nand Tiwari and Mrs Sita Devi, sisters Dipti and Tripti for their love and support. Without them, I would not have been able to finish my goal successfully.

Above all, I owe it all to Almighty God for granting me the wisdom, health and strength to undertake this research task and enabling me to its completion. Thanks for *everything* that helped me get to this day.

Synopsis

Thesis title: Studies of the structural-biochemical behaviour of individual (and combined) extracellular domains of human E- and N-cadherins

Introduction

Cadherins are important cell adhesion molecules (CAMs) which are responsible for cell-cell interactions. There are various types of cadherins in which classical cadherins are the most studied. This class of cadherins is characterized by a multidomain extracellular structure which is responsible for dimerization at the cell surface, a small transmembrane domain and a cytoplasmic domain which does the signalling. The two important representatives of classical cadherins; epithelial (E) and Neuronal (N) cadherins are identified by the presence of an extracellular domain comprising five 110 amino acids-long repeats (EC1 to EC5) and a highly conserved tryptophan residue at the 2^{nd} position in the extracellular domain 1 (EC1). During cell-cell interactions, when cells come close to each other, the cadherins from each cell surface interact and undergo dimerization. During the process of dimerization, tryptophan 2 of EC1 binds into a hydrophobic pocket of the EC1 of cadherin from the opposing cell, and the tryptophan 2 from EC1 of the opposing cell binds the hydrophobic pocket of the EC1 of the first cell, and this mode of binding is called domain or strand swapping.

Objectives

Different models have been proposed to explain the process of dimerization. All of them hold the domain 1 (EC1) is involved in the process of dimerization and domain 2 (EC2) is involved in some cases. We asked some fundamental questions after a preliminary analysis of sequence and structural alignments of the domains, EC1 to EC5, which we call E1 to E5 for E-cadherin and N1 to N5 for N-cadherin.

- What are the roles of domains 3, 4 and 5 of E- and N-cadherins? Do they perform any specific roles during cell-cell interaction?
- Which EC domains interact with which other ones?
- Are there any E domain -N domain interactions (given the extreme structural similarities)?
- Which EC domains are structured and soluble when produced singly?

- Which EC domains are poorly structured, and show a tendency to aggregate?
- Which EC domains form defined homo-multimers by themselves, e.g., E1, E2, E3 etc.?
- Which EC domains undergo structural changes (rigidification) upon calcium binding?
- Which combinations of EC domains (e.g., E1-E2 or E1-E2-E3) show binding to calcium?

So, we decided to make the domains (all of them, and in all possible combinations, with/without affinity tags), to examine their structural-biochemical behaviour, individually, in fusions and towards each other. Some of the questions posed above were addressed while some others have been left for later studies. On the other hand, new questions came up as a result of studies addressing some of the above questions. Therefore, the questions above may be taken merely to be an indicator of why we set about doing this work. As the results will show, we carried out several unplanned studies.

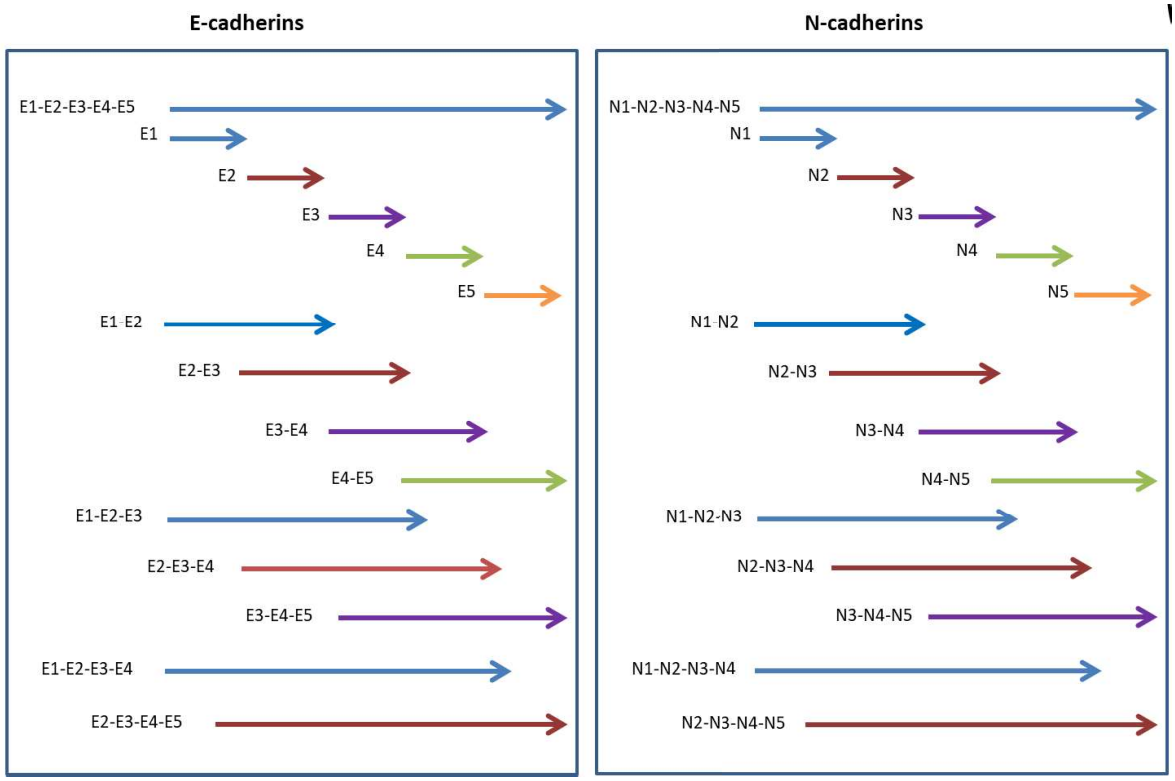


Figure 1: Schematic diagram representing truncation of all the isolated and fused domains

The extracellular regions of human E- and N-cadherins, comprising EC domains EC1-EC5 (hereinafter called as E1/N1, E2/N2, E3/N3, E4/N4 and E5/N5) were amplified into cDNA using appropriate primers and template DNA. To amplify these regions, the

primers were designed for cloning in-frame using the pET-23a vector in *E.coli* strain XL-1 Blue and for protein expression, the *E.coli* strain Rosetta DE3 was used.

Outcome of the study

Individual domains of E- and N-cadherins were studied first. A total of 10 individual domains and 2 fusion domains were studied in great detail. Protein purification was performed by affinity chromatography under both non-denaturing and denaturing conditions.

Discovery of anomalous mobility on SDS-PAGE. The SDS-PAGE analysis of the domain and domain fusion constructs showed that despite their being of comparable sizes (approx. 14 kDa), most of these domains purified under denaturing conditions showed unusual migration on SDS-PAGE. The apparent molecular weights were higher than the calculated molecular weights. After the confirmation of domain identity and after ruling out any possibility of peptide chain modifications, we came to the conclusion that it was the high negative charge due to an excess of acidic residues which resulted in poor binding of SDS, causing slower migration on SDS-PAGE. Our finding is that if the ratio of numbers of negative charge-carrying amino acids (NC) to positive charge-carrying amino acids (PC) ratio exceeds the value of '1.50', the protein displays anomalous mobility on SDS-PAGE. We have published this in the background context of all causes of such anomalous mobility, both known and conceivable.

Evaluation of secondary, tertiary and quaternary structures. The circular dichroism (CD) data shows that domains E1, E2, N1 and N2 are well folded into β-sheet-based structures whereas the EC domains 3, 4 and 5 are randomly coiled. None of the individual domains show any significant structural change upon incubation with calcium, when the linker joining them to the next domain along the polypeptide is present in fusion. We have not yet checked the effect of calcium binding when the linker preceding the domain is present in fusion. The fused domains E1-E2 and N1-N2 also were found to be folded into β-sheet-based structure and, in the cases of these fused domains, the folding of the chain was even more enhanced upon being incubated with calcium, suggesting that the presence of the next-neighbour domain allows the protein to bind calcium and display structural change as a consequence. The fluorescence emission studies showed that tryptophan residues were buried inside the hydrophobic cores. The quaternary structure

was determined by size exclusion chromatography, glutaraldehyde crosslinking and analytical ultracentrifugation and demonstrated to be dimeric in the presence of calcium, for the fusion domains E1-E2 and N1-N2.

Evaluation of thermal and chemical stability. The differential scanning calorimetry (DSC) studies of the protein constructs suggest that all the individual domains possess melting temperatures (T_ms) in the range of 45-65 ºC and all show a roughly ~5 ºC increase in T_m was observed in the presence of calcium. The EC domains of N-cadherins showed a relatively high degree of refolding ability (especially N2, which survived 22 cycles of unfolding-refolding with no hysteresis in the DSC data sets for unfolding and refolding); this suggests some physiological implications, e.g., in the case of neuronal cells, perhaps mechanical forces unfold domains during dissociation and reattachment of cells, and the refolding ability of domain N2 helps molecules to continue to function on the cell surface without replenishment, without losing function. The chemical denaturation studies of the E1 and N1 domains show that Gdm.HCl is very much more effective in comparison to urea in effecting denaturation (200 times in the case of E1 and over ten times in the case of other domains), which indicates that the structures of these domains are largely stabilized by electrostatic interactions. The fusion domains derived from E1 were also shown to be stabilized by mainly electrostatic interactions.

We have also shown how cadherin domains are able to fold independently with respect to their neighbouring domains. To begin with, we picked E2-E3 domain, in which E2 was soluble, and E3 expressed as soluble aggregates. We constructed a fusion domain with these two domains, and the resulting protein was soluble, as well as well folded, in the *E.coli* cytoplasm and showed enhanced structural content when allowed to interact with calcium. The data indicates that the neighbouring domains are also responsible for folding and important for facilitation of function. One interesting observation was that fluorescence emission spectra of E2-E3 with calcium showed a dramatic several-fold increase in fluorescence intensity arising on account of energy transfer between the tyrosine residues of E2 and the tryptophan residue of E3.

Conclusions

The five extracellular domains of human epithelial(E) and neuronal (N) cadherin were produced in recombinant form, both individually, as well as in molecular fusion in the form of many variants of different lengths. A total of nearly 30 constructs were produced.

The domain and domain fusion constructs were found to differ significantly in respect of their electrostatics, solubility upon expression in a heterologous host, nature and content of folded secondary and tertiary structure(s), thermal and chemical stabilities, tendency to aggregate, and the ability to bind to calcium, as well as in their relative affinities of calcium binding. Chemical denaturation studies suggested that all domains are significantly stabilized by electrostatic interactions. Without exception, domains 1 and 2 were well-structured, and soluble, while domains 3, 4, and 5 are poorly structured as well as poorly soluble, both individually and in fusion. Calcium binds to the inter-domain linker regions with significant effects on domains 1, 2 and 3. We are using these constructs to carry out interaction-studying assays involving issues of affinity versus avidity in cadherin-cadherin interactions. We are also studying whether calcium binding to the linkers between domains affect the structure and stability of the succeeding domain along the chain, since it is clear that calcium binding has no effect on the structure and stability of the previous domain alone, but has an effect when the linker is flanked by both the preceding and the succeeding domain.

Chapter 1: Introduction

Multicellular organisms are composed of small independent and composite units of function comprising growth and division, which are known as cells. These are further assembled in many ways, to form tissues, organs and eventually individual organisms. Several models have been proposed to elucidate the mechanism of structure formation, leading to progressively larger units of function. Out of these, the 'selective adhesion' hypothesis proposed by Townes and Holtfreter (1955) is one of those which is widely accepted (Townes and Holtfreter 1955). According to this hypothesis, cells of the same type aggregate together, and different types of cells segregate to form a tissue-like structure. However, the molecular mechanism underlying this was not clear for a long time. The primary question was apparently about the type of molecule, or protein, which could be responsible for the aggregation or adhesion activity of cells. Eventually, Takeichi proposed the Ca^{2+}-dependent cell adhesion hypothesis, and his group also identified the protein responsible for cell adhesion, which was given the name cadherin.

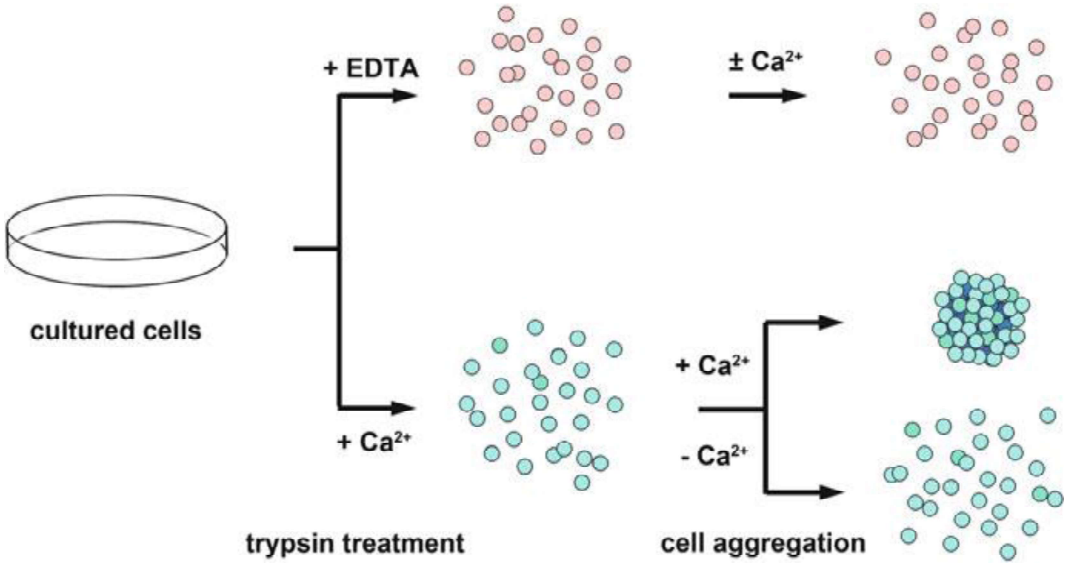


Figure 1.1. Takeichi's experiment showing Ca^{2+} protection of cell adhesion activity from trypsin.

In 1977, an experiment was performed by Takeichi's group in which they discovered an essential property of cell adhesion and demonstrated that cell-cell aggregation activity was easily destroyed by the proteolytic enzyme, trypsin, in the presence of EDTA, whereas the cell aggregation activity was significantly protected from trypsin digestion in the presence of Ca^{2+}. Based on these observations, a Ca^{2+} -dependent cell adhesion mechanism was identified, and the group also identified a protein candidate named cadherin (Urushihara and Takeichi 1980). Later, after the advancement of cloning technology, several groups discovered various types of cadherins and their role in cell-cell adhesion. Hyafil et al. (1981) discovered a protein called uvomorulin which is responsible for the compaction process during morula formation (Hyafil, Babinet et al. 1981). Ozawa et al., using the technique of immunoprecipitation, identified some proteins and revealed that they are bound to the specific site of the cytoplasmic domain of cadherin and were essential for adhesion activity (Ozawa, Baribault et al. 1989). They named the proteins 'catenins' and established that cadherins are firmly bound with β-catenin that further binds to actin filaments through a molecule called α-catenin, thus resulting in the formation of a complex which plays a vital role in cell adhesion.

Even after so much information had been gathered, a puzzling problem persisted. No one could show the formation of extracellular domains aggregation, or the formation of dimers of these cell adhesion molecules, despite the widespread impression (and assumptions) that the extracellular domains are responsible for the cell adhesion function of the cadherins. Over time, several experiments indicated that the domains had a role in interaction activity, but still, this remained difficult to establish with a strong foundation until results from X-ray crystallography revealed the structural basis of cadherin-cadherin interaction (Shapiro, Fannon et al. 1995). The conclusion was that an extracellular (EC) domain of cadherin interacts with another EC domain from another cell forming what are called 'trans' dimers, and the dimers interact in a side-by-side fashion, to also form a 'cis' interaction. This model was later called the 'zipper' model for cadherin interaction. Interestingly, this model does not require Ca^{2+} for interaction, although Ca^{2+} is an essential factor for the formation of an active conformation capable of interaction.

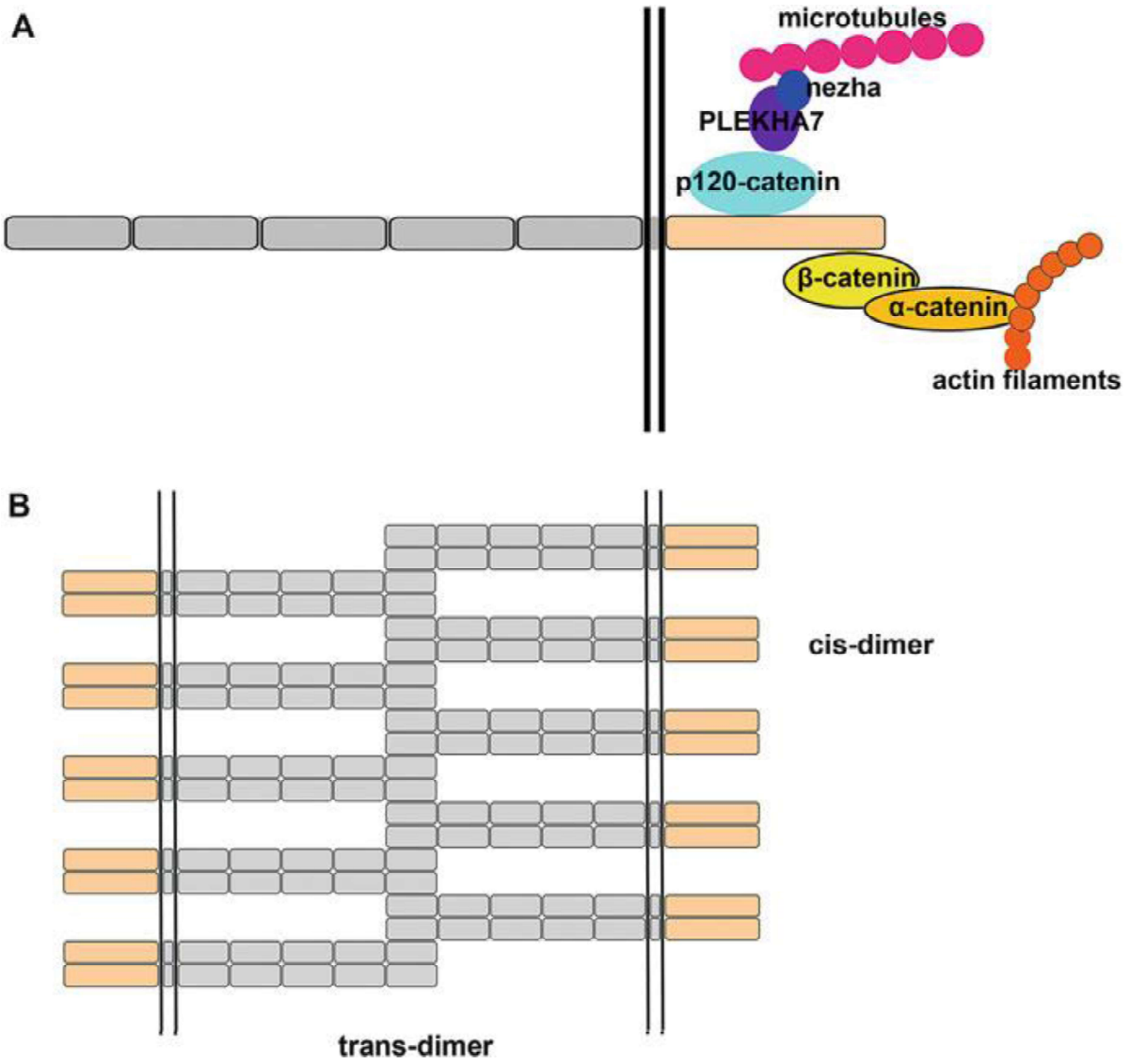


Figure 1.2. Cell adhesion complex of a classical cadherins *Panel A.* Basic cell adhesion complex of cell adhesion with interacting β-catenin, α-catenin, and actin filaments, *Panel B.* A model (zipper model) showing 'cis' and 'trans' dimers.

Soon after the discovery of the first cadherin, Takeichi's group discovered additional cadherins (Hatta and Takeichi 1986) and demonstrated that they actually represent a protein family, which came to be called by names such as E-cadherin, N-cadherin, and P-cadherin, according to their origin in tissues such as epithelia, neurons, and placenta. Later, in Suzuki et al. identified other cadherins by using degenerate PCR with primers designed against the highly conserved cytoplasmic region of the cadherins (which interacts with the catenins) (Tanihara, Sano et al. 1994) and classified cadherins into type I and type II cadherins based

on their amino acid sequences. Koch et al. in 1990 observed that desmosomes also contain some cadherin-like proteins (Buxton, Cowin et al. 1993) which are called desmogleins and desmocollins. The cytoplasmic domains of these molecules were initially thought to be highly conserved, but Mahoney et al., (1991) demonstrated that a protein in *Drosophila melanogaster*, named Fat, contains multiple repeats like cadherins but there was no significant sequence homology seen between its cytoplasmic domains (Mahoney, Weber et al. 1991). It is well known that a large number of proteins of various types contain the cadherin-specific extracellular domain motif and that these constitute a large cadherin superfamily. Protocadherins form the largest subfamily in vertebrates, and these were initially named to denote the nonclassical cadherins, but now most people use this term for any cadherin that contains six or seven repeats of the cadherin motif.

The cadherin superfamily is categorized into four major sub-families. These are (a) classical cadherins, (b) protocadherins, (c) desmosomal cadherins, and (d) atypical cadherins.

a) *Classical cadherins*. The classical cadherins are characterized by the presence of a five domains-long extracellular region, a transmembrane region, and a cytoplasmic region that interacts directly with p120 and β-catenin, and indirectly with α-catenin, to link up to actin filaments through protein-protein associations. The classical cadherins are further sub-classified primarily into types, I, II, III and IV. Type I and type II classical cadherins are present only in vertebrates. The classification is done on the basis of the tissue within which they were first identified. In vertebrates, there are 6 type I cadherins and 13 type II classical cadherins. Type III classical cadherins are found both in vertebrates and invertebrates, but not in mammals. The three types of cadherins are further described in more detail. Type I classical cadherins consist of proteins like epithelial cadherin or E-cadherin which is known as CDH1, and neuronal (or neural) cadherin, known as N-cadherin, or CDH2. They contain a conserved HAV tripeptide motif and a conserved tryptophan at the second position in the most distal domain, EC1, located farthest from the plasma membrane (Shapiro, Fannon et al. 1995). The type I cadherins have five main members, which include CDH1 (E-cadherin, epithelial), CDH2 (N-cadherin, neuronal), CDH3 (P-cadherin, placental), CDH4 (R-cadherin, retinal), and CDH15 (M-cadherin, myotubule) (Suzuki, Tsunekawa et al. 2016). Type II classical cadherins consist of proteins like vascular endothelial (VE)-cadherin which

is known as CDH5, and kidney (K)-cadherin which is known as CDH6 (Gumbiner 2005, Leckband and Prakasam 2006). Type II classical cadherins are characterized by the presence of two conserved tryptophan residues at the second and fourth positions (Trp2 and Trp4) in the most distal domain, EC1, but they lack the HAV tripeptide which is present in the type I cadherins (Shapiro et al. 2009). There are currently 13 named type II cadherins, these being CDH5 (VE-cadherin, vascular endothelium), CDH6 (K-cadherin, fetal kidney), CDH7, CDH8, CDH9 (T1-cadherin, testis), CDH10 (T2-cadherin, testis), CDH11 (OB-cadherin, osteoblast), CDH12, CDH18, CDH19, CDH20, CDH22, and CDH24. The type III classical cadherins possess a variable number of ectodomain repeats (Tanabe et al. 2004) and a conserved region called the primitive classical cadherin domain (PCCD) which lies between the cadherin repeats and the transmembrane helix. For the maturation of E-cadherin in Drosophila, proteolytic cleavage of PCCD is required (Oda et al. 1999), indicating that Type III and Type I cadherins interact. Type IV cadherins have seven ectodomains.

b) *Protocadherins.* More than 80 members of the cadherin superfamily together constitute another group of cadherins known as protocadherins. The protocadherins possess 6 or 7 ectodomain repeats. Their expression is dominated in the developing and mature vertebrate nervous system, although low levels of expression are also seen in the lungs and the kidney. These are highly conserved amongst the protocadherin subgroup, but show low sequence homology with other members of the cadherin superfamily (Hulpiau and Van Roy 2009). In addition to ectodomains, the protocadherins have a single pass transmembrane domain and a very distinct and specific cytoplasmic domain which, however, lacks motifs for catenin binding (Nollet, Kools et al. 2000, Vanhalst, Kools et al. 2005). These are further classified into clustered and non-clustered protocadherins based on their genomic organization.

c) *Desmosomal cadherins*. These types are characterized by a highly conserved extracellular region consisting of five repeat domains like the classical cadherins (Boggon, Murray et al. 2002, Delva, Tucker et al. 2009, Shapiro and Weis 2009). The cytoplasmic domain in these cadherins interacts with the β-catenin-related protein, armadillo, with plakoglobin, and also with the plakophilins which are associated with the intermediate filaments (Hatzfeld 2007, Al-Amoudi, Castaño-Diez et al. 2011). These cadherins known as the desmogleins and desmocollins are highly expressed in epithelial tissues and cardiac muscle (Nollet, Kools et

al. 2000, Hulpiau and Van Roy 2009). Desmosomal cadherins display both homophilic and heterophilic interactions (Green and Simpson 2007, Thomason, Scothern et al. 2010). Desmosomal adhesion is crucial for the stability of adhesion junctions in epithelial cell sheets and the regulation of epidermal differentiation (Garrod, Merritt et al. 2002).

d) *Atypical cadherins.* These are required for planar cell polarity signalling (Halbleib and Nelson 2006). Unlike classical and desmosomal cadherins, each of which has 5 extracellular or ectodomains (EC), Ds and Fat are characterized by the presence of 27 and 34 extracellular repeat domains, respectively. The cytoplasmic domains of Ds and Fat show sequence homology with the catenin-binding motifs present in classical cadherins (Mahoney, Weber et al. 1991). The main atypical cadherins are Dachsous (Ds), Fat, and Flamingo (Fmi). In mammals, Fat1 interacts with Ena/VASP, a family of proteins involved in regulation of actin cytoskeleton assembly and dynamics (Siemens, Lillo et al. 2004, Tanoue and Takeichi 2004). Fmi is unique as it contains a seven-pass transmembrane region and nine extracellular domain repeats (Nakayama, Gardner et al. 1998), causing it to be one of the most atypical cadherins yet known. Another much-discussed atypical cadherin is cadherin-23, or cdh23, which has 27 extracellular repeat domains, and plays a role in hearing involving stereocilia (Siemens, Lillo et al. 2004).

Invertebrate cadherins. After advancements in DNA sequencing techniques, sequence comparisons have helped to trace the origins of the cadherins. In organisms like Branchiostoma floridae (lancelet), Nematostella vectensis (sea anemone) and Trichoplax adhaerens (primitive placozoan) which occupy key positions in studies of metazoan evolution, sequencing reveals the presence of 30, 16 and 8 cadherin genes (or cadherins-like genes), respectively, in their genomes. The genome of the sea urchin, Strongylocentrotus purpuratus, contains 14 cadherin-like genes. The worm, Caenorhabditis elegans, contains has 12, and the fly, Drosophila melanogaster, has 17. In the closest known relative of the metazoans, Monosiga brevicollis (a unicellular nonmetazoan choanoflagellate), 23 putative cadherin-like genes have been identified (Murray-Stewart, Woster et al. 2014). Members of cadherin families, lefftyrin, coherin, and hedgling, were present in the last common ancestor of choanoflagellates and metazoans. Mainly present in choanoflagellates and sponges, these may have evolved by domain shuffling and lateral gene transfer. These genes are speculated

to have adhesive functions in these organisms (Nichols, Townsend et al. 2012). Cadherins containing extracellular domain repeats are linked to Src homology 2 (SH2), Hedgehog N-terminal peptide (N-hh), immunoglobulin (Ig) and von Willebrand type A domains are seen in M. Brevicollis and Amphimedon queenslandica. Cadherins which are now classified as Fat cadherins are also observed in sponges and sea urchins. The conserved cadherin cytoplasmic domain containing β-catenin binding sites is also observed in N. Vectensis (Abedin and King 2008). The function of cadherins in these unicellular organisms is mostly unknown, but they are found to be localized in the apical collar and basal pole of these cells. They play a role in cell shape and polarity and facilitate intracellular processes by taking cues from the extracellular environment.

Human cadherins. The human genome encodes 114 cadherins. Although many arise through alternative splicing of mRNA, the presence of such a sizeable repertoire of genes has caused them to be classified as 'cadherin main branch' and 'cadherin-related major branch'. The cadherin-related major branch mainly consists of protocadherins whereas the cadherin main branch consists of classical (type I) cadherins and atypical (type II) cadherins. Evolution of such a large superfamily of proteins appears to have mainly resulted from whole genome duplications, individual gene duplications and diversification of duplicated genes. Type I cadherins consist of CDH1/E cadherin/Epithelial cadherin, CDH2/N cadherin/Neuronal cadherins, CDH3/P cadherin/Placental cadherins, CDH4/R cadherin/Retinal cadherin, and CDH15/M cadherin/Myotubule cadherins. Each of these consists of the same number (five) of extracellular domains and a highly conserved tryptophan at position 2 (used for adhesion) and cytoplasmic domains are used for association with other proteins of the armadillo family and β catenin (Hulpiau and Van Roy 2010).

Origins of the five ectodomains. Bioinformatics-based analyses of DNA and protein sequences from divergent organisms reveals that an ancestral five repeat cadherins gene arose before divergence into paralogs. Repeated duplication of the extracellular domains appears to have led to the formation of a classical cadherin prototype in which introns were inserted because the introns in all cadherin genes are present in the same locations. After divergence of this basic linear gene structure of the cadherin gene into different organisms, mutations could have occurred at fixed rates. Of the five classical cadherins, N cadherins

show the least rate of change because of a selection pressure placed on it due to its presence in nervous systems. The existence of duplicates of gene paralogs could generate a broader scope for intragenomic recombination. It could also lower selection pressure on copies due to greater redundancy. The somatic morphology of organisms changes dramatically in vertebrates. So, the E cadherins are placed under much less selection pressure and appear to have evolved faster than N-cadherins (Gallin 1998).

The domains I and II of E-cadherins only share 25% sequence identity with each other whereas domains III, IV and V show no significant similarity in their sequences. Conservation is observed at different residue positions amongst different domains. Position A′5 is occupied by Glu in all sequences of domains I, II, and III, while this position is shared by Gln and Asp in case of domains IV and V. A'1 position in EC I of all cadherins is shared by hydrophobic residues and EC II share Gly and Ala residues. The position D1 is also conserved for hydrophobic residues for the domain I and conserved for hydrophobic and aromatic position for domains II and IV. Domain V is the least conserved domain amongst classical cadherins. The gaps (deletions and insertions) in sequences of ectodomains are almost always found at the borders of strands or helices (Kister, Roytberg et al. 2001). Various studies have suggested that the five and seven extracellular domains of type I and type IV cadherins have independently evolved from a common ancestral cadherin that is represented by type III cadherins. EC1 of type I cadherin and the EC6 of type IV cadherin appear to have evolved from the same extracellular cadherin domain in the common precursor (Uchida, Sivaraman et al. 2016).

Cadherin regulation. It is pronounced that cadherin-based adhesion is required to be very dynamic in order to perform its designated roles. During developmental processes, increase in cadherin expression mediate tissue condensation or compaction (Haas and Tuan 1999, Astick, Tubby et al. 2014), whereas decrease can lead to cell dissociation, for example during the EMT (Yang and Weinberg 2008, Thiery, Acloque et al. 2009). It is very interesting to note that losses of cadherin expression during the EMT are usually followed by an increase in expression of a different cadherin. Typically, for the transportation of epithelia, a switch from E- cadherin to N-cadherin (Wheelock, Shintani et al. 2008) and during the formation of the neural crest, a switch from N-cadherin to cadherin-6 (Park and Gumbiner 2010)

occurs. The EMT is a major change in the differentiation of the cell, and numerous other factors can control how the cadherins are utilized. It is also evident that cadherins also regulate cell signalling pathways and these differences can account for their different properties/cell behaviours.

Signalling by classical cadherins. It has been found that certain signalling events can induce cadherin internalization which could be a way to increase turnover due to decreased gene expression. The p120-catenin has been found to be an important regulator of cadherin internalization. Loss of p120-catenin expression leads to much-reduced levels of cadherin protein expression, apparently due to its control of cadherin endocytosis (Davis, Ireton et al. 2003, Xiao, Garner et al. 2005, Kang, Jenabi et al. 2007). The exact mechanism behind the removal cadherins from the cell junctions to allow internalization is not known.

Classical cadherins interact and stimulate signalling processes in many different ways (Figure 3). In order to migrate, it is required to control cell motility by allowing cells to dissociate from a tissue. It also includes facilitating the formation of other cell junctions, including tight junctions and desmosomes and associated generation of cell polarity, which in turn regulate many signalling events in the cell. As a direct interaction, classical cadherins interact directly to a major nuclear transcriptional effector protein, β-catenin (McCrea, Turck et al. 1991, Peifer, McCrea et al. 1992). The central regulator is the Wnt pathway, in which Wnts act as extracellular ligands via Frizzled and LRP transmembrane receptors to inhibit the degradation of cytosolic and nuclear β-catenin via a specialized destruction complex (Clevers and Nusse 2012). However, the levels of cytosolic and nuclear β-catenin can be controlled independently. Cadherins also interact directly with other proteins known to mediate different classical signalling pathways. Cadherin-catenin complexes have been found to associate with transduction modules, including numerous kinases, phosphatases, GTPases, and GEFs, and so on, that can transduce a variety of different signals (Wheelock and Johnson 2003, McLachlan, Kraemer et al. 2007, Dejana, Tournier-Lasserve et al. 2009, McCrea, Maher et al. 2015). Cadherins have often been found to interact specifically with certain types of cell surface signalling receptors, that is, E-cadherin with EGFR, N-cadherin with FGFR, and VE-cadherin with VEGFR (Carmeliet, Lampugnani et al. 1999, Suyama, Shapiro et al. 2002, Qian, Karpova et al. 2004, Curto, Cole et al. 2007, Rudini, Felici et al.

2008). Experiments with E-cadherin protein coated-bead attachment to isolated cells demonstrated that E-cadherin mediates contact inhibition independent of the formation of any other cell interactions (Perrais, Chen et al. 2007). This also showed that a homophilic bond is essential for contact inhibition signalling, in contrast to the inhibition of β-catenin by sequestering out of the nucleus (Gottardi, Wong et al. 2001, Simcha, Kirkpatrick et al. 2001). Cadherin-mediated contact inhibition has interesting implications for patterns of tissue growth and appears to work by shifting the dose dependence of cell-proliferation.

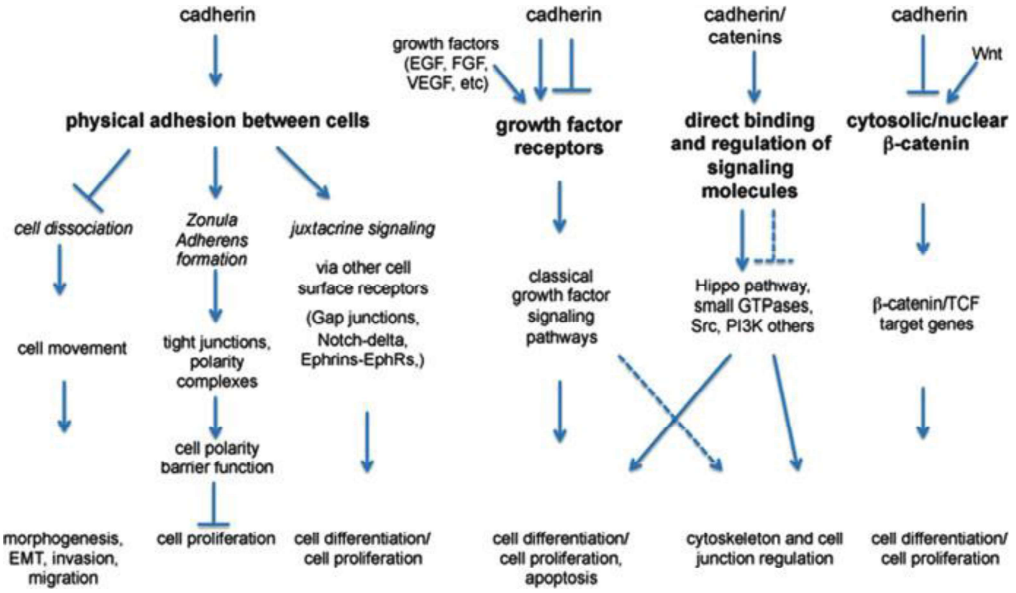


Figure 1.3. The diversity of ways that mediated cadherin signalling by physical adhesion, interaction through cell surface growth factor receptors and via catenins.

Structure and function of cadherin extracellular regions.

Extracellular regions of cadherins are characterized by the presence of distinctive protein domains of ~100 amino acids called extracellular cadherin (EC) domains (Hatta, Nose et al. 1988, Overduin, Harvey et al. 1995, Shapiro, Fannon et al. 1995). The topology of EC domains is similar to immunoglobulin domains although arrangements of their hydrophobic core residues are different (Shapiro, Fannon et al. 1995). Two β -sheets are formed by seven strands; with one sheet formed from strands ACFG, and the other by strands BED. The N-

terminal A strand enters at the "top" of the domain, whereas the C-terminus of the final G strand exits through the "bottom", with the long axis of the domain running parallel to the β-strands.

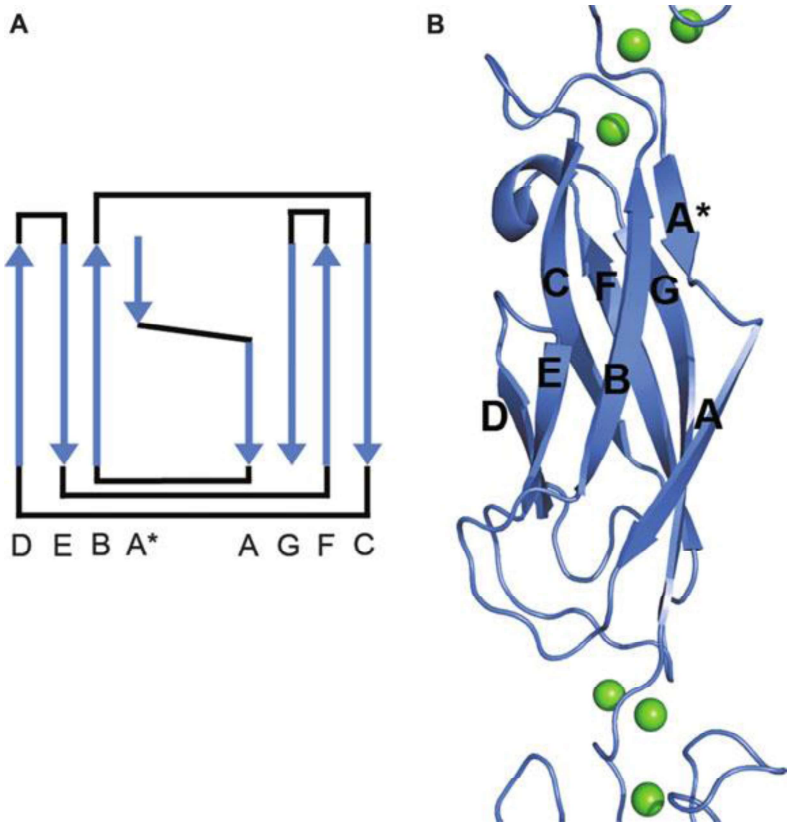


Figure 1.4. The folding topology of an EC domain is shown schematically in (a). EC1 is similar to an immunoglobulin domain (A strand can also associate with either sheet in different cases), and the immunoglobulin β-strand nomenclature is used. (b) shows

Vertebrate classical cadherins. Extracellular domains of classical cadherins interact with opposing cell surfaces and form trans homodimers through their N-terminal EC1 domain. The structure has been characterized by X-ray crystallography (Shapiro, Fannon et al. 1995, Boggon, Murray et al. 2002, Häussinger, Ahrens et al. 2004, Harrison, Jin et al. 2011). All classical cadherins share a common dimerization mechanism in which then a strand present at N-terminus (A strand) enters into the hydrophobic pocket of the EC1 domain of opposing cadherin and *vice versa*. This interaction is called '3D domain swapping'(Bennett, Schlunegger et al. 1995).Three Ca^{2+} binding sites are present at interdomain linkers between

each set of EC domains. Glutamate 11 is conserved in all classical cadherins which are responsible for binding of Ca^{2+} and created a strain the N terminus by Trp2 docking. This strain destabilizes the closed monomer and thus favours strand-swapped dimer formation (Harrison, Jin et al. 2011, Vunnam and Pedigo 2011).

Now, this is clear from the above reports that homodimerization is the central event in the particular cell-cell adhesion in case of classical cadherins. The mechanism in detail at the molecular level has been studied intensively in the case of E-cadherin (Brasch, Harrison et al. 2012, Troyanovsky 2012, Sivasankar 2013). An intermediate transient conformation called X-dimer introduces the formation of the strand swapped dimer which basically reduces the activation energy to form a strand-swapped dimer (Hong, Troyanovsky et al. 2011, Luo, Li et al. 2013). These finding has dramatically enhanced the understanding of the dimerization in general.

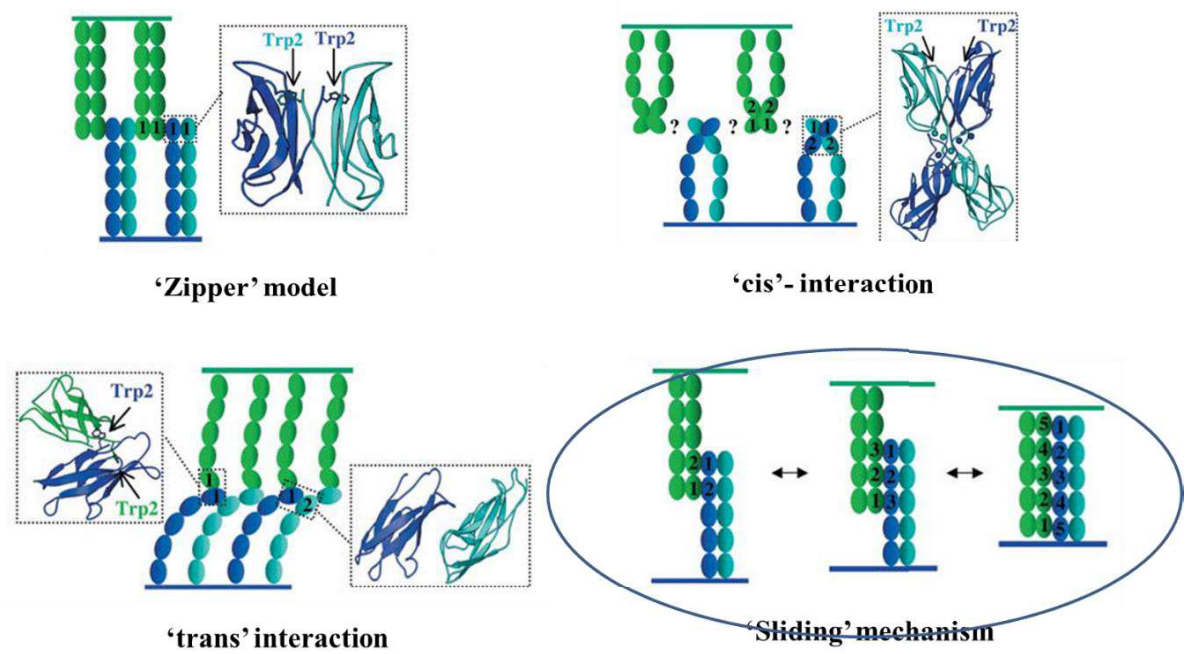


Figure 1.5. Evolution of structure-based models: cis and trans interactions

Several models have been proposed to describe the mode of dimerization depending on the available information present at that time. The 'zipper model was a famous model representing the interaction of domain 1 with another molecule with its domain 1, and simultaneously cis and trans conformations were discovered (Koch, Manzur et al. 2004).

In all the models, authors have explained the role of domains 1 and 2, but no one really has explained the roles of the other domains. Why are they present at all, and what are their possible roles during dimerization at the cell surface? Although it has been accepted by most that the only EC1 has a role in interactions and that the other domains are just there for supporting roles, there is no certainty regarding this. We also followed the same hypothesis initially, but the presence of the other domains made us think about their possible roles. Our lab has expertise in protein folding-unfolding, structure-function correlations, deciphering molecular mechanisms by looking into chemical and thermodynamic stability *et cetera.*

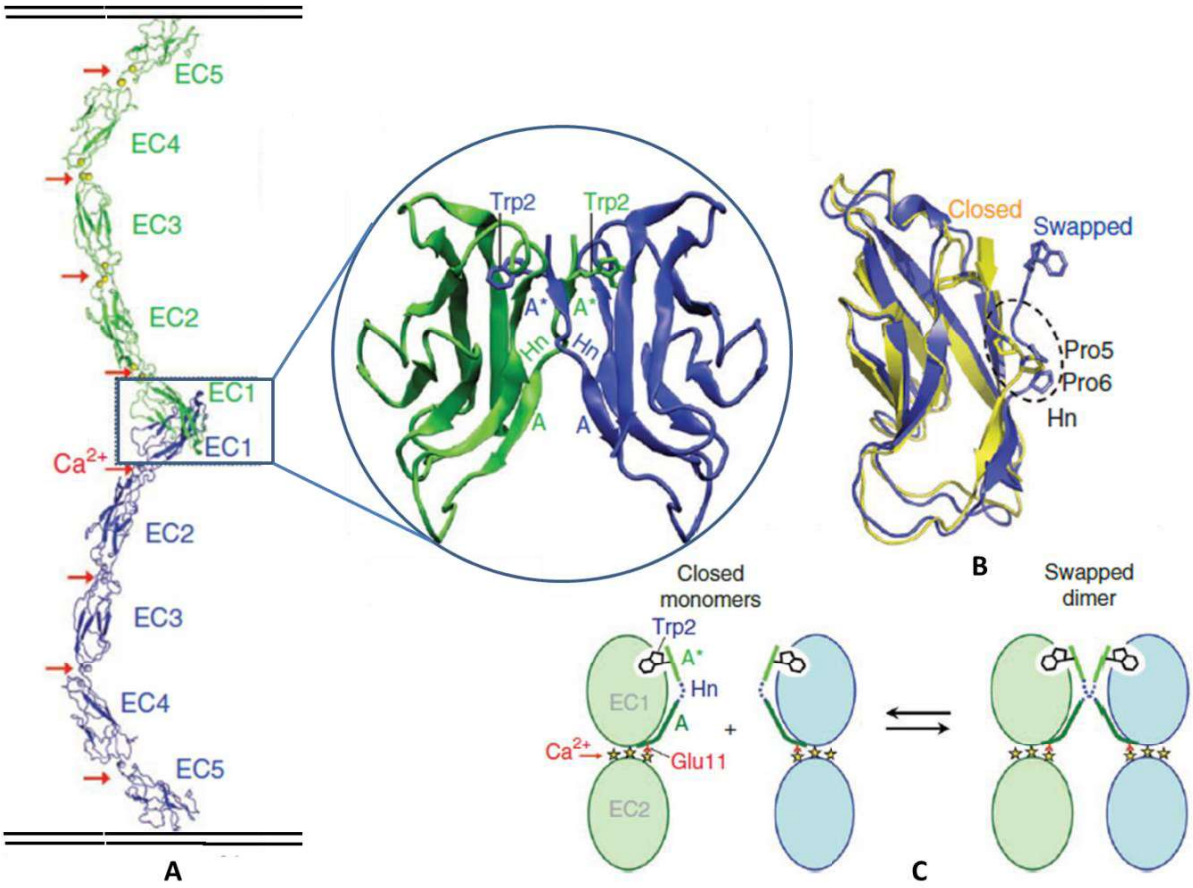


Figure 1.6. Domain/Strand Swapping is the only model of cis or trans-interactions with some supporting evidence *Panel A.* Crystal structure of E-cadherin dimer depicting domain swapping, *Panel B.* Closed and swapped conformations with the role of Prolines, *Panel C.* Schematic figure showing closed and swapped dimers.

Thus, looking towards these biophysical aspects, we decided to address the roles of all the 5 domains of E- as well as N-cadherins, by producing and studying them at the molecular level, in isolation, and also in the company of their neighbouring domains, and by examining in some depth their structure, stability and other biophysical properties.

We cloned and expressed all the individual domains and their fusion domains of E- and N-cadherins in various combinations.

Which EC domains interact with which other ones? From the crystal structure and recent finding from other research groups, it is clear that EC1 interacts with EC1 of opposing cell forming a strand-swapped dimer. EC2 also participates in X-dimer formation which is considered as an intermediate conformation before the final interaction. But, we were interested in knowing what the roles of the other 3 domains why they are there are? Do they participate *in* cis interactions or play a supporting role in placing the E1 domains further away from the cell surface?

Are there any E- domain -N domain interactions (given the extreme structural similarities)? Sequence and structure alignment data show that there is significant similarity in the amino acid sequence and the structures of E- and N-cadherins. Despite their origins in different tissues and their roles in various events, do they exhibit any interaction? In other words, apart from homophilic interactions, do cadherins also show any signs of heterophilic interaction?

Which EC domains are structured and soluble when produced singly? Upon expression and purification after production in *E. coli*, which cadherin domains are soluble and structured when produced as individual domain constructs? Do their neighbouring domains have any role in their folding and solubility? How are their biophysical properties affected in the presence of their neighbouring domains?

Which EC domains are poorly structured, and show a tendency to aggregate? Among all the 5 domains, which ones have poor structure upon purifying by native purification and which of them are not able to fold and remain aggregation-prone?

Which EC domains form defined homo-multimers by themselves, e.g., E1, E2, E3 etc.? Cadherin EC domains are known to form homodimers; keeping this in mind, do they form other associated species leading up to protein aggregation?

Which EC domains undergo structural changes (associated with presumed rigidification) upon calcium binding? Cadherin EC domains possess three calcium-binding sites in each interdomain region. Which ones of these show any binding to calcium and show structural changes thereupon?

Which combinations of EC domains (e.g., E1-E2 or E1-E2-E3) show binding to calcium? While designing the individual domains constructs we did not disturb any of the calcium binding sites. Therefore, we wished to see how individual domains are different from fusion domains in terms of the binding of calcium and the effect of this binding.

Do E1-E1 interactions occur with isolated E1 domains? Similarly, E2-E2? What about E2-E1? Using protein-protein interaction techniques we could potentially study whether homo-dimerization occurs and which domains play a crucial role. As we know, domains function independently. So, we wished to determine whether their binding with their counterpart-domains requires two domains constructs or higher (E1 or E1-E2, or E1-E2-E3). How is binding affected by longer fusion constructs?

What are the structural stabilities of the domains and combinations of domains like? How does the presence of the neighbouring domains affect the structural stability of any domain? Does their presence stabilize or reduce the stability?

Can these domains undergo facile unfolding and refolding? Is there a calcium-dependence? As we have discussed in the early part of this section, cadherins are designed to perform mechanistic as well as signalling functions, and for that, they would be required to be highly dynamic in terms of their replenishment on the membrane. We wanted to check whether the EC domains are able to refold back to their original structure when they get denatured in

physically stressing conditions (e.g., one cell pulling away from another) or under any other adverse conditions. If so, i.e., if refolding occurs, this could be very useful for cells since they would not need to produce fresh protein and replenish it each time some physical denaturation happened to occur.

Are homophilic contacts between N-cadherins less strong than those between E-cadherins? Despite reasonably large sequence and structural similarity between E- and N-cadherins, their physiological function requires different properties of the ectodomains. The surface area of neurons close to their neighbouring cells is smaller than that of epithelial cells and to make strong connections the affinity could be higher than that of E-cadherin. In case of E-cadherin as they have a larger contact surface like a 'velcro' if the affinity is lesser, then the avidity would be sufficient enough to tighten the connection without wasting much energy. Therefore, it would be interesting to study the protein-protein interactions and compare their dissociation constant values (k_d).

Are the different E- and N-domains different in their ability to unfold and refold? If the EC domains possess the ability to refold then how does this vary between E- and N-cadherin domains, as both the cadherin types perform very distinct function despite having significant sequence and structural similarities. Different kinds of cadherins have evolved for their function like neurons frequently make and break synaptic connections with other neurons whereas epithelial cells live longer in the same state and in some cases, they make the contacts which remains life-long. This biophysical property can be correlated in terms of the plasticity of EC1 and EC2 domains of E- and N-cadherins. This can be speculated that in order to make connections for memory formation, neurons frequently connect and separate. The rapid separation could result in mechanical forces which can cause partial unfolding of the extracellular domains of N-cadherin. By investigating their refolding ability, we could know whether this property can have a direct correlation with their physiology.

What might the answers to such questions tell us about the 'cis-dimer' or 'trans-dimer' or 'sliding-pentamers' hypotheses? During the process of cell adhesion, the same type of cadherin molecules EC1 domain interacts to the EC1 of opposing cell surface which is known as 'trans' dimer formation. However, a puzzling question still remains about 'cis' dimer formation whether it uses the same dimer interface which is used for 'trans' dimers or

a new dimer interface because both the types of dimers are formed simultaneously during the process of dimerization.

So, to address the above thoughts we decided to clone and purify the domains (all of them, and in all possible combinations, with/without affinity tags), to examine their structural-biochemical behaviour towards each other and also, to create 'artificial' cells for binding assays, from bead bearing only recombinant cadherins and 'bridge' cadherin constructs that could cause epithelial and neural cells to associate whether cadherin-bearing beads can be used in tissue-engineering and morphogenesis.

Chapter 2: Materials and methods

2.1. Materials

2.1.1 List of primers

All the primers which have been used for the cloning of any gene at any point of this thesis, are listed below in the table.

S. No.	Primer	DNA sequence (5'- 3')
1.	ECAD 1 F	ATAAGACATATGGACTGGGTTATTCCTCCGATC
2.	EEC 1 R	TATATACTCGAGCTGGGTGAATTCGGGCTTG
3.	EEC 1 R UT	TATATACTCGAGTTACTGGGTGAATTCGGGCTTG
4.	EEC 2 F	ATATAGCATATGGAATTCACGCAGGAGG
5.	EEC 2 R	ATTATTCTCGAGGGGATTGAAGACCGGAGG
6.	EEC 2 R UT	ATTATTCTCGAGTTAGGGATTGAAGACCGGAGG
7.	EEC 3 F	ATATATCATATGATCTTCAATCCCACCACGTAC
8.	EEC 3 R	ATATAGCTCGAGAGGCACAAAGATGGGGG
9.	EEC 3 R UT	ATATAGCTCGAGTTAAGGCACAAAGATGGGGG
10.	EEC 4 F	ATATACCATATGATCTTTGTGCCTCCTGAAAAG
11.	EEC 4 R	TATTATCTCGAGTGGTATGGGGGCGTTG
12.	EEC 4 R UT	TATTATCTCGAGTTATGGTATGGGGGCGTTG
13.	EEC 5 F	ATATATCATATGGCCCCCATACCGGAACC
14.	ECAD 5 R	TATATACTCGAGCTGTGCCTTCCTACAAACGCCAG
15.	EEC 5 R UT	TATATACTCGAGTTACTGTGCCTTCCTACAAACGCCAG
16.	NCAD 1 F	ATTATTGCTAGCGACTGGGTCATTCCTCC
17.	NEC 1 R	ATTAATCTCGAGGTGTAAGAACTCAGGTCTG
18.	NEC 1 R UT	ATTAATCTCGAGTAGGTGTAAGAATTCAGGTCTG
19.	NEC 2 F	ATTATAGCTAGCGAGTTCTTACACCAGGTTTG
20.	NEC 2 R	ATCTATCTCGAGGGCAGTAAACTCTGGAGG
21.	NEC 2 R UT	ATCTATCTCGAGTGAGGCAGTAAATTCTGGAGG
22.	NEC 3 F	AATATTGCTAGCGAGTTTACTGCCATGACG
23.	NEC 3 R	ATAATACTCGAGGGGGGCAAAATAAGGG
24.	NEC 3 R UT	ATAATACTCGAGTGAGGGGGCAAAATAAGGG
25.	NEC 4 F	ATTATAGCTAGCTATTTTGCCCCCAATCCTAAGATC
26.	NEC 4 R	ATTATACTCGAGCACTTGAGGGGCATTGTCATTAATATC
27.	NEC 4 R UT	ATTATACTCGAGTAGCACTTGAGGGGCATTGTCATTAATATC
28.	NEC 5 F	ATTATAGCTAGCGCCCCTCAAGTGTTACCTCAAG
29.	NCAD 5 R	ATAATACTCGAGCCTGTCCACATCTGTGCAGTCC
30.	NEC 5 R UT	ATAATACTCGAGTGACCTGTCCACATCTGTGCAGTCC

2.1.2 Vectors

Vector	Properties
pET23a	• C-terminal 6X-His tag • Ampicillin resistance • T7 promoter • Multiple Cloning Site (*Bam*H1-*Xho*I) • pBR322 origin • f1 origin

2.1.3 Bacterial Strains

Strains	Genotype/Properties
DH5α	• F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ– • Recombination-deficient • Endonuclease A deficient
XL-1 Blue (Stratagene)	• endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK-mK+) • tetracycline resistant (carried on the F plasmid)
TOP10	• F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ- • Streptomycin resistant
BL21(DE3) pLysS	• E. coli str. B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysS[T7p20 orip15A](CmR) • pLysS plasmid chloramphenicol resistant; grow with chloramphenicol to retail plasmid • The pLysS plasmid encodes T7 phage lysozyme (an inhibitor of T7 polymerase) which reduces and almost eliminates expression from a

	transformed T7 promoter containing plasmids when not induced.
Rosetta (DE3) pLysS	• E. coli str. B F– ompT gal dcm lon? hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysSRARE[T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL orip15A](CmR) • Chloramphenicol resistant • pLysS contains tRNA genes. The rare codons AGG, AGA, AUA, CUA, CCC and GGA are supplemented.

2.1.4 Chemicals and kits

Chemicals	**Company name**
DNA polymerases Restriction enzymes T4 DNA ligase	New England Biolabs (NEB), USA
DNA ladders	BR BIOCHEM Life Sciences
FastDigest Restriction enzymes	Thermo Fischer Scientific
Plasmid isolation kit, Gel extraction kit, PCR purification kit, Ni-NTA agarose	QIAGEN, Germany
Protein ladders	Thermo Scientific Pierce

2.1.5 Media (Luria Bertani)

Components	**Amount (for 1 L)**
Tryptone	10 g
Yeast Extract	5 g
Sodium Chloride	10 g
LB agar (for agar plates)	2%

The media was sterilized by autoclaving (15 psi for 15 minutes at 121°C).

2.1.6 Antibiotics and IPTG

All the stocks were prepared to 1000X and sterilized by syringe filters of 0.22 µm pore size. These antibiotics were stored as aliquots at -20°C.

Antibiotics	**Stock concentration**
Ampicillin	100 mg/ml in water
Chloramphenicol	35 mg/ml in methanol
Kanamycin	25 mg/ml in water
Tetracycline	12.5 mg/ml in 70% ethanol
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	1 Molar

2.1.7 Buffers used in molecular biology

Buffers used in preparation for chemically competent cells

Calcium chloride	60 mM
Glycerol	15 % v/v
PIPES	10 mM

The pH of the buffer was adjusted to 7.0 and sterilized by 0.22 µm syringe filters and further sterilized by autoclaving. The solution was stored at 4 °C.

50X TAE (1 L)

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

The pH of the buffer was adjusted to 8.0 and sterilized by 0.22 µm syringe filters which were further sterilized by autoclaving. The solution was stored at room temperature

6X DNA loading dye

Bromophenol blue	0.25 %
Glycerol	30 %

Ethidium bromide stock solution (1 % w/v)

Ethidium bromide	0.1 g
Deionized water	10 ml

2.1.8 Buffers and stock solutions for SDS-PAGE

Acrylamide solution (100 ml)

Acrylamide	30 g
N, N'-Methylene bisacrylamide	0.8 g

Acrylamide was carefully weighed inside the fume hood, and after it was mixed thoroughly, bis-acrylamide was mixed. The volume was made to 100 ml and stored at 4°C in amber coloured bottles.

Lower Tris (4X)

Tris	18.17 g
10 % SDS	4 ml

The pH was adjusted to 8.8, and the volume was made to 100 ml and after filter sterilization stored at room temperature.

Upper Tris (4X)

Tris	6.06 g
10 % SDS	4 ml

The pH of the solution was adjusted to 6.8.

5X Sample loading buffer

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

Laemmli buffer (100 ml)

Tris buffer	3.00 g
Glycine	14.4 g
SDS	1 g

Gel staining solution (100 ml)

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

Gel destaining solution (100 ml)

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

2.1.9 Buffers used for protein purification

All the buffers used for non-denaturing (native) and denaturing protein purifications were made according to Qiagen standard protocols given in the handbook.

For purification under native conditions

Lysis buffer	50 mM NaH2PO4 300 mM NaCl 10 mM Imidazole
Wash buffer	50 mM NaH2PO4 300 mM NaCl 20-50 mM Imidazole
Elution buffer	50 mM NaH2PO4 300 mM NaCl 250 mM Imidazole

The pH was adjusted to 8.0 using NaOH.

For purification under denaturing conditions

Lysis Buffer (pH 8.0)	100 mM NaH2PO4 10 mM Tris-Cl 6M GuHCl/8M Urea
Wash buffer (pH 6.3)	100 mM NaH2PO4 10 mM Tris-Cl 6M GuHCl/8M Urea
Elution buffer (pH 4.5)	100 mM NaH2PO4 10 mM Tris-Cl 6M GuHCl/8M Urea

Due to dissociation of urea, the pH of the buffers for denaturing buffers was adjusted immediately before use.

2.2. Methods

2.2.1 Polymerase chain reaction (PCR)

Standard PCR reactions were set according to NEB protocol:

Component	**Stock concentration**	**25 µl reaction**
10 X Standard *Taq* Reaction buffer	10X	2.5 µl
dNTPs	25 mM	0.25 µl
Forward primer	100 µM	0.25 µl
Reverse primer	100 µM	0.25 µl
Template DNA	variable	variable
DNA Polymerase enzyme	-	0.1 µl -0.25 µl
Nuclease-free water	-	To 25 µl

Steps	**Temperature**	**Time**
Initial denaturation	95 °C	5 minutes
25-35 cycles	95 °C 55-65 °C 72 °C	30 seconds 30-45 seconds 1 minute per kb
Final extension	72 °C	10 minutes
Hold	4 °C	forever

The PCR reaction was set using autoclaved microcentrifuge tubes on ice. Then the reaction was gently mixed by pipetting in and out. The PCR reactions were analyzed by running 1% agarose gel.

2.2.2 Agarose gel electrophoresis

1% agarose powder was weighed and mixed into 1X TAE buffer by boiling it for 1-2 min. The solution was allowed to cool at room temperature. Ethidium bromide (0.5 µg/mL) was added to the solution to stain nucleic acid. The gels were visualized under UV-transilluminator, and then the images were documented by EZ images (Biorad).

2.2.3 Extraction of DNA fragments from the gel

The desired DNA bands were excised with a fresh scalpel and collected into a 2 ml microcentrifuge and following steps were performed.

1. 3 volumes of QG buffer was added according to the dry weight of the gel pieces. (e.g., for 100 mg. 300 µl of QG buffer) and incubated at 50 °C until the gels pieces have dissolved completely.

2. one gel volume of isopropanol (100%) was added to the sample and mixed by inverting the tube.

3. The spin column was placed in the collection tube. For the binding of the DNA to the column, the sample was passed through the column and centrifuged for 1 min. The flow-through was discarded.

4. All centrifugation steps were carried out at 13,000 rpm in a conventional microcentrifuge.

5. 750 µl Buffer PE was added and centrifuged it for 1 min. The flow-through was discarded. The spin column was placed back to the collection tube and centrifuges for 2 mins to remove any residual wash buffer.

6. The spin column was placed into a fresh 1.5 ml autoclaved centrifuged tube.

7. The elution was done by adding 50 µl of autoclaved milliQ/ TE buffer to the centre of the spin column and again centrifuged for 1 min. The purity of the DNA was checked on 1% agarose gel and stored at -20 °C.

2.2.4 Quantification of DNA

The purified DNA fragments were quantitated by either measuring absorbance at 260 nm or visually by running on an agarose gel along with a DNA ladder of known concentration. In the case of the absorbance method, any RNA/protein contamination was checked by collecting the ratio of OD_{260}/ OD_{280}, which should be in the range of 1.8-2.0.

2.2.5 Restriction Digestion

The gene of interest (insert) and the vector both were digested by appropriate restriction enzymes to create a staggered cut and the ends. The enzymes called FD (FastDigest) were purchased from ThermoScientific. The standard reaction is as follows.

Components	**Volumes**
10X FD buffer	3 µl
Vector/insert	1 µg /100 ng
Enzyme I	1 µl
Enzyme II	1 µl
MilliQ	Adjusted to 30 µl

The reaction was incubated for 30 minutes at 37 °C. Afterwards, 0.8 % agarose gel was run to separate the samples, and the eluted samples were purified and quantitated.

2.2.6 Ligation

Once the cohesive ends were generated, the ligation reaction was performed to join the cohesive ends of vector and insert. T4 DNA ligase catalyzes the formation of a phosphodiester bond between 5’ phosphate and 3’ hydroxyl termini in double-stranded DNA or RNA.

Components	Volume (20 µl)
T4 DNA buffer	2 µl
Vector DNA (3.6 kb)	50 ng
Insert DNA (1 kb)	variable
T4 DNA Ligase	0.5 µl
Nuclease-free water	variable

The ligation reaction was set up by keeping a molar ratio 1:3 vector to insert.

$$\text{ng of insert required} = \frac{50 \text{ ng (vector)} * 1 \text{ kb (insert)} * 3}{3.6 \text{ kb (size of the vector)}}$$

The ligation reaction was kept for incubation at 25 °C for 1 hr.

2.2.7 Preparation of *E.coli* competent cells (chemical)

All the competent cells were made by using calcium chloride method which is following.

1. The required strain was inoculated into 10 ml LB media with appropriate antibiotics and kept overnight at 37 °C in shaking incubators.
2. 1% of the fully-grown culture was inoculated into 200 ml LB media and incubated at 37 °C until the OD_{600} reached 0.3-0.4.
3. The culture was kept on ice for 20 minutes (from now on the cold form was maintained).
4. The cells were collected by centrifugation in pre-chilled bottles at 1500 g for 10 minutes, and the supernatant was discarded.
5. 10 mL of pre-chilled calcium chloride solution was added to the cells and re-suspended by gentle pipetting in and out.
6. The cells were again collected by centrifuging at 1500 g for 5 minutes at 4 °C.
7. Step 5 was repeated, and the cells were kept on ice for 30 minutes.
8. The centrifugation was done at 1100 g for 5 minutes to settle the cells.
9. Finally, the cells were re-suspended again in 1.5 mL of calcium chloride solution. 100 µl was aliquoted into 1.5 mL autoclaved and pre-chilled 1.5 mL MCTs. The aliquots were immediately stored at -80 °C till further use.

2.2.8 Bacterial Transformation

This process involved the introduction of foreign DNA (in this case ligation product) into a bacterial competent cell. Specific treatments are given to increase the transformation efficiency and make bacteria more susceptible for either chemical or electrical based transformation these are commonly referred to as 'competent cells'. This process involved the following steps.

1. The competent cells were taken out of -80 °C and thawed on ice (approximately 15-20 minutes).
2. The ligation product was mixed into 100 µl of competent cells. Then the microcentrifuge tube was gently mixed by flicking the bottom of the tube by fingers for 3-4 times.
3. The mixture was incubated for 20-30 minutes on ice.
4. Heat shock step was performed by placing the bottom of the MCT into a preheated 42 °C water bath for 90 seconds.
5. The tube was kept back on the ice for 2-3 minutes.
6. 900 µl of autoclaved LB media (without antibiotic) was added to the tube and incubated in 37 °C shaking incubator for 45-60 minutes.
7. The tube was centrifuged for 5 minutes at 2500 rpm to settle down the bacteria. 950 µl of the supernatant was discarded by micropipette, and the remaining pellet was resuspended gently and plated on the appropriate LB agar plate with antibiotics.
8. The plate was incubated at 37 °C overnight.

2.2.9 Screening and confirmation of transformants

The isolated colonies appearing on the plate were screened by a form of PCR called colony PCR. The concept is just like a routine PCR but the colonies are used as a template (including genomic and plasmid DNA present within cells), and the primers are vector specific primers. The PCR products are run on agarose gel.

After the successful transformation, confirmed by colony PCR, transformants were further confirmed by restriction digestion. In this experiment the plasmid was isolated for the

screened transformation and digestion was set up using the two restriction enzymes which were initially used to clone the gene.
Once the insert of the correct size was confirmed on an agarose gel, the DNA sequence was confirmed by DNA sequencing.

2.2.10 Protein induction and expression of recombinant proteins in *E.coli*

In all cases, the plasmids containing the desired genes were transferred into the expression host Rosetta DE-3. A glycerol stock was prepared for all the confirmed cloning as well as expressing clones where1500 µl of overnight grown culture was mixed with autoclaved 60 % glycerol and stored at -80 °C.
The level of expression was checked by adding different concentrations of IPTG in secondary cultures at an $OD6_{00}$ value of 0.6. Then the level of induction and the solubility of the proteins were compared on SDS-PAGE by running the pellet and supernatants of un-induced as well as induced samples. In most of the cases, 0.5 mM IPTG was used to induce the expression and reduced the temperature to 25 °C post-induction for 8 hours.

2.2.11 Improving the solubility during protein expression

Proteins undergo competition between folding and aggregation, as in many cases overexpressed recombinant proteins accumulate intracellularly as insoluble aggregates which are also called inclusion bodies. Usually, they are inactive and misfolded, but to make them functional, various strategies to refold polypeptides extracted from inclusion bodies have been attempted, which are described below.
1. Lowering the growth temperature after induction to decrease the rate of protein synthesis.
2. Lowering the inducer (e.g., IPTG) concentration.
3. Adding 1 % glucose to reduce the induction of the lac promoter by lactose which is present in LB media.
4. A soluble fusion partner can be attached to the N-terminus of the heterologous protein.

2.2.12 Protein purification (Native/non-denaturing)

The following steps were followed for the protein purification.

1. The bacterial cells were harvested by centrifuging at 8,000 rpm for 7 minutes.
2. The cells were re-suspended in lysis buffer at 2-5 ml per gram wet weight. Alternatively, lysozyme was added to 1 mg/ml and incubated on ice for 30 minutes.
3. The partially lysed cells then subjected to sonication equipped with a microtip.
4. The lysate was then centrifuged at very high speed (12000 rpm) for 45 minutes to pellet the cellular debris.
5. The supernatant was separated in a fresh 50 ml falcon and passed through Ni-NTA column containing resins. Flow-through was collected.
6. Washing was done with native wash buffer, and fractions were collected.
7. The protein was eluted by passing elution buffer which was almost 4 times the volume of resin. The fractions were collected for SDS-PAGE analysis along with other samples.

2.2.13 Protein purification (Denaturing)

1. The cell pellet was re-suspended into buffer B at 5 ml gram wet weight and lysed them gently by vortexing and taking care to avoid foaming.
2. Centrifuged at 12000 rpm for 45 minutes to pellet the cellular debris.
3. The supernatant was collected in a fresh falcon and passed through a pre-equilibrated Ni-NTA column.
4. Washed with buffer C for 5-10 column volumes.
5. The bound protein was eluted in buffer E for 4-5 column volumes. The proteins usually elute in the second and third column volume.

2.2.14 *In vitro* denaturation and refolding of proteins

When all the efforts of expressing the target protein in soluble form result in the formation of inclusion bodies (IBs) despite such attempts, then the last option is to denature and refold the protein *in vitro*. This procedure involves solubilization and denaturation of the IBs in the

presence of a strong denaturing agent (8M urea or 6M guanidine) under reducing conditions (e.g., 20 mM DTT). Then the protein is refolded by removal of the denaturant by one of the following methods.

On-column refolding.

The protein was allowed to bind to the Ni-NTA column in the presence of denaturant. During the washing steps the concentration of denaturant was gradually lowered to 2M, 1M, 0.5M and finally to 0.1 M. In the buffer containing 0.5 M and 0.1 M, 5 % glycerol was maintained to avoid protein aggregation. The protein was eluted in the native elution buffer which also included 5 % glycerol.

Flash dialysis/dialysis.

This method is the most common method to remove for protein refolding. The denaturant concentration was slowly decreased which allows the protein to refold optimally. The concentration of urea in the dialysis buffer was reduced in a step-wise manner (8M, 4M, 2M, 1M, 0.5M and finally to 0.1M). Below 1 M urea concentration, 5 % glycerol was maintained to avoid protein aggregation. In some cases, 250 mM imidazole was also maintained in the final dialysis buffer in order to improve protein stability.

2.2.15 Protein concentration measurement

The concentration measurement was done by UV-absorption method. The spectrum was collected in the range of 250-600 nm using Cary 50 UV-Vis spectrophotometer, and the scattering factor was removed by subtracting A_{280} from A_{310}. The measured absorbance was calculated by entering the protein sequence in ProtParam (ExPASy online server), and the protein concentration was calculated by putting observed A_{280} in the equation.

2.2.16 Confirmation of protein mass and identity (Mass spectrometry)

SYNAPT G2S-HDMS mass spectrometer (Waters) was used for intact mass determination and *de novo* protein sequencing. Intact masses of proteins were confirmed in ESI-TOF mode

by injecting the protein sample in water. The protein identity was confirmed by Peptide Mass Fingerprinting (PMF) where the denatured protein was subjected to tryptic digestion and the peptides generated were used for peptide mass matching with predicted masses, and also sequencing in MALDI-TOF mode was carried out for some tryptic peptides. The samples were mixed in 1:1 ratio with a matrix CHCA (α-Cyano-4-hydroxycinnamic acid) and spotted on a MALDI plate. The samples for peptide sequencing were prepared following standard protocol provided by Sigma Aldrich kit.

2.2.17 Peptide Mass Fingerprinting (PMF)

In- solution digestion

A ratio of 1:100 to 1:20 (w/w) of protein or peptide to the enzyme is recommended. The substrate was dissolved in 100 mM Ammonium bicarbonate (ABC) pH 8.5. Trypsin was added according to substrate amount, mixed well and incubated at 37 °C for 16-18 hours.

In-gel digestion

The desired protein band was excised out of a gel and cut into small pieces. The pieces were transferred to un-autoclaved microcentrifuge tubes.

1. These pieces were then dehydrated with 40-100 µL of 1 M Acetonitrile (ACN) solution. This step was repeated thrice.
2. 40-100 µL of 30 % ACN solution in 25 mM ammonium bicarbonate (ABC) was added to destain the gel pieces; The samples were incubated at 30 °C with mild shaking for 30 minutes. This step was again repeated three times.
3. After removing the solution mentioned above, the gel pieces were further dehydrated once with 40 µL of 1 M ACN.
4. Then, 50 µL of 25 mM Dithiothreitol (DTT) was added and incubated at 56 °C for 20 minutes.
5. After cooling the samples to room temperature and pipetting off the residual solution, 50 µL of freshly prepared 55 mM iodoacetamide (IDA) in 25 mM ABC was added for alkylation. The samples were alkylated for 20 minutes in the dark.

6. After this, the gel pieces were washed with de-ionized water to remove IDA and further dehydrated using 200 µL of 50 % ACN in 25 mM ABC for 5 minutes and then with 100 % ACN for 30 seconds.
7. To these de-stained and dehydrated gel pieces, 50 µL of trypsin was added and incubated at 37 °C for 16-18 hours.
8. After this, the supernatant obtained by centrifugation at 9500 rpm for 1 minute was subjected to MALDI-MS analysis.

2.2.18 Size exclusion chromatography

The eluted proteins obtained were subjected to size exclusion chromatography using AKTA purifier system from GE. The columns Superdex 75 10/300 and Superdex 200 10/300 GL were equilibrated with the appropriate buffer and the 500 µl sample volume was injected by running at a flow rate of 0.5 ml/min. The elution profile was compared with the standard calibration curve to determine the mass and the oligomeric population in the solution.

2.2.19 Ion exchange chromatography

Anion exchange chromatography was performed for the untagged proteins (without affinity tag).
The Q-Sepharose resin was taken using AKTA purifier system from GE. The protein was allowed to bind the column in the appropriate buffer and eluted by passing a salt gradient (1M NaCl in the buffer). The salt was removed by dialysis in the required buffer.

2.2.20 Circular Dichroism (CD) spectroscopy

CD is a sensitive tool to study the secondary structure and the conformational changes of macromolecules during various chemical and physical treatments. It is based on measurement of the differential absorptions of left- and right-handed circularly polarized light by optically active compounds at different wavelengths. This method was used to determine the protein secondary structure content (α-helix, β-sheet and random coil) as well as the thermal and chemical stability of the proteins. The spectra were collected using the

CD instrument Chirascan in the range of 200-250 nm. The sample was put in a cuvette of 0.1 mm path-length. The raw ellipticity was converted into mean residual ellipticity by putting in following formula

$$[\theta] = \frac{\theta * 100 * \text{MW}}{1000 * \text{protein concentration (mg/ml)} * \text{path length (cm)}}$$

Where

$$\text{MRW is Mean Residue Weight} = \frac{\text{The molecular weight of the protein (Da)}}{\text{Total number of amino acids in a protein}}$$

2.2.21 Fluorescence spectroscopy

Fluorescence is a process of photon emission which occurs during molecular relaxation from electronically excited states. The fluorescence emission is highly sensitive to the biochemical environment of the fluorophore, and the spectral changes resulting from solvent relaxation of natural fluorophore are important reporters of protein structure and folding. A natural fluorophore in proteins is tryptophan, and its environment in the protein provides insight into tertiary structural aspects of the protein.

The tryptophan emission spectra were collected in the range of 300-400 nm with the excitation wavelength of 295 nm using Cary Eclipse fluorimeter.

2.2.22 Differential Scanning Calorimetry (DSC)

DSC is a technique used to characterize the stability of a protein in its native form. It measures the heat change associated with the molecules when the thermal denaturation of a molecule happens at a constant rate. MicroCal VP-DSC instrument was used to perform the experiments. The reference and the sample cells were filled first with buffer in which the protein is purified and stored. The absorption of heat occurs when a protein unfolds, and this causes a difference in the rate of rising in temperature between the cells, which is compensated for by the instrument. The enthalpy of protein unfolding is the area under the concentration normalized DSC peak, and its unit is calories (joules) per moles. In some

instances, thermodynamic models can be fitted to the data to obtain Gibb's free energy (ΔG), the calorimetric enthalpy (ΔHcal), the van't Hoff enthalpy (ΔH_vH), the entropy (ΔS) and the change in the heat capacity (ΔCp) associated with the transition.

2.2.23 Isothermal titration calorimetry (ITC)

ITC is widely used to study a wide range of biomolecular interactions. In our experiments, we measured the interaction between cadherin domains and calcium, cadherin domains with other cadherin domains. The MicroCal ITC 200 from GE/Malvern was used in all interaction studies. The system directly measures the heat directly released or absorbed during biomolecular interactions. Heat is released or absorbed resulting from redistribution and formation of the non-covalent bond. ITC measures these heat changes by measuring the differential power, applied to the cell heaters, required to maintain zero temperature difference between the reference and the sample cells.

The ligand was injected into the sample cell as a pulse of 1 to 2µl injections until the saturation reached. Each injection of ligand results in a heat pulse that is integrated with respect to time and normalized for concentration to generate a titration curve of kcal/mol vs molar ratio (ligand/sample). The resulting isotherm is fitted to a binding model to generate the affinity (KD), stoichiometry (n) and enthalpy of interaction (ΔH).

2.2.24 Analytical Ultra Centrifugation (AUC)

Beckman Coulter XL-I Ultracentrifuge was used to perform the AUC experiments. It provides information on mass, size, and shape of the particles, whether they are composed of spherical or rod-like shapes. During an interaction event when a macromolecule changes its conformation, this slight change in conformation can also be measured by using this technique. AUC can perform an analysis of the concentration of the sample during centrifugation with the help of light detection devices which are preinstalled. At very high speed, due to the difference in force applied to the components primarily caused by mass and shape differences, the components separate out in layers forming boundaries in the solution.

The cells of the centrifuge rotor were cleaned with water and detergents and allowed to dry. The cells were then assembled, and the sample was loaded along with the buffer in their respective chambers. The cell and the counterweight both were weighed and placed in the rotor. The absorbance was measured at 280 nm, and the run was started at 40,000 rpm.

2.2.25 Dynamic Light Scattering (DLS)

Zetasizer Nano ZS90 model (Malvern Instruments) was used to determine the hydrodynamic radii of proteins studied, using dynamic light scattering. It is a non-invasive technique for measuring the size of particles and molecules in suspension. This technique measures the speed of particles undergoing Brownian motion which is influenced by particle size, sample viscosity, and temperature, by monitoring correlations in scattering and the rate at which they change over time, to infer particle diameter. The larger the particle is, the slower the Brownian motion becomes. Hydrodynamic diameter can be defined as the diameter of a hard sphere that diffuses at the same speed as the particles being measured. Hydrodynamic diameter will not only depend on the size of the particle but also the surface structure. The sample was placed in a cuvette and number of scans was determined by the software.

2.2.26 Glutaraldehyde crosslinking

Protein-protein interactions comprise the molecular mechanisms of complex biological processes. Chemical crosslinking offers a direct method of identifying protein-protein interactions. This technique involves the formation of covalent bonds between two proteins by using bifunctional reagents containing reactive end groups that react with functional groups such as primary amines and sulfhydryls of amino acid residues. Glutaraldehyde which is a homo-bifunctional reagent was used to study crosslinking of proteins. For this, 50 to 100 µg of protein was treated with 5 µl of 2-3 % freshly prepared solution of glutaraldehyde and incubated for 2-5 minutes at 37 °C (note: the reaction can be terminated by addition of 10 µl of 1M Tris pH 8.0, as the Tris reacts with the remainder of the crosslinker). The cross-linked proteins were solubilized by adding an equal volume of

Laemmli sample loading buffer and heated at 100 °C for 10 minutes, and SDS-PAGE was conducted.

2.2.27 Chemical and thermal denaturation

To examine the transitions between a folded state and unfolded state, chemical, as well as thermal denaturation, have been used in our experiments. Chemical denaturation studies were performed to determine the Cm of the protein, which is the concentration at which half the molecules in the population being monitored have undergone unfolding. The samples were incubated with progressively increasing concentrations of denaturant (8M urea/6M GdmCl) in separately prepared samples and incubated overnight at room temperature. After incubation, both CD and fluorescence data were collected and plotted to determine the extent of protein unfolding.

Chapter 3: Structural and sequence bioinformatic analysis of intra- and inter-cadherin domain relationship

Analysis of sequence and structure allows us to find conserved positions and residues specific to the structures for a given family. The conserved sequences have a crucial role, and sometimes these are also called as sequence determinants of a particular protein. In this section, we have described a comparison of sequence and structure of the five extracellular domains of E-and N-cadherins. The sequence alignment was done using Clustal X, and Emboss Needle server was used to calculate the percentage identity/similarity.

3.1 How similar are the domains of E-cadherin to each other?

The amino acid sequence alignments of cadherin EC domains was done. This alignment was started with the domains E1 to E5 together, where two conserved amino acids glutamate (E) at 14^{th} position at N-terminus and aspartate at 108^{th} position of C-terminus were found. These are present in all the 5 EC domains at the same position, near the linker regions, and are considered to be the sites of calcium binding. However, a few additional aspartates (D) and glutamate (E) amino acid residues are also present, but these are not believed to be involved in calcium binding. Two proline residues are present at the N-terminus of all the domains except E2. Hydrophobic amino acid residues at positions, 22, 25 and 27 in each domain are also conserved at the N-terminus, and in regions 84-90, and 101-106 at the C-terminus of each domain. The presence of these hydrophobic residues could be responsible for the similarity of fold amongst the EC domains of the E-cadherins. The sequence alignment of domains with each other in a pair-wise fashion showed that the identity is highest for E1 vs E1 and lowest between E1 vs E5. Since the five domains are considered to have arisen from the gene duplication of an ancestral domain, this analysis suggests that E5 might have evolved last, during evolution.

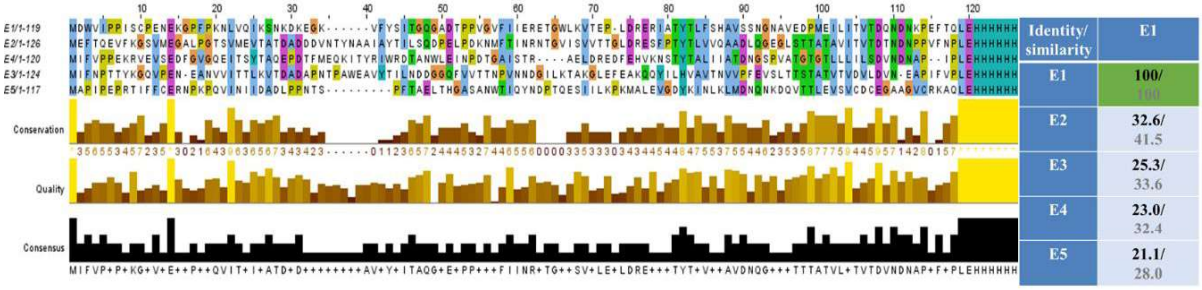


Figure 3.1. Comparison of the amino acid sequence of the five EC domains of E-cadherin with each other

3.2 How similar are the domains of N-cadherin to each other?

Like the EC domains of the E-cadherins, N-cadherins also showed conserved acidic amino acid residues, but in this case, the positions are entirely different. Glutamate residues are present at the 16th position and aspartate residues at positions 31-33 at the N-terminus, but at the C-terminus, a few aspartates are present but not consistently so in all five domains. Glycine is present at the 66th position in all the EC domains of N-cadherin, which is almost in the middle of the peptide length of each domain, potentially facilitating flexibility of the backbone's trajectory in structural terms. Hydrophobic residues are also conserved at various places and especially crowded together towards the C-terminus of the peptide chain length comprising each domain. The table showing percentages of identity demonstrates that N2 and N3 shared the highest identity and N5 showed the least identity, which once again indicates that N5 probably evolved last during evolution of the N-cadherins, as was inferred in the case of E5 for the E-cadherins.

Figure 3.2. Comparison of the amino acid sequence of the five EC domains of N-cadherin with each other

3.3 How similar are the corresponding domains of E-and N-cadherin?

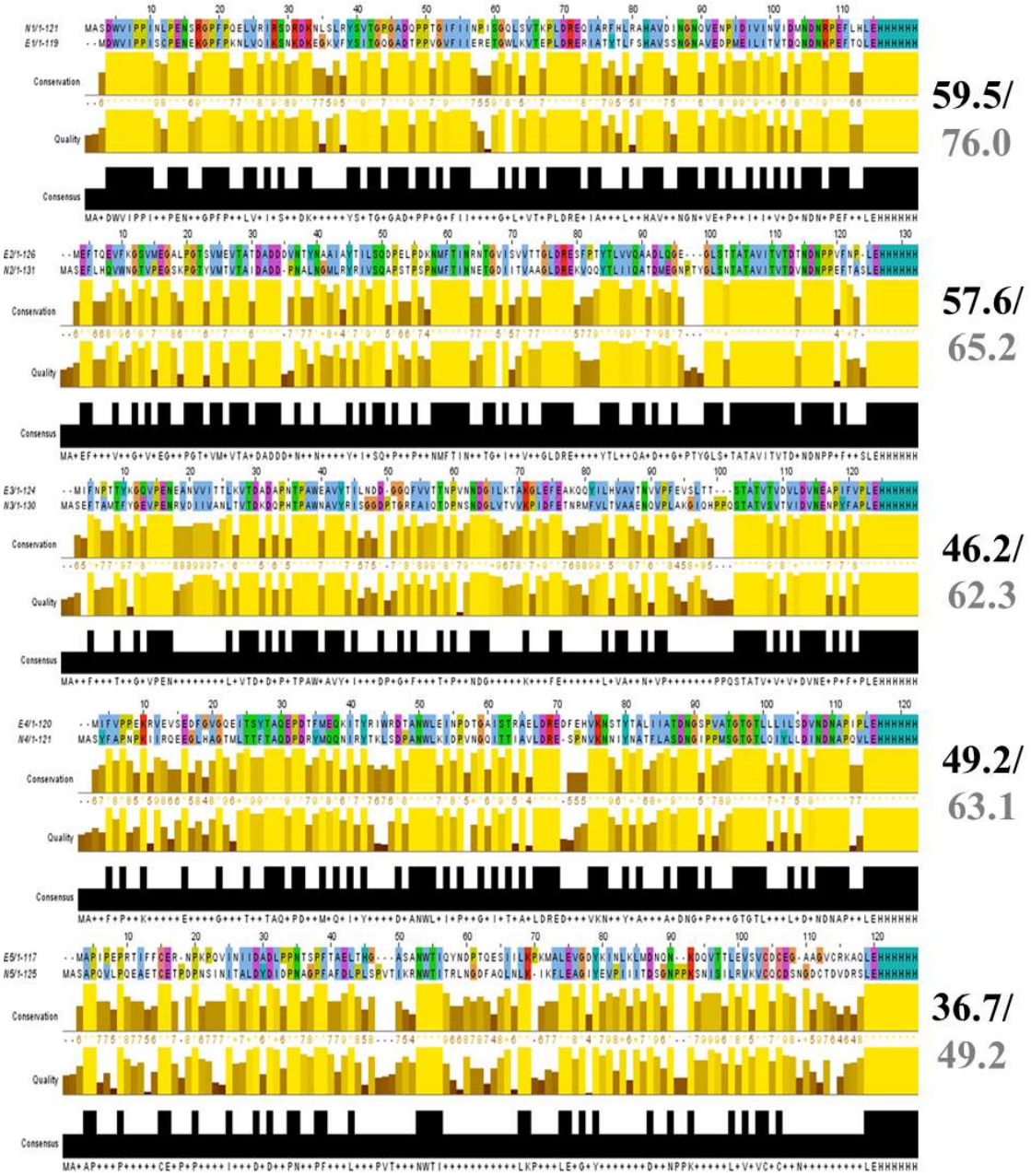


Figure. 3.4. Comparison of amino sequence alignment of E- and N- cadherins EC domains separately with the information of sequence identity and similarity (%)

The sequence alignment with the percentage identity/similarity for all the domains of E and N-cadherins in all possible combinations have been mentioned in Figure__. It is evident that the first two domains of E- and N-cadherins share a very significant identity with each other, while the other three domains show less than 50 % overall identity. It can probably be proposed that both the cadherin types are very closely related in terms of their domains despite their being known to perform very functions involving different cell types. This also suggests that one might have evolved from the other, and it appears likely from the analyses of similarities and identities of sequence that N-cadherins appeared first during evolution and E-cadherin evolved later, although the analysis provides no real clarity about whether the five domain structures of the two evolved first before the domains diverged, or whether the diversity arose in an ancestral domain with subsequent convergent evolution producing the five domain architecture. The first 30 amino acids from the N-terminus are identical and, with few gaps, residues 38 to 110 are also identical. This suggests that E1 domain could have evolved from N1 since there is a possibility that during evolution the formation and diversification of neural tissues occurred before that of epithelia.

Similarly, with the case with domain 2, it is found that E2 and N2 share 57.6 % identity and a significant number of position have highly conserved residue identity. The domains 3, 4 and 5 are far less conserved, and also have folds which are considerably different from those of the first two domains. The locations which involve calcium binding sites are also conserved in almost all the domains.

Clustal X Default Colouring		
Residue at position	**Applied Colour**	**{ Threshhold, Residue group }**
A,I,L,M,F,W,V	BLUE	{+60%, WLVIMAFCHP}
R,K	RED	{+60%,KR}, {+80%, K,R,Q}
N	GREEN	{+50%, N}, {+85%, N,Y}
C	BLUE	{+60%, WLVIMAFCHP}
C	PINK	{100%, C}
Q	GREEN	{+60%,KR}, {+50%,QE}, {+85%,Q,E,K,R}
E	MAGENTA	{+60%,KR}, {+50%,QE}, {+85%,E,Q,D}
D	MAGENTA	{+60%,KR}, {+85%, K,R,Q}, {+50%,ED}
G	ORANGE	{+0%, G}
H,Y	CYAN	{+60%, WLVIMAFCHP}, {+85%, W,Y,A,C,P,Q,F,H,I,L,M,V}
P	YELLOW	{+0%, P}
S,T	GREEN	{+60%, WLVIMAFCHP}, {+50%, TS}, {+85%,S,T}

Figure 3.5. Default colouring used in Clustul X sequence alignment

Identity/ similarity	E1	Identity/ similarity	N1	Identity/ similarity	N1	N2	N3	N4	N5
E1	100/ 100	N1	100/ 100	E1	59.5/ 76.0	34.8/ 45.2	26.2/ 39.6	25.5/ 34.0	22.1/ 32.5
E2	32.6/ 41.5	N2	36.6/ 52.7	E2	29.6/ 40.7	57.6/ 65.2	37.1/ 47.9	33.6/ 38.6	22.4/ 34.3
E3	25.3/ 33.6	N3	32.9/ 43.4	E3	23.9/ 39.4	30.9/ 44.6	46.2/ 62.3	20.0/ 28.8	16.9/ 22.0
E4	23.0/ 32.4	N4	25.4/ 39.9	E4	26.0/ 39.0	31.1/ 43.2	27.5/ 39.4	49.2/ 63.1	19.9/ 34.9
E5	21.1/ 28.0	N5	14.9/ 26.2	E5	23.4/ 37.5	24.0/ 30.5	23.5/ 35.9	15.8/ 21.6	36.7/ 49.2

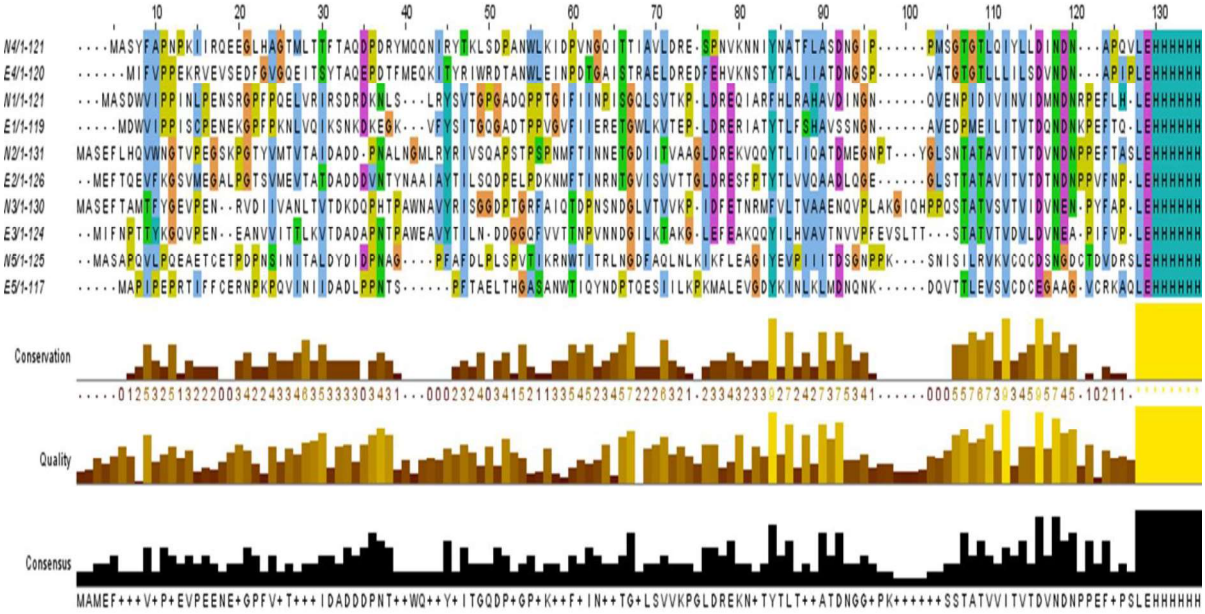


Figure 3.6. Comparison of the amino acid sequence of the all the five EC domains of E- and N-cadherins together

3.4 Comparison of root mean square deviation (RMSD) values of domain structures by truncation and superimposition of domains derived from the structures of full-length E- and N-cadherin, using pairwise combinations

To perform the structural alignments, 3q2v (mouse E-cadherin) and 3q2w (mouse N-cadherin) were downloaded. Pymol software was used to truncate and overlay the domains

from the crystal structure of the full-length EC domains, and the rmsd values are listed in the table below.

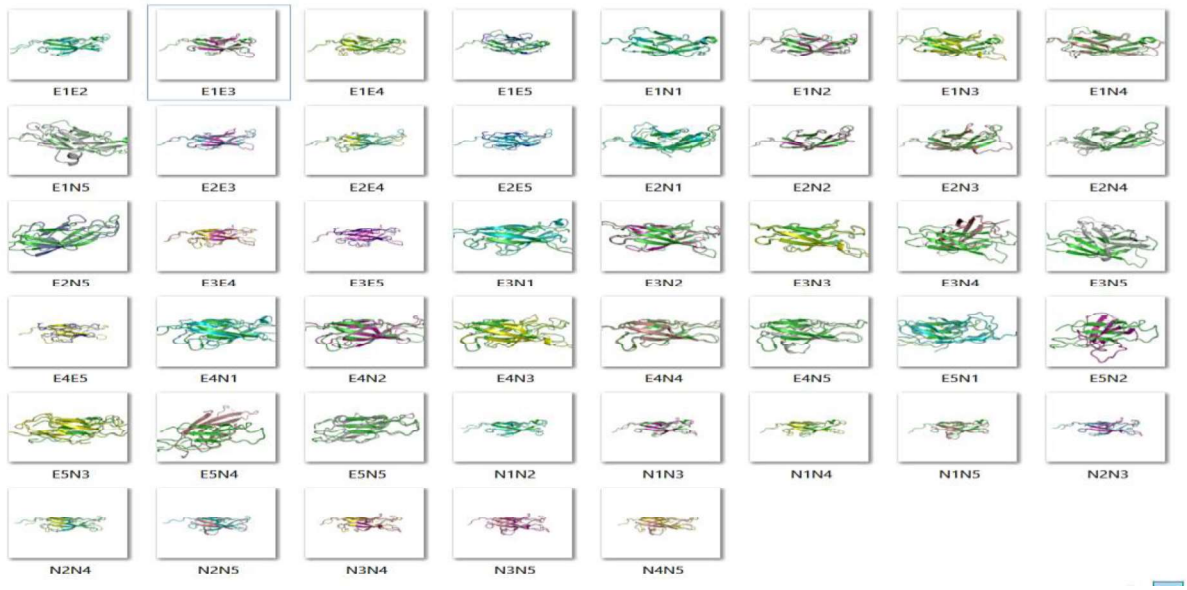


Figure 3.7. Structural alignment of EC domains of E- against EC domains of N-cadherins in all the possible combinations using Pymol.

The structural fold of E1 vs N1, E1 vs E2 and N1 vs N2 are quite similar, which is also supported by the sequence alignment data. As the figure demonstrates, the fold is very different in the case of E1 vs E5, E5 vs N4 and E5 vs N2. The amino acid sequences are also very dissimilar in case of E5 vs N4 which suggests that these two domains are distantly related.

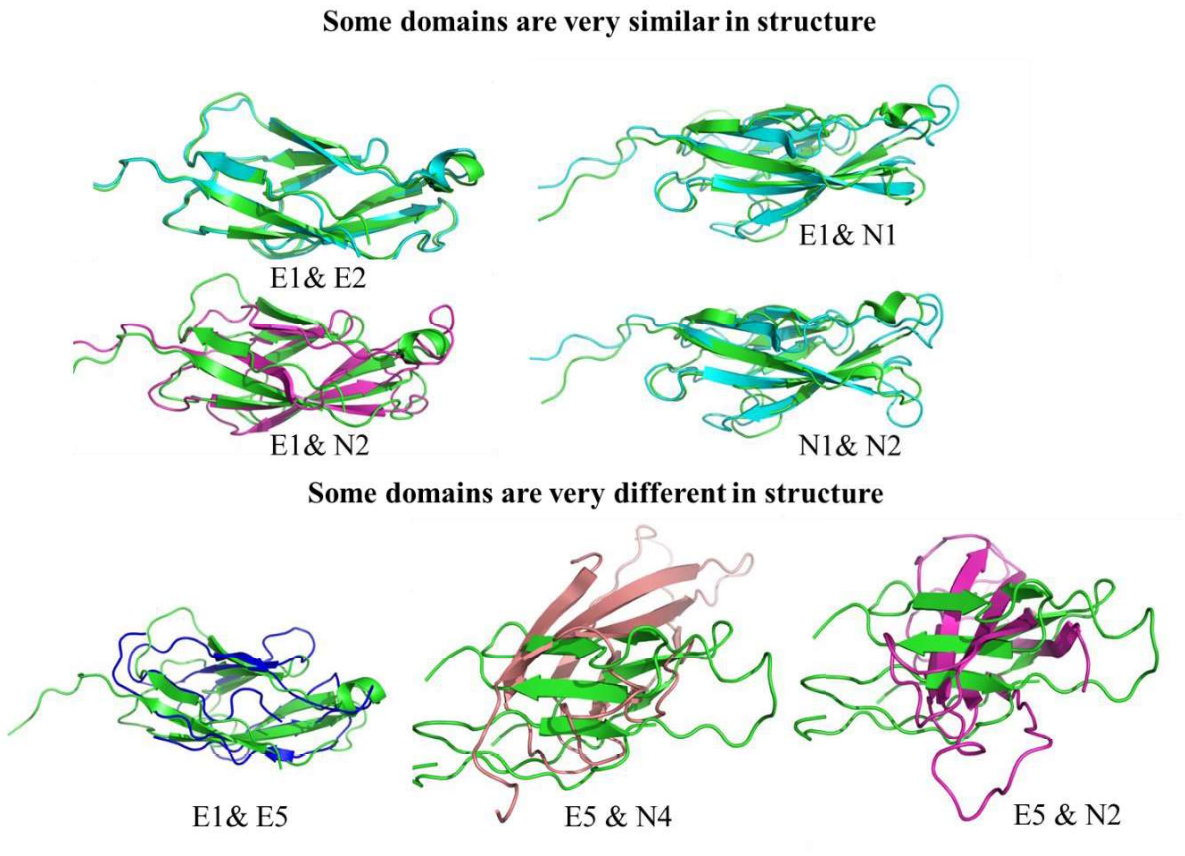


Figure 3.8. Structural alignment of some important domains showing significant resemblances and dissimilarities and of EC domains

	E1	E2	E3	E4	E5
E1	0	1.77	1.78	1.65	1.97
E2	1.77	0	1.50	1.52	1.88
E3	1.78	1.50	0	1.78	1.84
E4	1.65	1.52	1.78	0	1.90
E5	1.97	1.88	1.84	1.90	0

	N1	N2	N3	N4	N5
N1	0	1.65	1.78	1.74	2.13
N2	1.65	0	1.47	1.67	1.81
N3	1.78	1.47	0	1.83	1.75
N4	1.74	1.67	1.83	0	1.85
N5	2.13	1.81	1.75	1.85	0

	N1	N2	N3	N4	N5
E1	0.80	1.70	1.86	1.70	2.20
E2	1.70	0.83	1.44	1.74	1.79
E3	1.83	1.60	1.17	1.80	1.83
E4	1.61	1.58	1.67	1.11	1.97
E5	2.28	1.84	1.67	1.77	1.55

Figure 3.9. RMSD values of E- and N-cadherins listed in all combinations for comparison among the EC domains from the structural alignments

The results from the above sequence and structural alignments are beneficial, as important insights can be obtained. These analyses will help in understanding the functions of individual domains and whether their biochemical properties are similar or different despite their extreme sequence and structural similarities. One thing to note is that the analyses in

this entire chapter are based on the sequences and structure of mouse cadherins. The mouse cadherins are almost identical to human cadherins, with very few differences. As the structures of the human cadherins are not determined, this analysis of the mouse cadherin sequences and structures serves as a substitute for an analysis of human cadherin sequences and structures.

Chapter 4: Creation and confirmation of identity of identity of individual and fused domains of E- and N-cadherins

4.1 Cloning of genes

4.1.1 Extracellular domains of E- and N-cadherins

The recombinant gene encoding the extracellular domains of human E-cadherin consists of amino acid residues 152-697 from the 882 residues-long polypeptide product of the *cdh1* gene. The residues preceding 152 comprise the signal and pro-peptide regions, which are removed during secretion. The residues following 697 comprise the transmembrane and cytoplasmic regions of the protein. We amplified the regions comprising residues 152-697 by using a cDNA template prepared from human breast cancer tissue and the appropriate primers. The extracellular domains EC1-EC5 will hereinafter be called E1, E2, E3, E4 and E5. The primers were designed to incorporate restriction sites, NdeI and XhoI, to facilitate digestion and in-frame cloning into the pET-23a vector, for transformation into *E.coli* strain XL-1 Blue.

Similarly, the gene encoding the extracellular domains of human N-cadherin consists of residues 160-714 from the 904 residues-long polypeptide product of the *cdh2* gene. The amplification was done using a cDNA template prepared from HEK-293 cell line and incorporating restriction sites NheI and XhoI, through the use of appropriate primers. The extracellular domains of N-cadherin, EC1-EC5, will from now on be called N1, N2, N3, N4 and N5.

The PCR amplification of both E-and N-cadherins was performed at different annealing temperatures (shown in Figure 1). The amplicon size was confirmed by running agarose gel electrophoresis along with a 1kb DNA ladder, and the correct DNA bands were purified using a gel extraction kit.

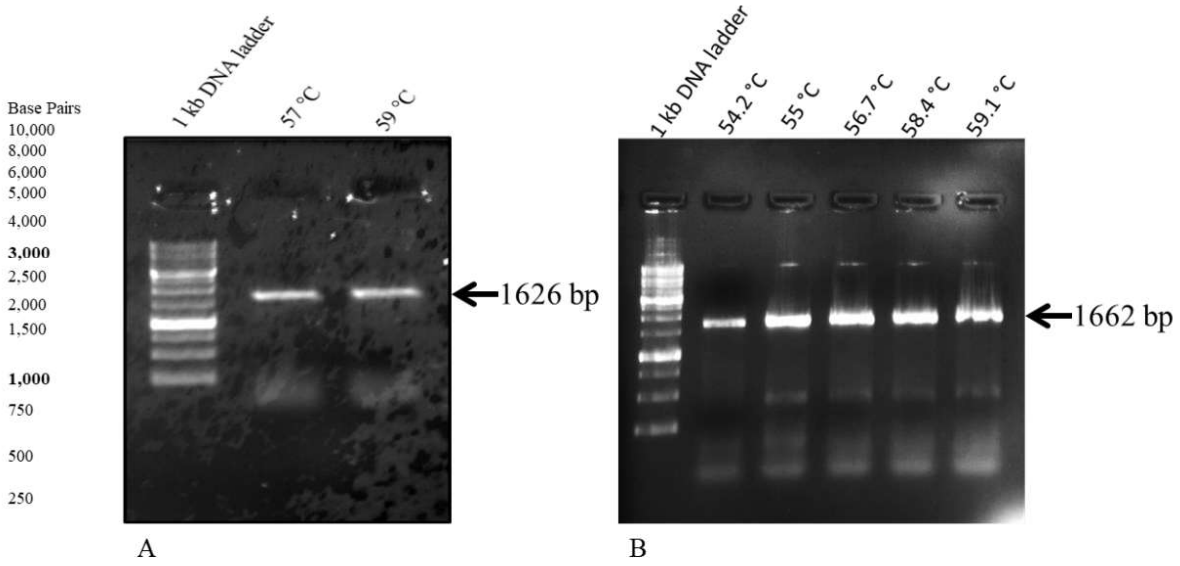


Figure 4.1. *Panel A*. PCR amplified product of E-cadherin extracellular domain, *Panel B*. PCR amplified product of N-cadherin extracellular domain

The purified PCR products and the vector pET-23a were subjected to digestion to create sticky ends using a set of enzymes (NdeI & XhoI for the E-cadherin encoding gene, and NheI & XhoI for the N-cadherin encoding gene). After the digestion, ligation reactions were performed for both the genes, and the ligation products were transformed into XL-1 Blue competent cells. The colonies were screened by colony PCR, using T7 vector-specific primers and the positive clones were inoculated into LB media with appropriate antibiotics. Plasmid purification was done from the inoculated cultures, and these were digested with the previously used set of enzymes, to confirm the sizes of the inserts. In Figure 2, panel A represents an insert size for E-cadherin which is 1.6 kb while panel B shows the same for N-cadherin. The gene sequences of both the clones were confirmed through DNA sequencing (1st BASE DNA sequencing service) with vector-specific primers (T7 forward and T7 reverse universal primers).

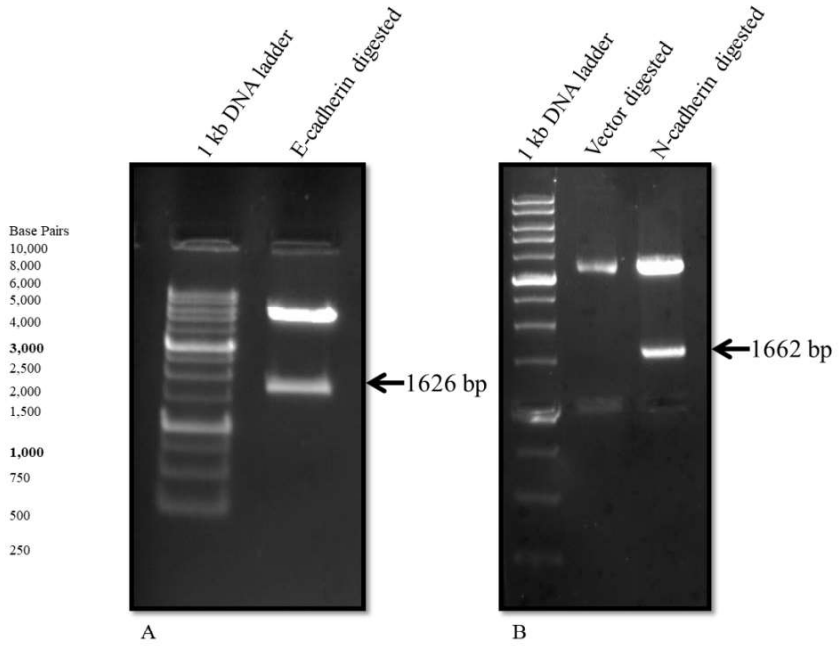


Figure 4.2. Confirmation of insert by restriction digestion, Panel A. Restriction digestion of E-cadherin clone by NdeI and XhoI, Panel B. Restriction digestion of N-cadherin clone by NheI and XhoI

The DNA constructs encoding the entire length of five extracellular domains of E- and N-cadherins were used as templates to sub-clone all the individual domains and various fusions of adjacent domains, as well as groups of domains. The information present determined the boundaries of the five individual extracellular domains of E- and N-cadherins in UniProt, and the domains consisted of these sequences along with those of the linkers located on the C-terminal side as well as two preceding amino acids (where available) and two succeeding amino acids. To illustrate this, we can take an example of the E1 domain, where the sequence encoding E1 was followed by the sequence of the linker present between E1 and E2, with two additional residues after the linker's boundary and so on. The complete scheme of designing individual domains and various domain fusions is shown in Figure 3.

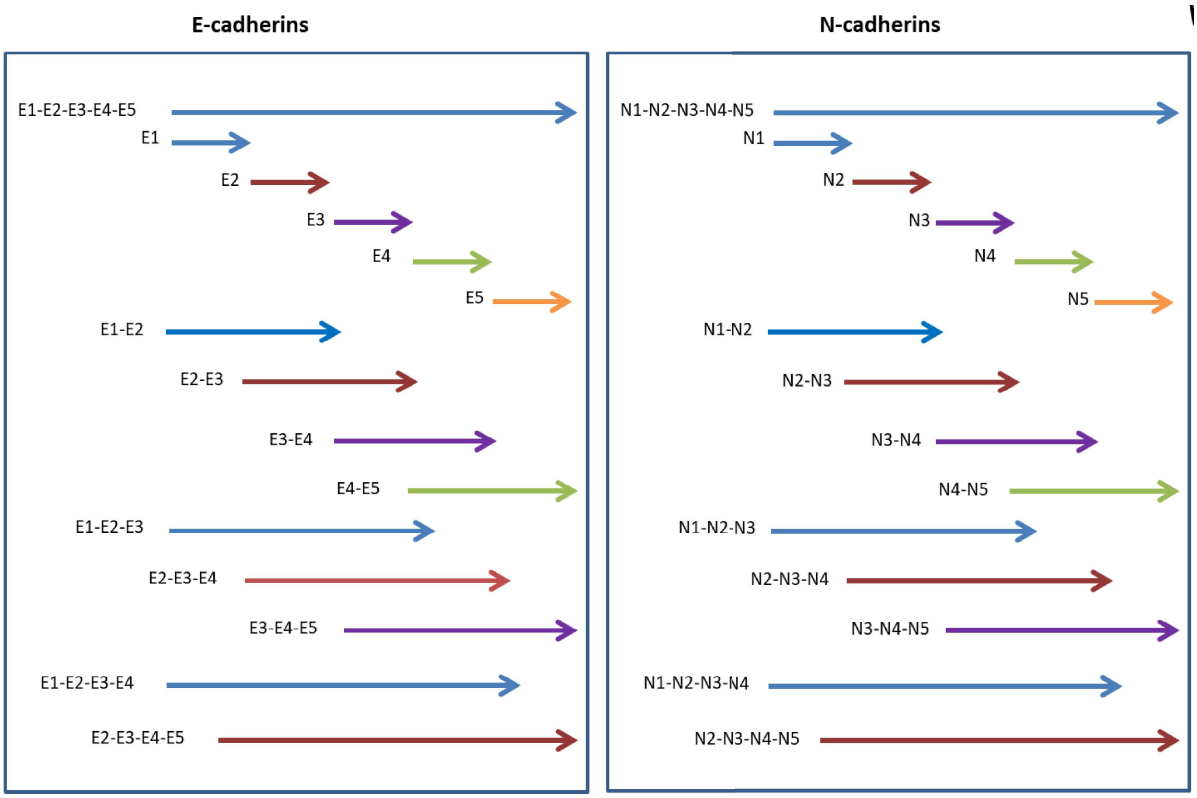


Figure 4.3. Schematic diagram representing truncation of all the isolated and fused domains

4.1.2 Cloning of individual (single) domains

Since we wanted to study the characteristics of the all the domains; we initially started cloning all the individual domains of E-and N-cadherins simultaneously. In this section, we have described the PCR amplification of representative domains containing affinity tags (6xHis tags), along with the untagged (without His-tag) forms of the same domains.

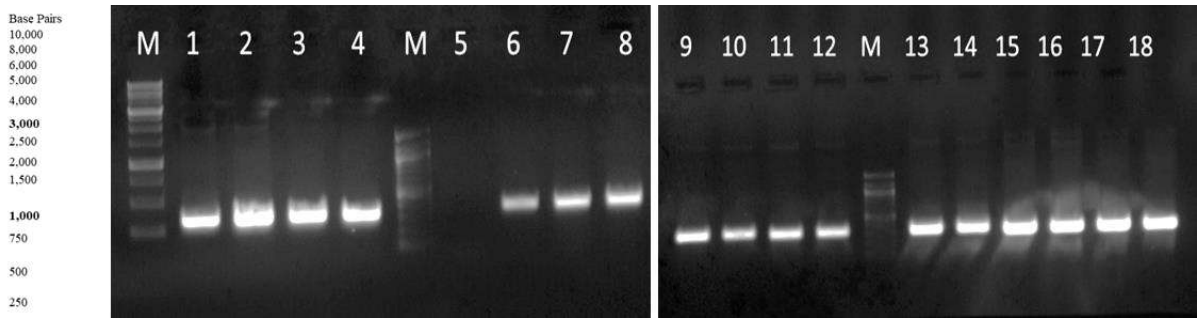


Figure 4.4. PCR amplification of single domain constructs (1-2 E1, 3-4 E1UT, 5-6 E2UT, 7-8 E3, 9-10 E3UT, 11-12 E4, 13-14 E4UT, 15-16 E5, and 17-18 E5UT)

The results of colony PCR performed to confirm cloning have been shown in Figure 5. All the PCRs were done using T7 universal primers, with the template consisting of the colonies to be confirmed, i.e., its entire content of DNA.

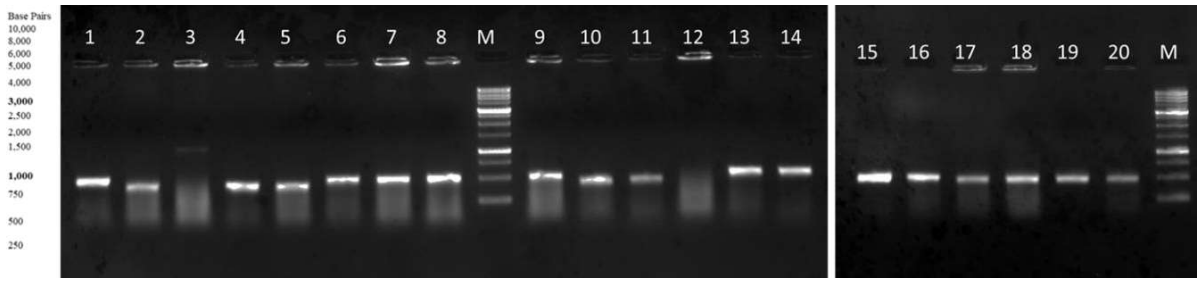


Figure 4.5. Colony PCR for screening the transformants (1-2 E1, 3-5 E1 UT, 6-7 E2, 8-9 E2 UT, 10-11 E3, 12-14 E4, 15-16 E4 UT, 17-18 E5, and 19-20 E5 UT)

One or two positive colonies for each construct were inoculated, and their insert sizes were confirmed by restriction digestion, which is presented in Figure 6. The lanes 1-9 show the vector and the insert which was released after restriction digestion.

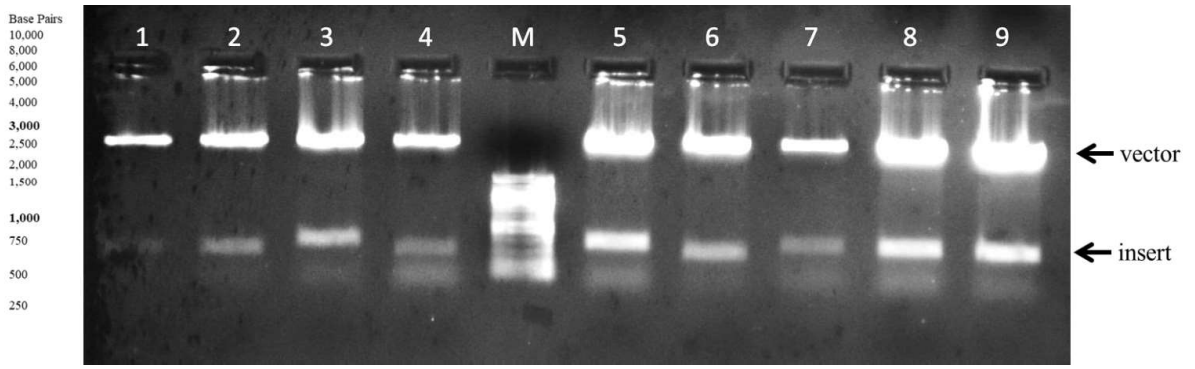


Figure 4.6. Confirmation of clones by restriction digestion (1 E1, 2 E1 UT, 3 E2, 4 E2 UT, 5 E3, 6 E3 UT, 7 E4, 8 E4 UT, and 9 E5)

Finally, the sequences of the clones were confirmed by DNA sequencing (1st Base DNA sequencing). The chromatograms obtained were visualized and analysed using the FinchTV software and the DNA sequences inferred were translated into amino acid sequences using the GeneRunner software. The sequence alignment was done for both the DNA as well as protein sequences using the Multialign web server (Multiple sequence alignment by Florence Corpet).

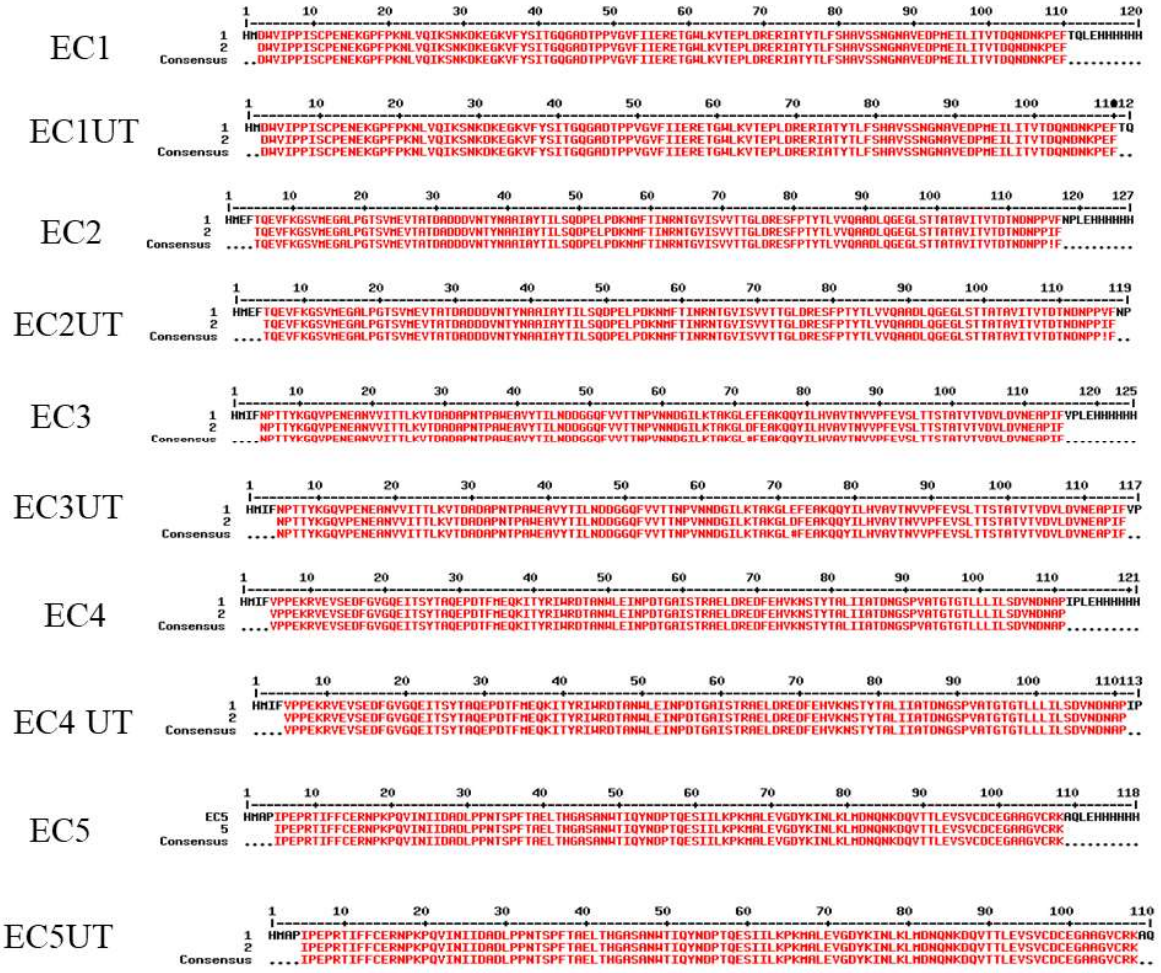


Figure 4.7. Sequencing results of single domain constructs (His-tagged and untagged)

Since the cloning of all the constructs was done simultaneously, we also prepared various fusions of neighbouring domains as well as groups of domains. Figure 3, which shows the schematic diagrams of individual domains also shows schematic diagrams of all the fusion domains which were constructed alongside, with the individual domains. The genes encoding the entire length of extracellular domains of E-and N-cadherins were used as templates for the PCR amplification to amplify fusion constructs using ap9propriate primers. Figure 9 represents PCR amplification of various fusion constructs.

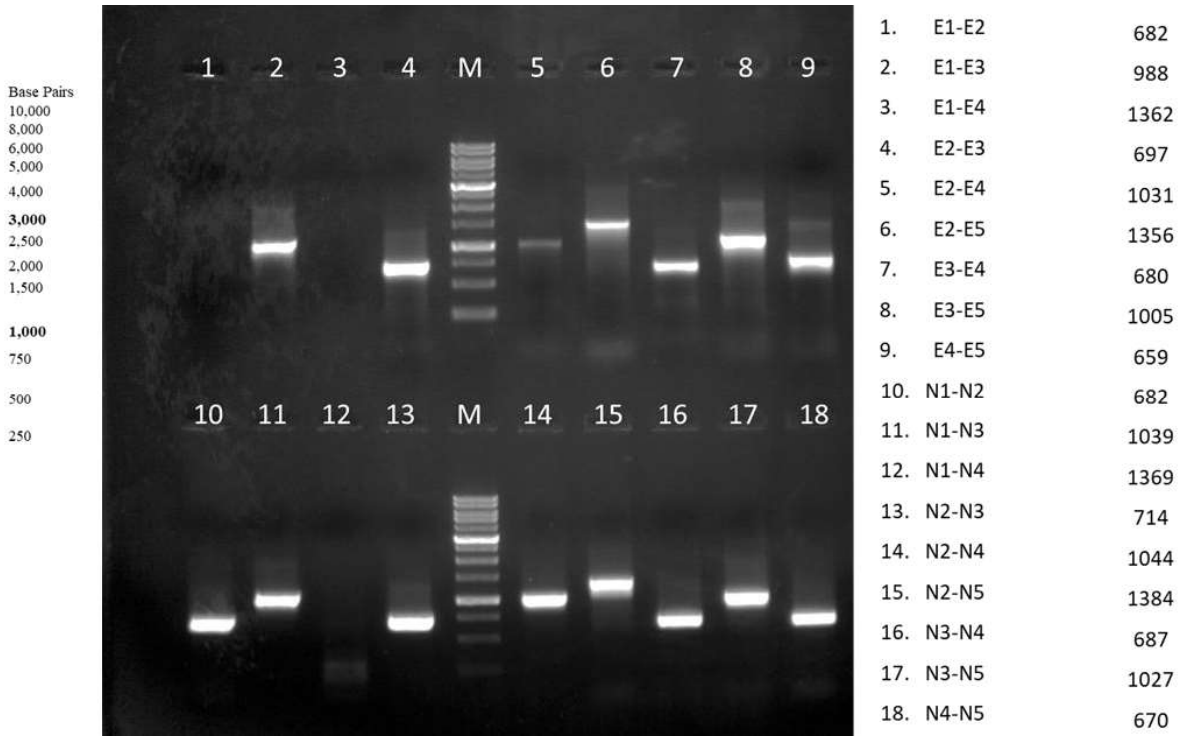


Figure 4.8. PCR amplification of fusion domains of E- and N -cadherins

Similarly, as with the individual domains, the fusion domain PCR products were purified from agarose gels and subjected to restriction digestion with restriction enzymes. In all the E-cadherin fusion constructs, NdeI and XhoI enzymes were used, while NheI and XhoI enzymes were used for all the N-cadherin fusion constructs. Ligation reactions followed digestion and all the ligation products were transformed into XL-1 Blue competent cells and plated on LB agar plate with appropriate antibiotics. The colonies which appeared were screened through colony PCR using T7 universal primers, and the positive colonies for each clone were inoculated into cultures, for plasmid isolation**.** Figure 10 represents colony PCR results showing positive colonies for various fusion domain constructs.

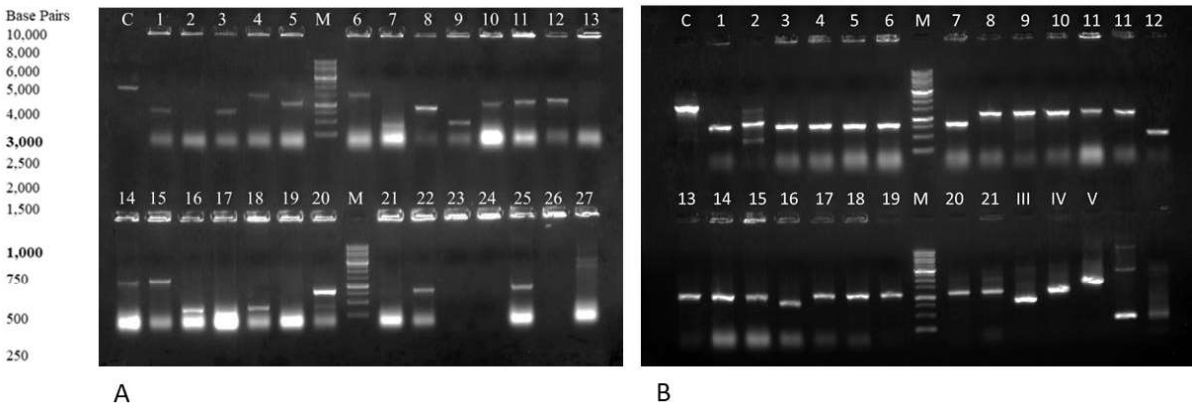


Figure 4.9. Colony PCR results of fusion domain constructs. *Panel A.* C +ve control, 1-3 E1-E2, 4-6 E1-E2-E3, 7-9 E2-E3, 10-12 E2-E3-E4 13-15 E2-E3-E4-E5, 16-17 E3-E4-E5, 18 E4-E5, 19-21 N1-N2, 22 N2-N3, 23-24 N2-N3-N4-N5, 25-27 N3-N4, *Panel B.* 1-7 E3-E4, 8-12 N1-N2-N3, 13-18 N2-N3-N4, 19-23 N3-N4-N5, 24-27 N4-N5, III E1-E2-E3-E4, IV E2-E4, V E2-E5

The plasmids isolated from the positive colonies were further confirmed by restriction digestion and sequences were confirmed by DNA sequencing.

4.1.3 Protein expression and purification

In all cases, plasmids containing the designed gene constructs were transformed into *E.coli* strain Rosetta DE-3 for protein expression. The production of proteins through expression of the corresponding encoding genes was induced in secondary cultures of the relevant bacterial clones, at an OD_{600} value of 0.6, by the addition of the inducer, IPTG (0.5 mM), with simultaneous reduction of the temperature from 37 °C down to 25 °C, for a period of 7-8 hours, after the induction of expression through addition of IPTG. Non-denaturing purification in buffers lacking denaturants, as well as denaturing purification (in denaturing buffers, in the presence of 8 M urea) was performed according to per standard protocols (Qiagen). In all cases, for non-denaturing (native) purification, the lysate was passed through a Ni-NTA column, and this was followed by washing the column using 20-50 mM imidazole to remove non-specifically adsorbed protein, and the specifically-bound protein was finally eluted at the end through use of buffer containing 250 mM imidazole, offering competition to the Ni-NTA resin, for binding to the 6xHis affinity tag. For purification under denaturing

conditions, bound protein was eluted using a buffer of pH 4.5. In every case, the protein obtained was subjected to dialysis against 50 mM tris of pH 7.4, containing 3mM $CaCl_2$.

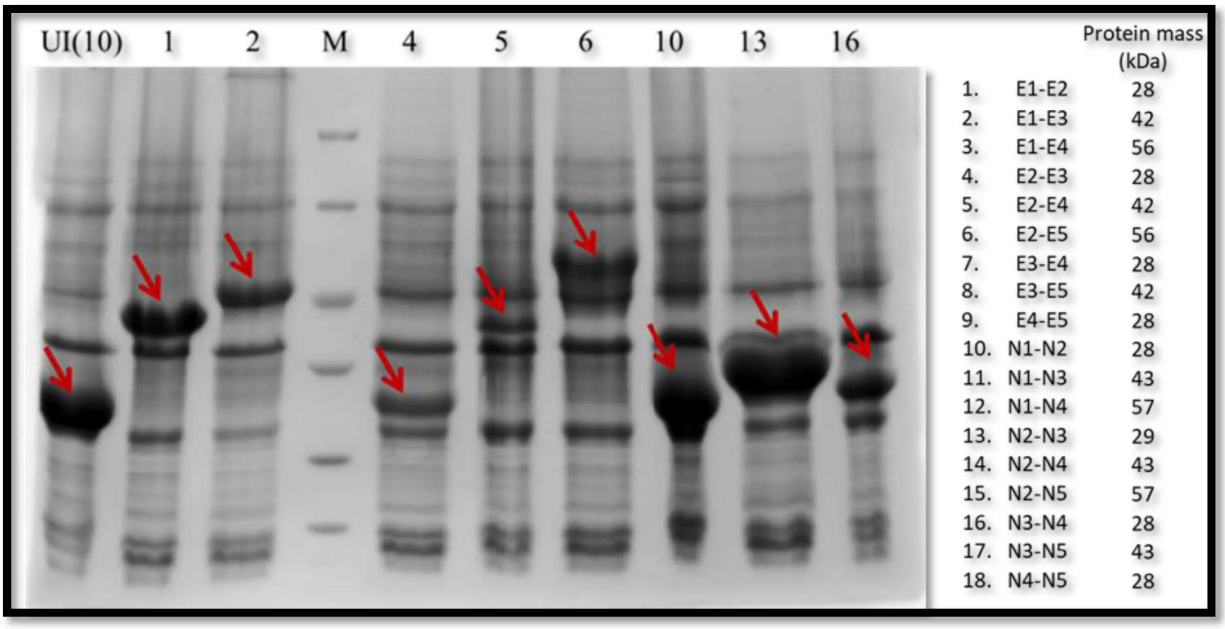


Figure 4.10. Protein **e**xpression check-in Rosetta DE3 expression host after induction

The protein expression profiles of some representative domain constructs are shown in Figure 11 where the clones were inoculated and grown in 10 ml LB culture containing appropriate antibiotics. 1ml cultures of each grown bacterial were centrifuged, and the bacterial pellets were boiled at 99 °C in the presence of SDS sample loading buffer, for electrophoresis using SDS-PAGE. In the Figure below the first lane shows the un-induced lane, i.e., the lane corresponding to cultures grown without any addition of the inducer (IPTG). The other lanes (i.e., lanes 1, 2, 4, 5, 6, 10, 13 and 16) show the induced protein, where arrows represent the overexpressed protein.

4.1.4 SDS-PAGE analysis of single domain constructs

Proteins produced from all 10 single domain constructs were purified by Ni-NTA affinity chromatography and electrophoretically analysed on SDS-PAGE, to confirm their masses and purity. Figure 12 depicts the electrophoretic patterns of all 10 single domain constructs which were purified by denaturing methods. All the domains are approximately 110 amino acids long and produce similarly-sized protein products, but on SDS-PAGE their electrophoretic mobilities were not the same. This mobility has been discussed in a separate section.

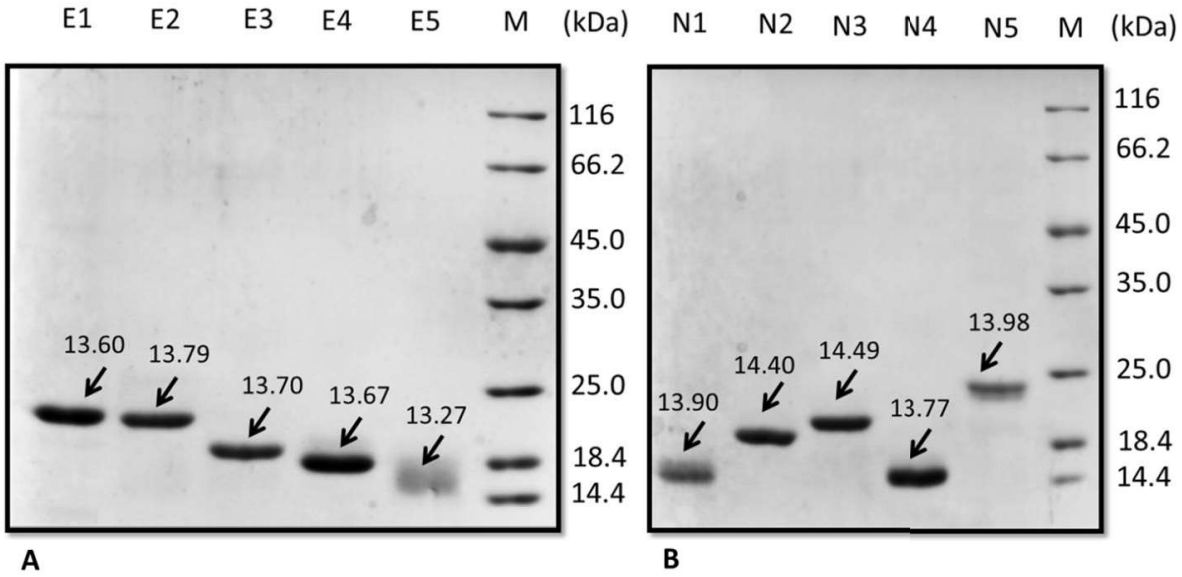


Figure 4.11. SDS-PAGE showing Ni-NTA affinity purification of *Panel A.* five extracellular domains of E-cadherin. *Panel B.* five extracellular domains of E-cadherin. The values marked next to the arrows point out the actual molecular weight(s) of each domain.

4.1.5 SDS-PAGE analysis of fusion domain constructs

In addition to the 10 single domain constructs, 20 further fusion domain constructs were also cloned and made. In this thesis, only a few of them have been studied, and these will be described in detail is a separate section.

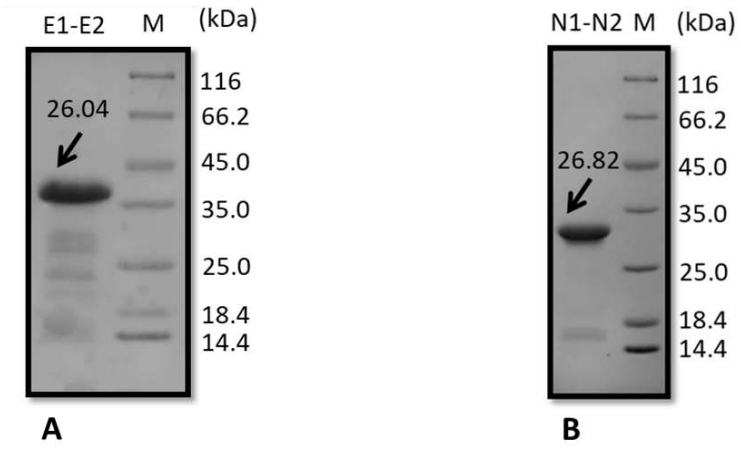


Figure 4.12. SDS-PAGE showing Ni-NTA affinity purification of *Panel A.* E1-E2 fusion domain. *Panel B.* N1-N2 fusion domain

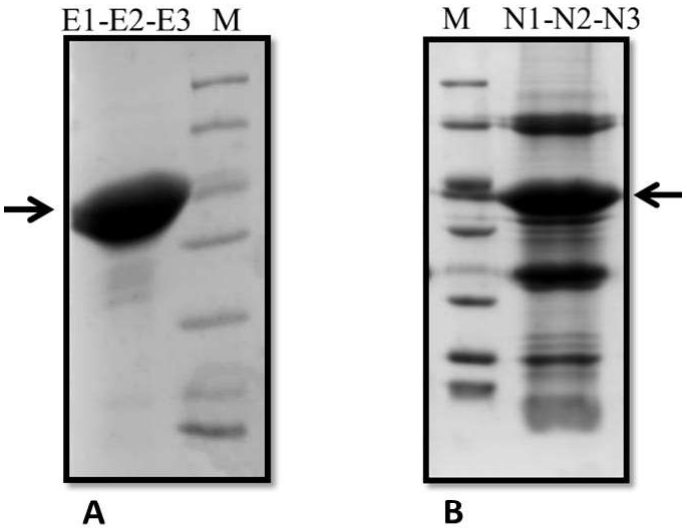


Figure 4.13. SDS-PAGE showing Ni-NTA affinity purification of *Panel A.* E1-E2-E3 fusion domain. *Panel B.* N1-N2-N3 fusion domain

4.2 Identity confirmation of single domains from gel band through mass spectrometry

The polypeptide chain integrity and sequence identity of all single and fusion constructs were confirmed by mass spectrometry. Standard methods of MALDI ionization and analyses were used for peptide mass fingerprinting analysis of purified protein constructs Intact mass analysis was performed by electrospray ionization (ESI) mode while peptide mass fingerprinting, using MALDI mode as the ionization source.

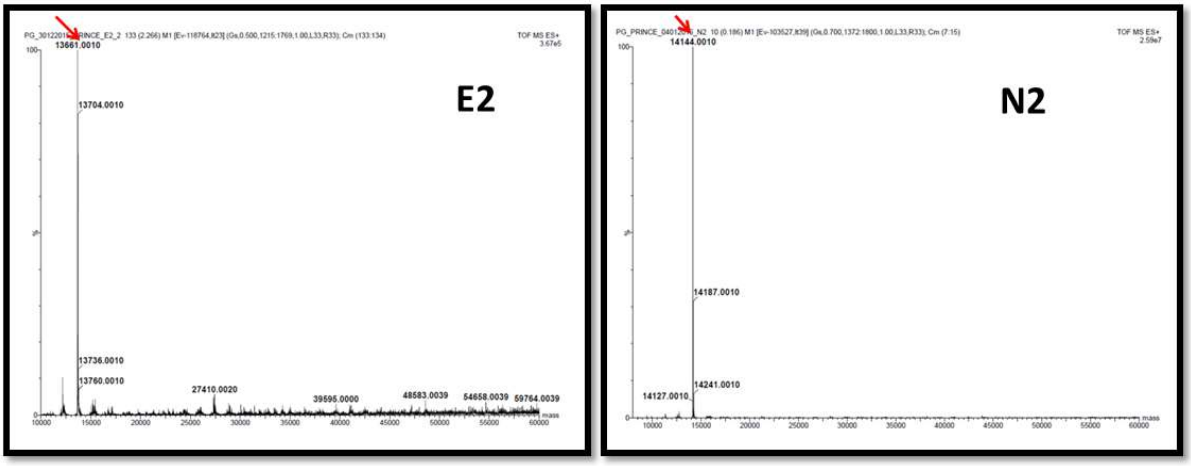


Figure 4.14. Intact mass analysis of representative cadherin single domains (E2 and N2)

The intact mass analysis confirmed that the domains did not undergo any post-translational modifications, cleavage or any other modifications. Further to verify the identities of the constructs, peptide mass fingerprinting was performed in which peptides released from the protein constructs through digestion by trypsin were matched with the MS profiles of peptides expected to be released through tryptic digestion *in silico*. A gradient of collision energy was also applied to break some of the tryptic peptides further (through MS-MS), and the masses of the resulting fragments were analysed and re-stitched into sequences by using MassLynx software, yielding the sequences of some of these tryptic peptides.

The matched peptides are highlighted with red arrows. The MS-MS approach has then used the peaks with the highest intensity, to do peptide sequencing for further confirmation of identity.

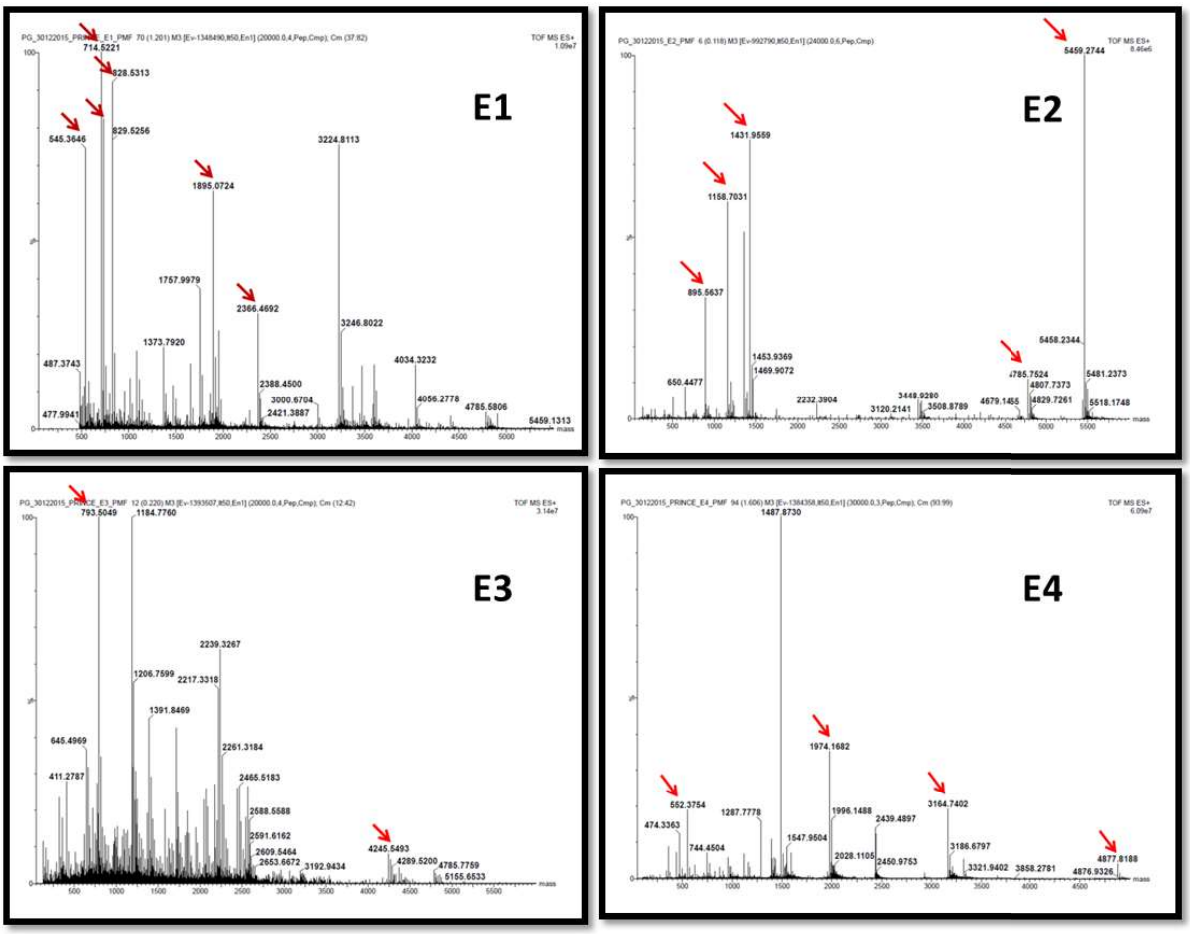


Figure 4.15. Diagnostic tryptic peptide masses (arrows) and sequence determination of peptides of E1, E2, E3 and E4 domains (The arrows in panels denote the tryptic fragment mass peaks which were used to identify the cadherin domains).

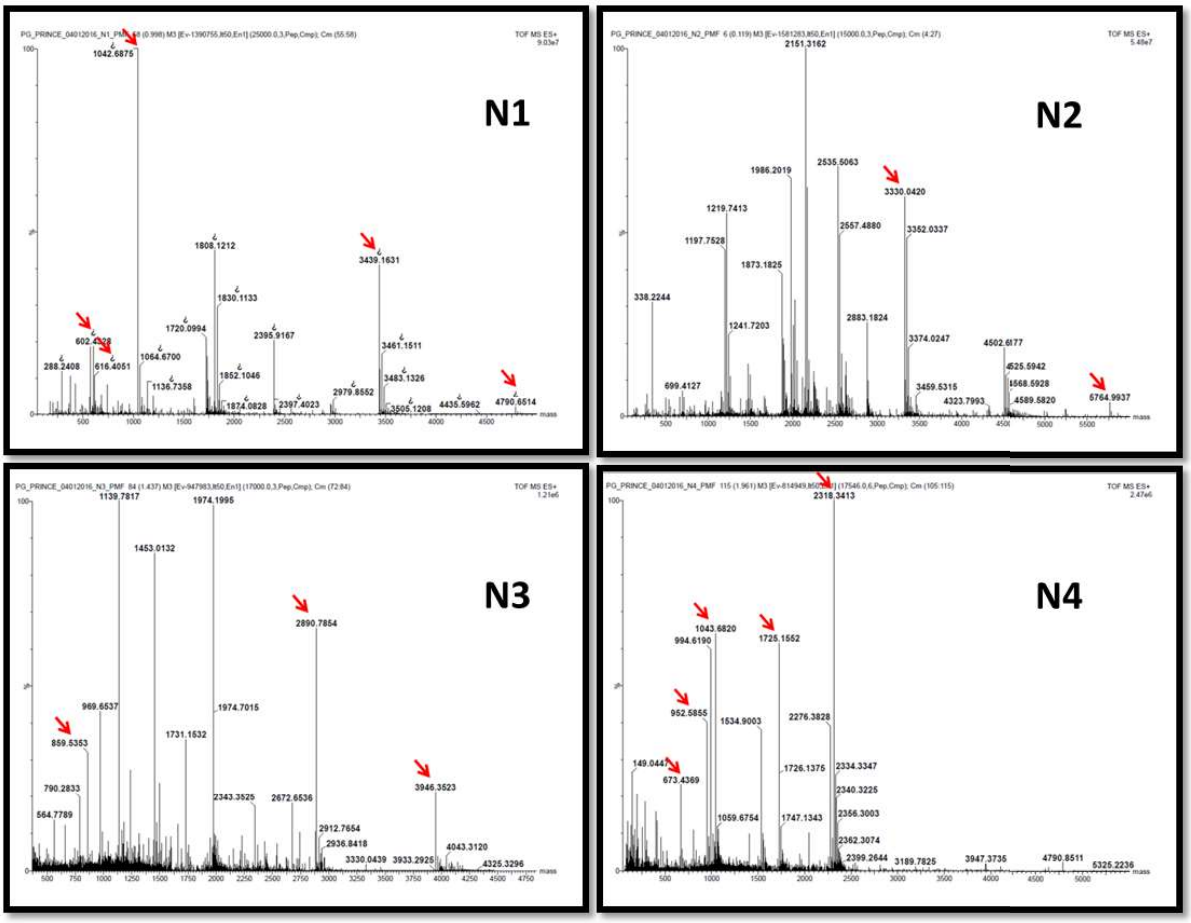


Figure 4.16. Diagnostic tryptic peptide masses (arrows) and sequence determination of peptides of N1, N2, N3 and N4 domains (The arrows in panels denote the tryptic fragment mass peaks which were used to identify the cadherin domains).

The figure shown below is a representative snapshot of the screen output from the analysis software, Masslynx, in which the sub-routine, Biolynx, was used to determine the sequences of the selected peptides.

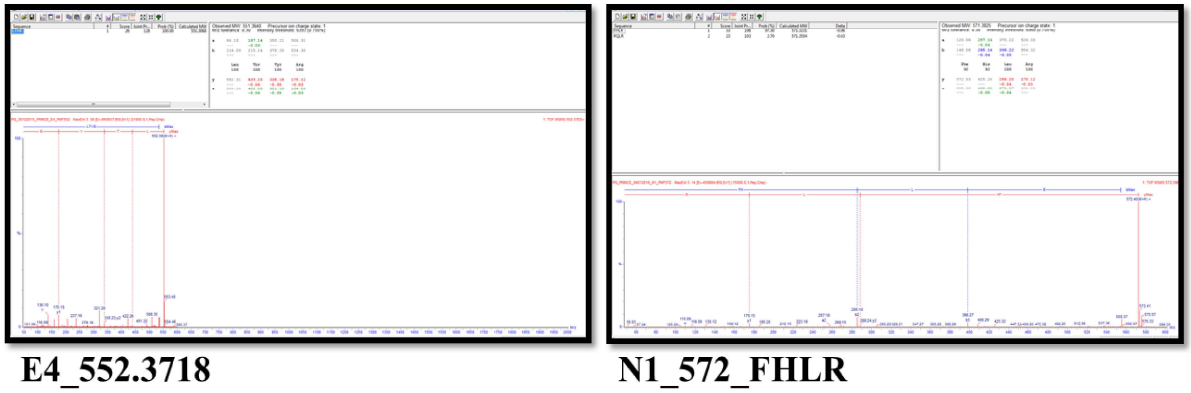


Figure 4.17. Sequence determination of representative peptides

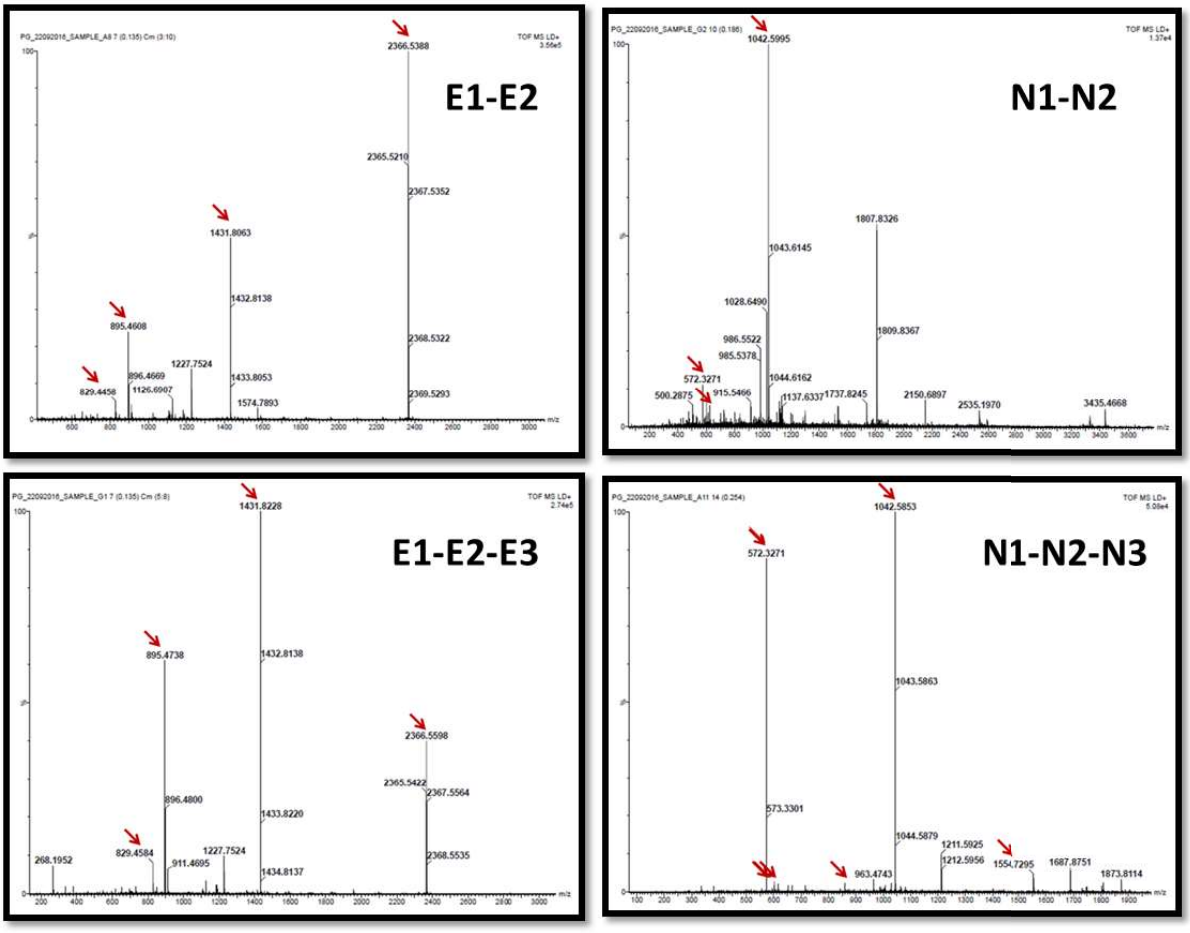


Figure 4.18. Diagnostic tryptic peptide masses (arrows) and sequence determination of peptides of fusion domains (The arrows in panels denote the tryptic fragment mass peaks which were used to identify the cadherin domains).

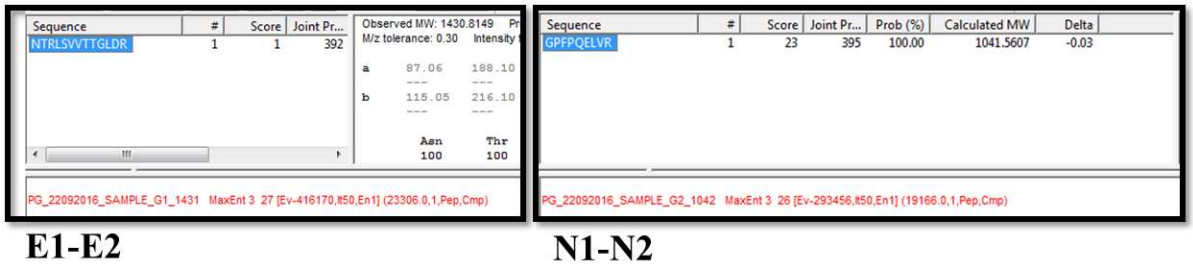


Figure 4.19. Sequence determination of representative peptides

The mass spectrometry data confirm that the sequences, and chain integrity, of all the domains used for the studies presented in this thesis, were found to be correct, and intact, respectively, in terms of both their calculated masses and verified amino acid sequences of some tryptic peptides. Further, intact mass analysis also confirmed that the domain constructs had not undergone any post-translational modifications (PTMs). It may be noted that all the cadherin domains used here were made in a prokaryotic system, i.e., in *E. coli* in which there is no significant PTM other than phosphorylation, in any case.

4.3 Comments on the yields and solubilities of the various protein domain constructs

Recombinant proteins are best expressed and purified in their natively folded forms, in order to perform their functions, although they can also be purified under denaturing conditions and then refolded, where this is feasible. It was found that different cadherin domain constructs show different levels of expression, abundance, and also solubility in the *E.coli* cytoplasm. Initially, we used BL21 DE3 as the expression host but found that the levels of expression were low. We then checked the codons used in the encoding genes with what is compatible for high-level expression in *E.coli* and found several codons whose corresponding charged tRNAs are not abundantly present in *E.coli*. Therefore, we shifted to using Rosetta DE3 as the expression host and found that the expression levels were noticeably higher. Thus, we used the Rosetta DE3 expression host for all the constructs thereafter. The solubilities and levels of expression depend on various factors like the origin of the gene, the amino acid sequences, the state of folding in terms of burial of hydrophobic regions, the PTMs (post-translational modifications) which may be required, and various

other factors. From the literature, we knew that there are four glycosylation sites present in the 4th and 5th extracellular domains of the five-domain extracellular regions of the cadherins, but also that these do not have any reported effect upon the folding of domains. We thus did a comparative analysis of the solubility of domains of E-cadherin with their counterpart domains of N-cadherin.

E1, E2, N1 and N2 domains.

Domain 1 and 2 of both the cadherin types were soluble and purified in good concentration by non-denaturing purification. The total protein yields for domains E1 and E2 were lesser than those for N1 and N2, but N2 and N1-N2 were obtained in the highest concentration among all domains.

E3,E4,E5,N3, N4 and N5.

Upon expression in *E.coli,* most of the fraction of these protein domains was seen as being present in the bacterial cytoplasm in insoluble form. However, a small fraction was managed to be obtained in the soluble fraction, by attempting various treatments such as lowering of the temperature after induction of expression, and the addition of glucose after initiation of induction, during bacterial expression. With the domains, in particular, which were less soluble than the other domains, and also purified only in very low amounts, two common contaminants happened to be co-purified. These two contaminants were later confirmed to be the proteins, SlyD, and Arna, by mass spectrometry (note: these proteins possess a natural affinity for Ni-NTA on account of exposed clusters of histidine residues). In addition to these two proteins, we also found that Glutamine--fructose-6-phosphate aminotransferase was also co-purified, although in much smaller amounts, whenever the overexpression was less, or the solubility was less.

The possible explanation for the presence of the contaminants in the purified protein samples could be the following: In proportion with the lowering of the amount of soluble domain constructs, higher numbers of the binding sites of the Ni-NTA column could be expected to remain unoccupied. These would then be occupied by these *E.coli* contaminants, since they also possess stretches of multiple histidine residues in their amino acid sequences, because

of which they can bind, and co-elute during the elution step. None of these contaminants were observed during denaturing purification, presumably because all the binding sites on the resin were occupied by the desired proteins, either because the levels of expression were high or because most of the expressed protein was recovered into solution from inclusion bodies through the use of the denaturant and, therefore, available for binding and offering competition.

In the case of E5, the solubility was better than that of its counterpart N5. The construct N5 was almost wholly insoluble.

Biophysical characterization of any protein requires reasonable amounts and concentrations in solution, and this depends on the expression and purification yields. The domain constructs and fusion domain constructs which were soluble (E1, E2, N1, N2, E5, E1-E2 and N1-N2) have been studied in great details.

4.4 Comments on non-denaturing vs denaturing purification and refolding

In most cases, the expressed proteins became insoluble and tend to form inclusion bodies (IBs). To recover proteins from inclusion bodies, denaturing purification is done followed by refolding by various methods. Proteins purified in their natively folded form are usually the best one study, especially if proteins do not refold back to their native structure from denaturants, after purification in an unfolded condition. Heterologous expression of proteins in bacteria may or may not produce proteins in their native forms, depending upon the nature of their amino acid sequences, folding tendencies and much else. If a protein of interest is an enzyme, it becomes possible to confirm the native-like nature of its fold by performing enzyme reactions and monitoring kinetics, because as long as the recombinant enzyme can convert its substrate into a product, it can be considered to be in native conformation. Even if the stability and efficiency of the enzyme were to be compromised through heterologous overexpression, the native fold would still be present and detectable. In contrast, as is the case with these cadherin domains, since the proteins are not enzymes and their structures cannot be solved by protein crystallization studies or NMR for 30 different constructs, it is not an easy task to ensure the presence of the native fold. However, it is possible to ensure

whether the constructs are folded, through different kinds of spectroscopic, calorimetric and other experiments.

Therefore, we relied on spectroscopic experiments for structural-biochemical information as well as on their tendency to function (through dimer formation) to determine whether our cadherin domain constructs were folded and in native-like conformation. The dimerization of cadherin domains does not involve all the domains. Therefore, we made calcium binding ability as an initial method to confirm whether the calcium binding causes any structural and stability change.

Chapter 5: Results and discussions

5.1 Structural – biochemical characterization of E-cadherin and its truncated (single and fused domain) forms: structure, stability, refolding ability and calcium binding behaviour.

5.1.1 Analysis of gel electrophoretic behaviour

The SDS-PAGE profiles of single domain cadherin constructs showed an unusual migration pattern where the apparent molecular weights were higher (up to ~24 kDa) than their calculated molecular weights despite their being of comparable molecular weights (i.e. ~ 14 kDa). The similar kind of migration pattern was also observed in the fusion constructs derived from these single domain constructs. Initially, we started questioning whether the identities of our constructs could be wrong, but the intact mass, as well as the protein sequencing results from mass spectrometry, confirmed that the clones and proteins were correctly made. Further, the intact mass results also excluded the possibility of there being any post-translational modifications which could have occurred during protein synthesis. We then thought of the alternative possibility that the domains were forming higher order oligomeric species which are resistant to SDS during SDS-PAGE analysis since some of these domains are reported to form dimers. To rule out this possibility, the proteins were purified under denaturing conditions by overnight incubation of cell lysates in 8 M urea to destroy the structure and tendencies of domains to associate. Proteins were then boiled at 99 °C for 15-20 minutes in sample loading buffer for SDS-PAGE. This did not change the observed migration patterns on SDS-PAGE. The proteins purified under denaturing conditions also showed the same unusual (anomalous) migration behaviour (Figure 1).

When other reasons were all ruled out, we analyzed the sequences and composition of amino acid residues, and certain other parameters (refer to Table 1). We found that the domains showing an anomaly in migration, manifesting as molecules that migrate slower than anticipated are overloaded with high negative charge content.

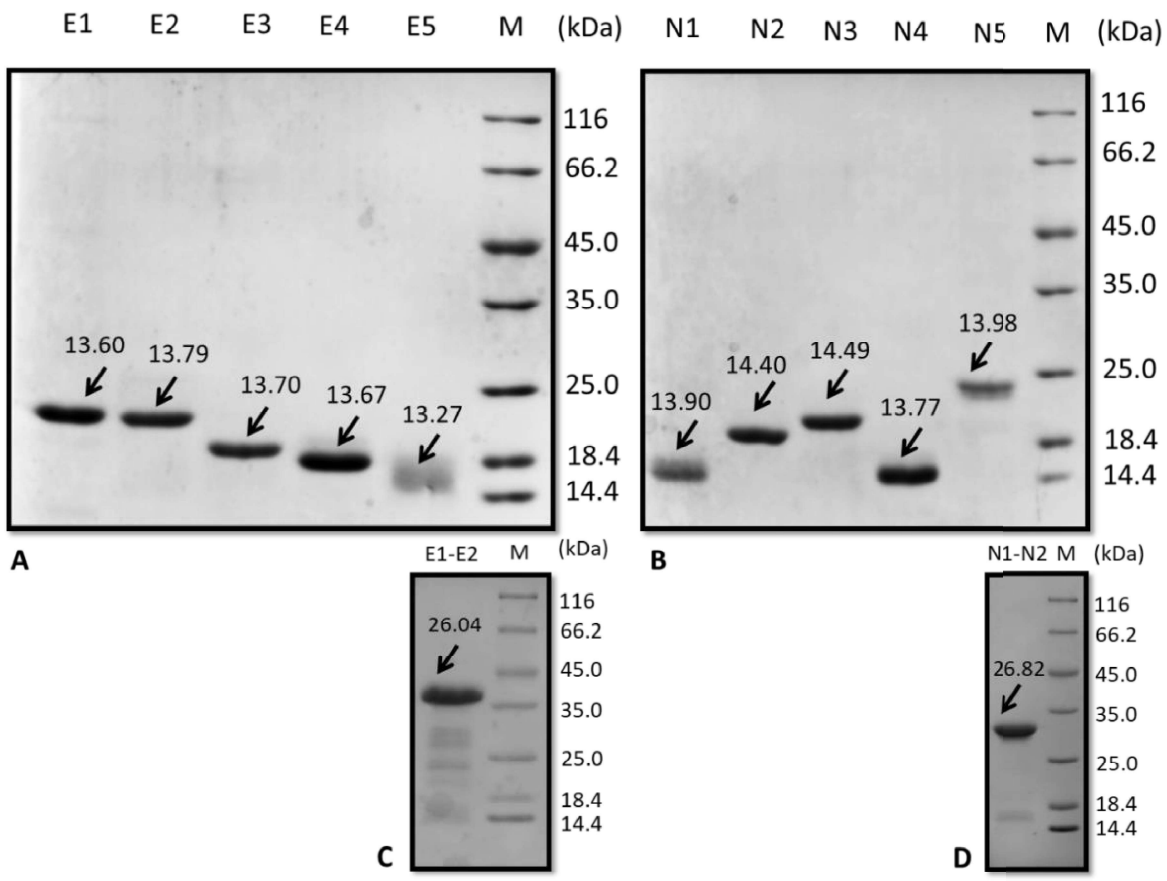


Figure 5.1. SDS-PAGE profiles of cadherin constructs. *Panel A.* AM-DRE behaviour of the five extracellular domains of E-cadherin (E1, E2, E3, E4, and E5). *Panel B.* AM-DRE behaviour of the five extracellular domains of N-cadherin (N1, N2, N3, N4, and N5). *Panel C.* AM-DRE behaviour of the fusion construct incorporating E-cadherin domains, E1 and E2. *Panel D.* AM-DRE behaviour of the fusion construct encoding N-cadherin domains, N1 and N2. The actual molecular weights of the domains and fusion constructs are shown on the gels, with arrows.

Interestingly, the ratio of negative charges (NC) to positive charges (PC) in the amino acid sequence of each construct (which is a key determinant of pI of the protein construct) could be directly corrected with this unusual behaviour. We called this unusual SDS-PAGE behaviour, AM-DRE (or 'anomalous mobility' during denaturing and reducing

electrophoresis), and use this term to refer to the electrophoretic migration of proteins on SDS-PAGE which is less or more than that which is expected.

We concluded that all constructs display AM-DRE if the ratio of NC: PC is greater than 1.50. On the other hand, if the ratio is smaller, proteins show normal mobility during denaturing and reducing electrophoresis (NM-DRE), with the extent of AM-DRE being in proportion with the AM-DRE. Exceptions were observed in the case of E1 and N5 in which a disproportionately high degree of AM-DRE could be observed, despite their NC: PC ratios being below 2.00. The reason for this exception could be differences in the actual locations of the negatively charged residues, because both E1 and N5 have a substantial number of 'clusters' of negatively charged residues which are located either next to each other (e.g., ED) or separated by one residue (e.g., EXE or EXD). Such clusters are less abundant, or inconsistently present, in other domains. Therefore, the presence of these charge clusters can bias the correlation. It was also observed that the fused domains (e.g., E1-E2, N1-N2 etc.) which are made up of domains showing AM-DRE also obeyed the correlation and *vice-versa.*

How the is highly negative charge content causing anomalous mobility? To answer this, we could think of the sample preparation step where SDS binding to the denatured polypeptide chain occurs. The high NC: PC ratio can cause insufficient binding of SDS molecules to the polypeptide chain by repelling similar charged entities and thus potentially affect the ability of SDS to bind to the protein to the reasonable extent (one SDS molecule for every two amino acids along the chain). SDS molecules are negatively charged, and they could thus be disproportionately attracted to, or repulsed by, regions of a protein's polypeptide chain which are rich in negatively charged, or positively charged, amino acid residues. Similarly, the overall charge present on a protein could, in theory, determine the ease with which SDS approaches it. The high negatively charged content of the cadherins presumably contributes to the overall negatively-charged aspect of the surfaces of cells, where the charges present on the head groups of lipids in the membrane also contribute to cell surface negative charge.

5.2 Evaluation of structural contents, stability and calcium binding of single domains and double domain fusions of E-cadherin

5.2.1 Evaluation by secondary structure by CD spectroscopy

The far-UV CD spectrum of any protein indicates the contributions of peptide bonds in various secondary structures present in that protein. The CD spectra of E1, E2, and N1 (Figure 2) establish that the purified domains are well folded, with formed secondary structure, indicative of a β-sheet pattern or at least a structure dominated by beta sheets. The CD spectra display negative peaks at 216-218 nm with a relatively low mean residue ellipticity (MRE) in the range of -1500 to - 4000 cm^2 $dmol^{-1}$. In the case of N2, the minor contribution evident as a peak at 202 nm is indicative of random coil content, present along with β-sheets (evident from the peak at 217 nm).

There are three calcium binding sites present at the junction of each domain with the domain following it along the chain. Calcium binding to cadherin extracellular domains is presumed to promote dimerization, which is an essential function in cell-cell interactions. Therefore, we also checked the effect of calcium on the secondary structure of the domains by CD spectral measurements, where 0.0, 0.1, 1.0 and 10 mM $CaCl_2$ were used, and domains were incubated for 8 hours with calcium. We did not notice any significant change in secondary structure in any of the 10 individual domains constructs studied. The reason could be either than calcium binding subtly alters the structure of the small inter-domain region but not the structure of the domain preceding it, or that the calcium binding itself is compromised by the fact that the next (neighbouring) domain along the cadherin's sequence is not present in the construct. In separate studies, we show that there is an effect of calcium on the domain stability, but in this section, we are focusing on whether there were changes in structure.

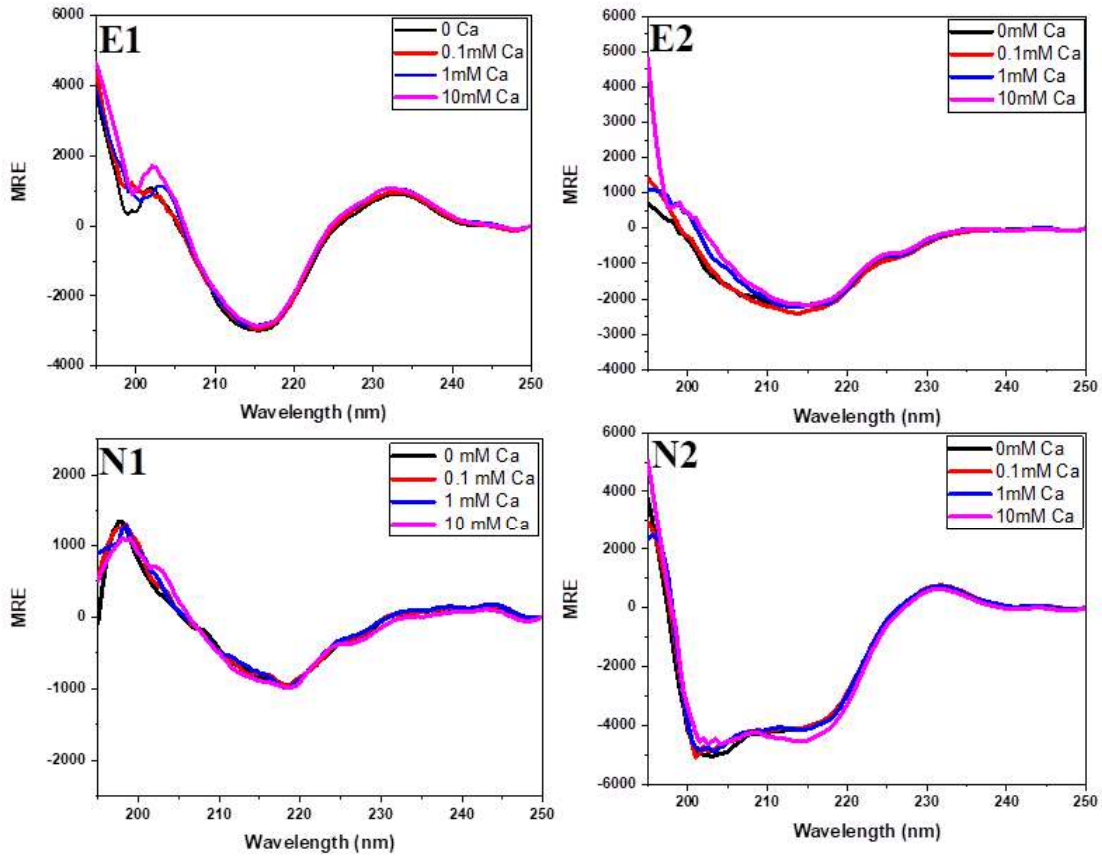


Figure 5.2. Circular Dichroism spectra of individuals domains 1 and 2 of E-and N-cadherins

The E3, E4, and E5 domains showed a sharp peak at 198-200 nm indicative of the contribution from random coil content, suggesting the populations are primarily unstructured (Figure 3). These domains are mainly unstructured, and during purification also these proteins were found in a mostly insoluble fraction. A unique peak observed in the case of E5, where a positive peak near 330 nm was present. Usually, this peak could be the contribution from any disulfide bond as E5 contains four cysteine residues which might be contributing to this positive peak. Like other individual domains, domains 3, 4, and 5 also did not show any change in secondary structure when treated with calcium.

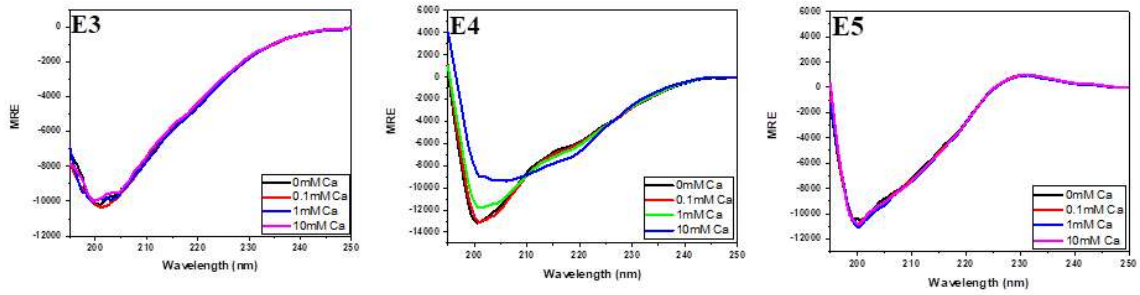


Figure 5.3. Circular Dichroism spectra of individuals domains 3, 4 and 5 of E-cadherin

The fusion domains E1-E2 and N1-N2

The CD spectrum of E1-E2 shows a 218 nm peak and a strong peak at 198 nm indicative of β-sheet mixed with random coil contribution. Upon addition of the calcium, the β-sheet contribution gets enhanced, as is evident from the shifting of the 198 nm to 218 nm. This suggests that neighbouring domains are required for binding of calcium, which causes critical structural changes as are necessary for dimerization.

Similarly, N1-N2 seems to be well-folded with formed secondary structure indicative of β-sheet content, with spectral shape peaking at 216 nm (Figure 4). The N1-N2 fusion also shows structural change upon binding of calcium, with shifting of a 205 nm band towards the development of a strong 216 nm band in the presence of calcium. This confirms that binding of calcium, occurring in the presence of the neighbouring domain immediately following, or preceding, E1, N1, E2, or N2, results in a significant change in domain structure.

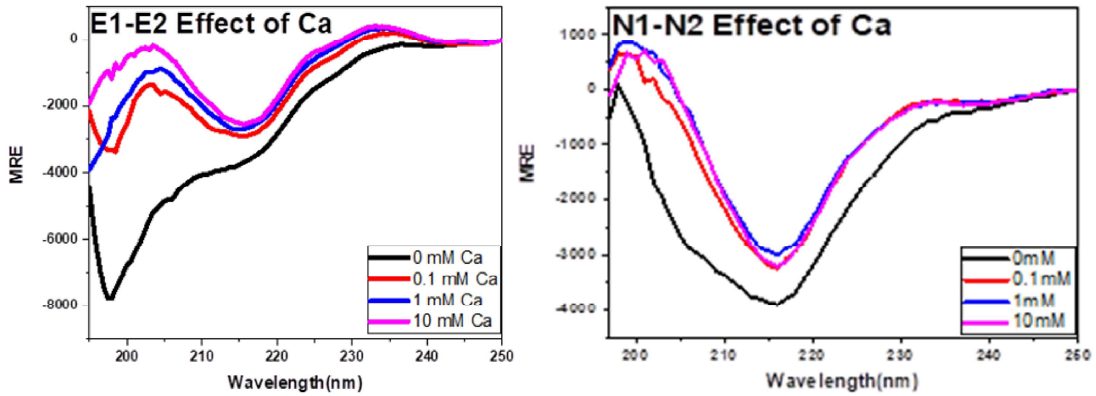


Figure 5.4. Circular Dichroism spectra of fusion domains E1-E2 and N1-N2

5.2.2 Evaluation of tertiary structure formation by fluorescence spectroscopy

The emission spectra of E1, N1, and N2 were found to be in the range of 337-340 nm, well below the value of 353 nm which is indicative of a tryptophan residue which is completely exposed to the aqueous solvent. This confirms that the three proteins are folded and that their tryptophan residues are buried significantly inside their hydrophobic cores. E2 does not contain any tryptophan as only tyrosine residues are present. This results in the emission spectrum peaking at 307 nm (a wavelength of maximal fluorescence emission which is characteristic of tyrosine) irrespective of the polarity of the tyrosine residue's environment.

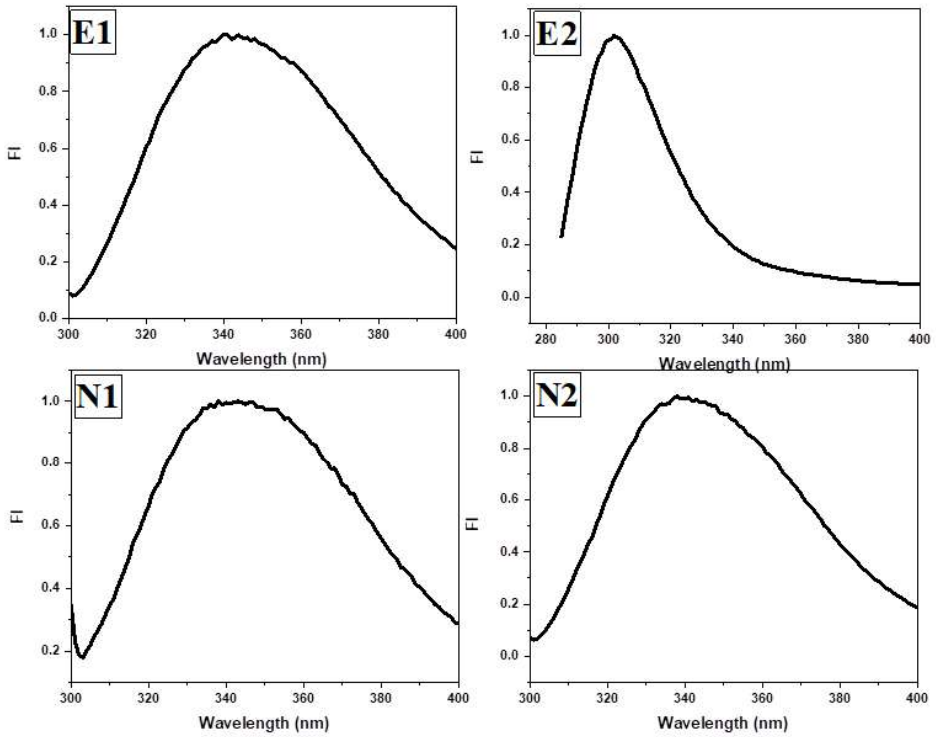
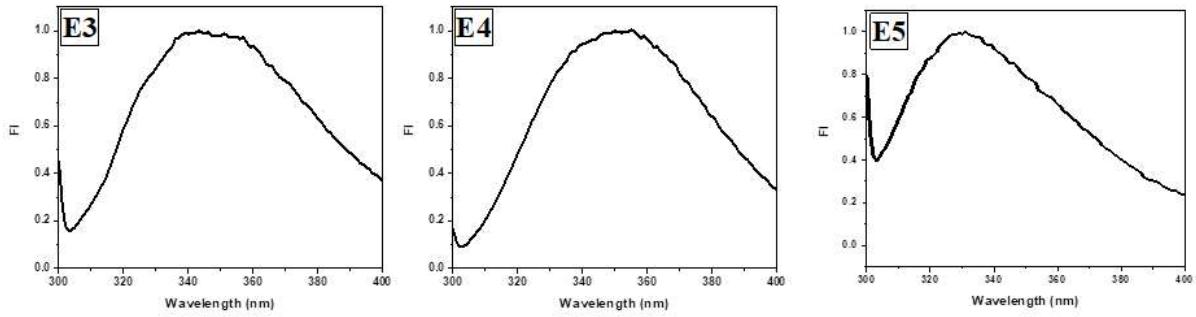


Figure 5.5. Fluorescence emission spectra of individual domains 1 and 2 of E-and N-cadherins

The E3 domain shows a mixed emission spectral envelope in which peaks at 340 nm and 356 nm are both presents which indicates the partial burial of tryptophan residue. Therefore, E3 contains both the unfolded and folded population which also correlated with CD spectra (dominated by the random coil). The folding of E4 is even worse as we can see the emission maximum at 357 nm which means the protein is predominantly unfolded. E5 shows an emission maximum at 330 nm indicating well-folded protein, confirmed by its high solubility inside *E.coli* cytoplasm.

Figure 5.6. Fluorescence emission spectra of individual domains 3, 4 and 5 of E-cadherin



5.2.3 Evaluation of quaternary structure by gel filtration chromatography and light scattering analysis

Cadherins form dimers in solution as they undergo homo-dimerization on the cell surface to form both *cis* dimers amongst cadherins on the same cell, and *trans* dimers between cadherins facing each other on juxtaposed cells. To check the oligomeric status of individual domains, purified proteins were run using a Superdex-75 gel filtration column pre-equilibrated with 50 mM Tris buffer, pH 7.5.

The elution volumes of E1 and E2 domains are 11 and 10.75 ml respectively, which corresponds to masses of 51 and 58 kDa, as can be seen in Figure 7. This suggests either the formation of tetramers or an overestimation of mass owing to the domains having an ellipsoidal (rather than spherical) shape in dimeric form since this can lead to an larger hydrodynamic volume. However, glutaraldehyde crosslinking data showed the presence of monomers and dimers, but no tetramers, as can be seen in Figure 8. This suggests that the populations seen eluting from gel filtration are mainly dimers. It may be noted that the low probability of inter-subunit crosslinking of polypeptides by glutaraldehyde (compared to intra-subunit crosslinking) always leads to underestimation of the higher order forms, due to dissociation of subunits lacking a minimum of one inter-subunit crosslink, by SDS, during denaturing gel electrophoresis. Also, it may be noted that glutaraldehyde can cause the formation of intramolecular crosslinks. During sample preparation for SDS-PAGE, heating and SDS are unable to denature the population of proteins in which such intramolecular crosslinks form, and this can result in the observation of a sub-population which migrates

faster than the bulk of the population (appearing to possess a lower molecular weight). Evidence of this is seen in Figure 8, in which virtually all samples also have a band running below the monomer which consists of such intramolecularly cross-linked species.

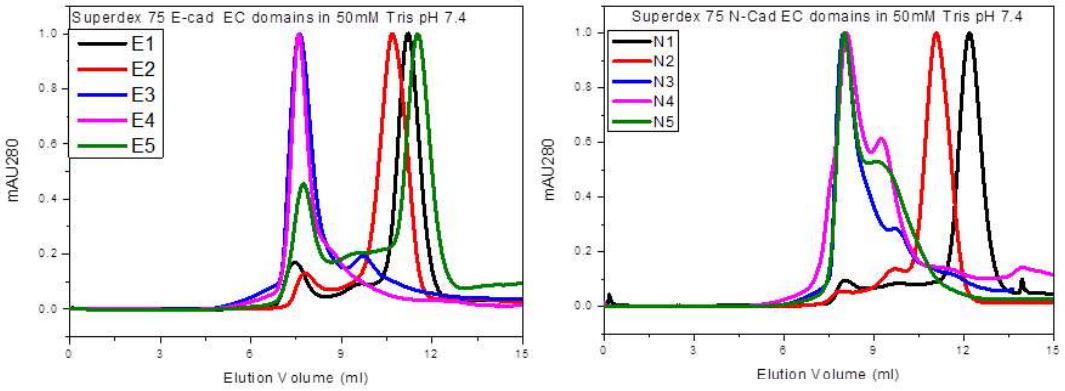


Figure 5.7. Size exclusion chromatography (SEC) profiles of all the 10 individual domains of E- and N-cadherins using Superdex 75 column

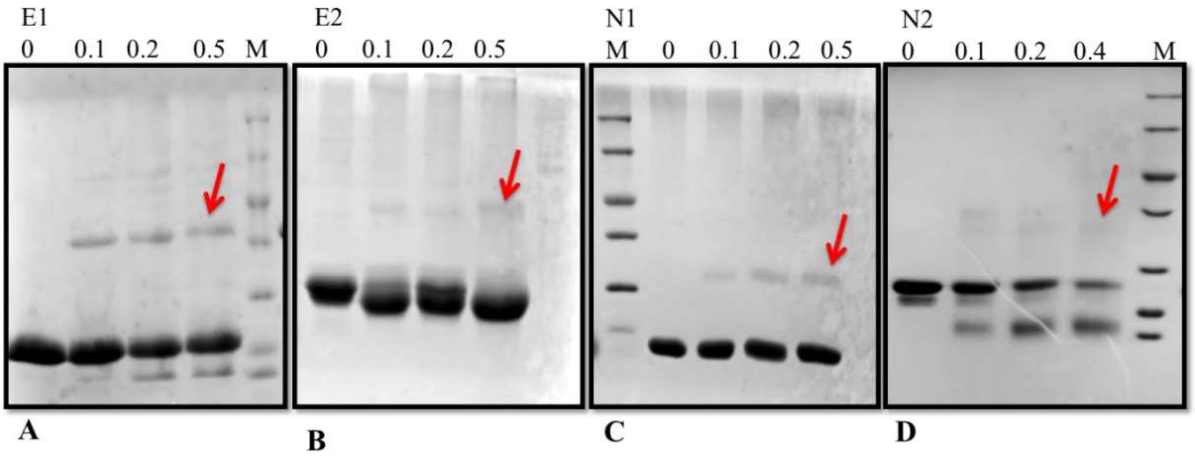


Figure 5.8. Glutaraldehyde crosslinking results for individual domains 1 and 2 of E-and N-cadherins (Panel A-D)

N1 elutes at 12.1 ml from the Superdex-75 column, which corresponds to a mass of approximately ~29 kDa. Here too, crosslinking shows the presence of dimers. Similarly, N2 also elutes at 11.75 ml which correspond to approximately 58 kDa mass suggesting that the

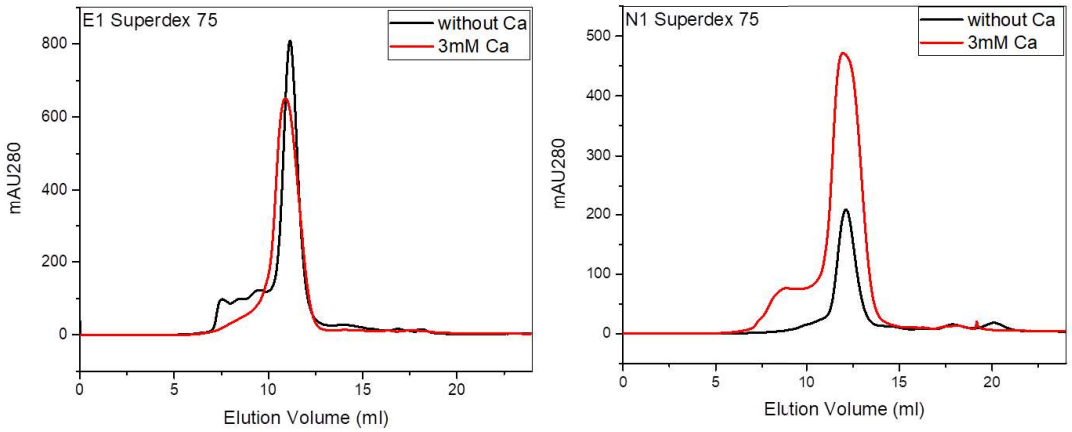
population is dominated by tetrameric species. N2 also did not show a significant presence of any higher order oligomers.

Part of the reason for the lack of clarity regarding the gel filtration data and what it indicates stems from the fact that cadherin domains form a rod-like shape, rather than a globular shape, due to the nature and arrangement of β-sheet-based structures. Size exclusion chromatography is suitable as a technique for the determination of quaternary structure for proteins possessing a globular shape, but not very suitable for rod-shaped proteins. As already mentioned above, the hydrodynamic volume of a protein depends on its shape and not on the mass, and a rod-shaped protein can thus migrate like a protein with greater mass than it actually possesses because the rotational motion of the protein causes it to behave like a sphere with a diameter slightly larger than its longest rod-like dimension. Therefore, E1, E2, N1 and N2 do not seem to form oligomers, and the early elution through the column was due to their rod-shaped structure in solution and their existence presumably as a mixture of dimers and monomers, with some fraction of the dimers being captured by the glutaraldehyde experiment.

E5 elutes at an elution volume of ~11.9 ml from the Superdex-75 column, which corresponds to a mass of ~30 kDa. The result suggests that the population is dominated by mainly by dimeric species; however, contrary to this, the crosslinking experiments did not reveal any dimeric forms of E5, although this could also be because there are no glutaraldehyde-reactive sidechains at the dimeric interface if indeed the dimer exists. The reason was similar to that of other individual domains where rod-shaped structure exhibits a larger hydrodynamic volume than the expected. E5 does not form soluble aggregates during purification which also got confirmed by gel filtration and CD results.

The domains E3, E4, N3, N4, and N5 were eluted in the void volume, suggesting the presence of soluble aggregates in solution. This property of forming soluble aggregates was also confirmed by other observations like CD spectra and presence of protein in an insoluble form (usually an outcome of the formation of soluble aggregates which can tend to grow larger and become insoluble).

The gel filtration profiles of E1 and N1 domains were compared in the presence of calcium, but no significant change was seen, as is evident from the gel filtration chromatograms in Figure 9. This confirms that these individual domains do not respond to binding of calcium in the absence of their neighbouring domains.



E1 without Ca 11.1 ml 49 kDa
3mM Ca 10.9 ml 54 kDa

N1 without Ca 11.9 ml 32 kDa
3mM Ca 12.1 ml 29 kDa

Figure 5.9. Effect of calcium of E1 and N1 (Gel filtration profiles)

Size exclusion (gel filtration) chromatography was also performed for the E1-E2 fusion domain, both with and without calcium present, to observe the effect of calcium upon the fusion construct which possesses three calcium-binding sites in the inter-domain region. On a Superdex-200 column, the elution volume of the construct was 15.8 ml in the absence of calcium, which corresponds to a mass of ~54 kDa. However, a small sub-population eluted in the void volume suggesting the presence of soluble aggregates in equilibrium with the ~54 kDa population. This data indicates that the population is mainly dimeric (with a monomeric fusion construct mass of ~24.04 kDa). Satisfyingly, during crosslinking experiments also, a significant sub-population showed dimeric status, in consonance with the dimeric status reported by the gel filtration studies. If the two rod-shaped E1-E2 fusion constructs were to happen to line up with each other in parallel, as would be expected during the formation of a *cis* dimer of cadherins interacting amongst themselves, on the surface of the same cell, then the rod-shape would become much more globular and this could explain the greater

consonance between elution volume (which depends upon hydrodynamic diameter) and molecular mass.

When the experiment was performed in the presence of calcium, no significant elution volume was observed at the column's void volume; instead, all the proteins were eluted at 15.8 ml, with little evidence of the formation of soluble aggregates, as can be seen in Figure 10. The difference in the results obtained with and without calcium suggests that, in the presence of calcium, the equilibrium between natively formed dimers and soluble aggregates shifts entirely to dimers. Glutaraldehyde crosslinking results also showed an increased intensity of dimeric population in the presence of calcium, as can be seen in Figure 11. This conclusion was also supported by circular dichroism spectroscopy results which indicate structural changes of the E1-E2 fusion construct in the presence of calcium.

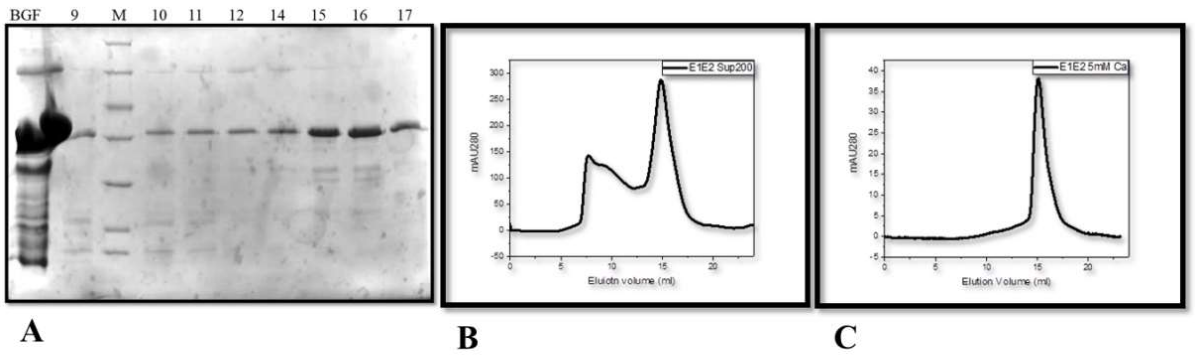


Figure 5.10. Size exclusion chromatography of E1-E2 using Superdex 200 column Panel A. SDS-PAGE profile after the run, Panel B. SEC profile in the absence of calcium, Panel C. SEC profile in the presence of calcium

With the N1-N2 fusion construct, as can be seen in Figure 12, the elution of protein occurred at 15.8 ml in the absence of calcium which suggests that a dimeric population is pre-existent. The main peak gets shifted to 17 ml when calcium is added, with a shoulder at 16 ml suggesting that an equilibrium exists between monomers and dimers. However, glutaraldehyde crosslinking experiment showed increased in dimeric population in the presence of calcium. This indicates that calcium is required for proper folding and

dimerization of both E1-E2 and N1-N2 fusion constructs, which is a crucial event in cell-cell interaction.

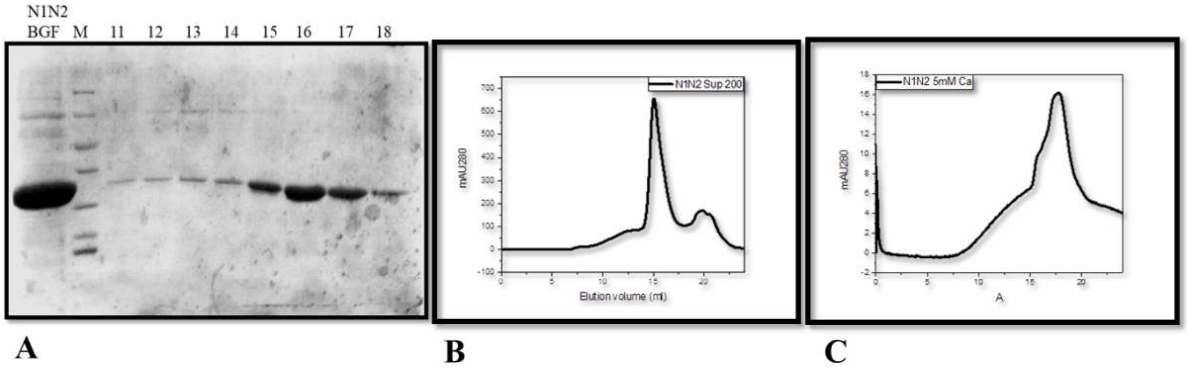
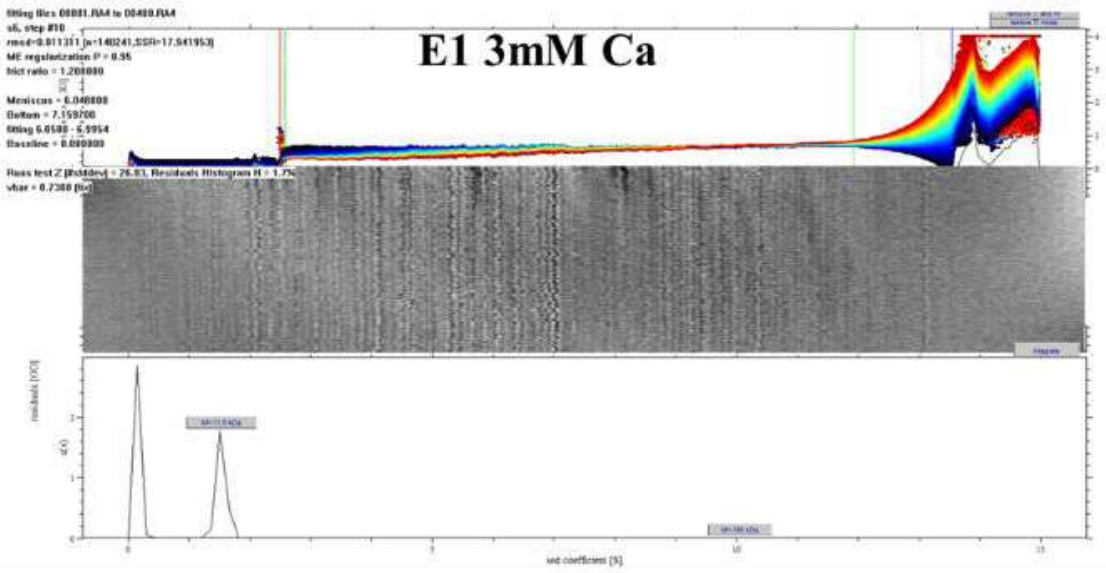


Figure 5.11. Size exclusion chromatography of N1-N2 using Superdex 200 column Panel A. SDS-PAGE profile after the run, Panel B. SEC profile in the absence of calcium, Panel C. SEC profile in the presence of calcium

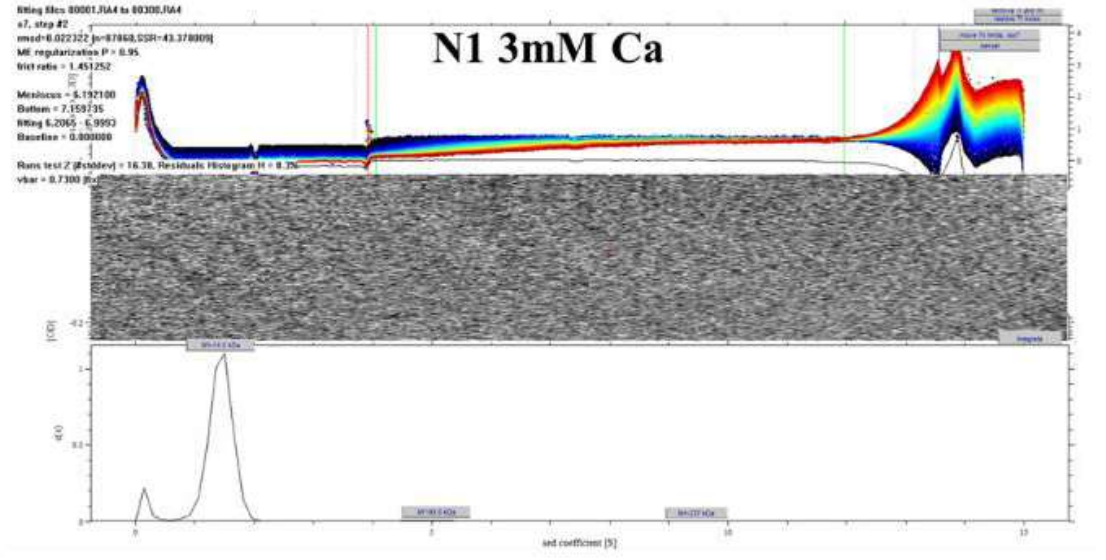
5.2.4 A comparison of the oligomeric status of E1-E2 and N1-N2 fusion domain constructs by analytical ultracentrifugation (AUC)

As discussed above, cadherin EC domains appear to form dimers, and the dimerization process is enhanced in the presence of calcium, especially when dealing with double-domain fusion constructs like E1-E2 or N1-N2. Initially, the conclusions were drawn from size exclusion (gel filtration) chromatography coupled with glutaraldehyde crosslinking studies. However, all analyses and results did not appear to be strictly correlated. Therefore, we decided to use an additional technique of analytical ultracentrifugation (AUC) to examine the formation of dimers. We performed AUC experiments in which the measured sedimentation coefficient is directly proportional to the molecular weight of the proteins being analyzed. Thus, for rod-shaped protein subunit structures, AUC is a more suitable technique for analysis of oligomerization.

From the existing literature, it is well established that the first EC domains of classical cadherins, i.e., domains E1 or N1, are involved in dimerization interactions amongst E- or N-cadherins, respectively. Here, we proceed to examine whether N1 or E1 are independently capable of dimerizing, or also require the presence of the next EC domain in the chain, i.e., E2 along with E1, an N2 along with N1, respectively. We performed AUC for E1 as well as for N1 in the presence of 3 mM calcium, as can be seen in Figure 13. The data were fitted using the SEDFIT program. The AUC results demonstrate the existence of only monomeric populations of E1 and N1. Notably, no dimeric species are seen even when the AUC is carried out in the presence of calcium (Figure 13). The control experiments without calcium also showed the presence of monomer (data not shown), but the data could not be fitted very well.



A



B

Figure 5.12. Analysis of oligomeric status of E1 and N1 using AUC with 3 mM calcium, *Panel A.* AUC profile of E1 showed only monomeric species, *Panel B.* AUC profile of N1 also showed the only monomeric population in the presence of calcium. In both the cases one more population was seen smaller to monomeric size but that is just an artefact of the data fitting.

It was observed previously, however, that dimers are formed by the fusion constructs, E1-E2 and N1-N2. We did AUC experiments with fusion constructs E1-E2 and N1-N2. As can be seen in Figure 14, both the fusion domains formed dimers in the presence of calcium. The absence of calcium appeared to completely prevent the formation of dimers since no dimers were detected in the AUC data. Notably, the extent of dimer formation was slightly lower in the case of E1-E2 (Figure 14), than in N1-N2 (Figure 15). Also, there appeared to be a close overlap between the monomer sub-population and the dimer sub-population. Probably, both the sub-populations co-exist in equilibrium and not able to separate completely, in both the fusion constructs.

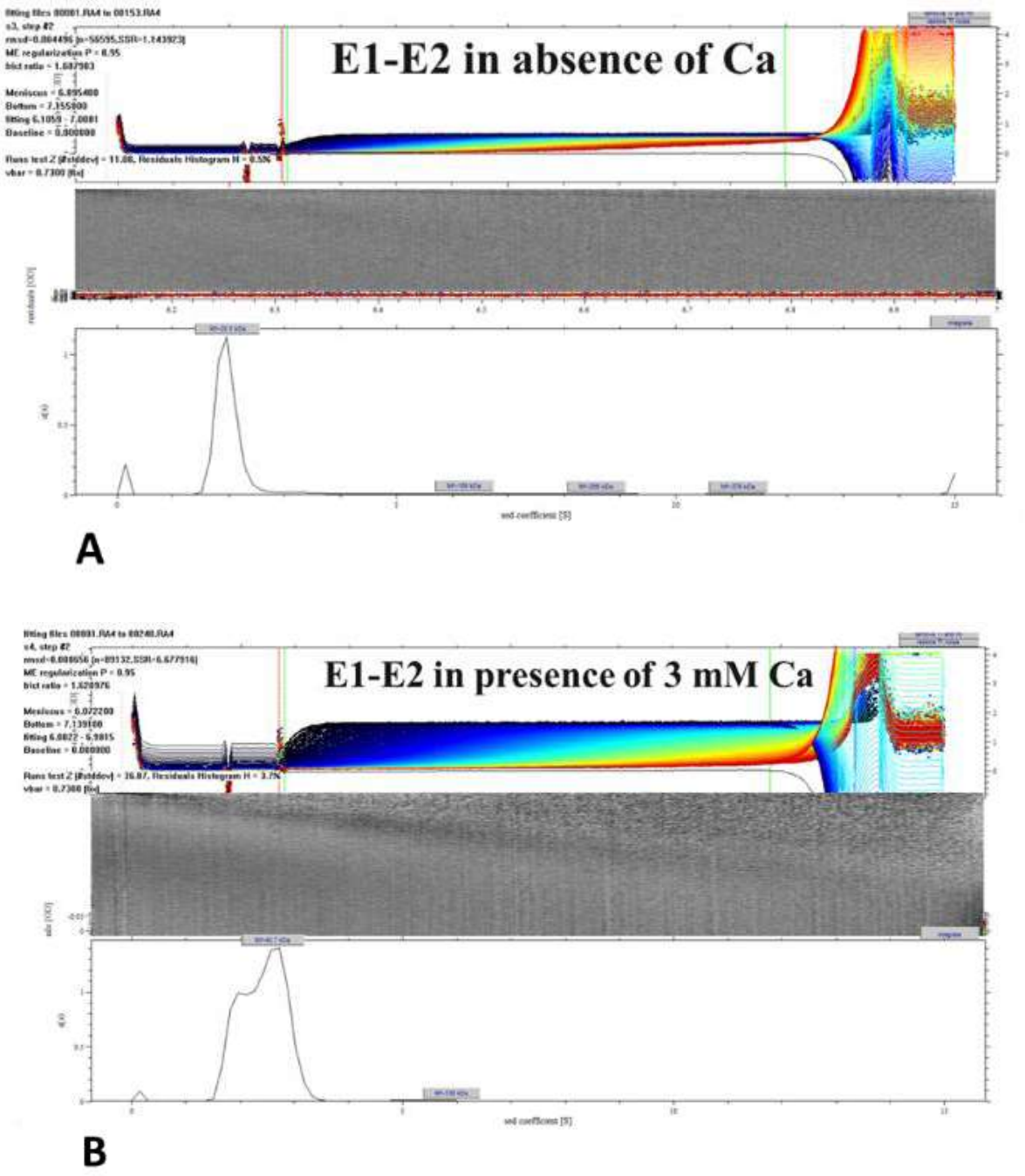


Figure 5.13. Analysis of oligomeric status of E1-E2 using AUC, *Panel A*. AUC profile showed only monomeric species in the absence of calcium, *Panel B*. AUC profile showed an equilibrium exists between monomers and dimers in the presence of calcium.

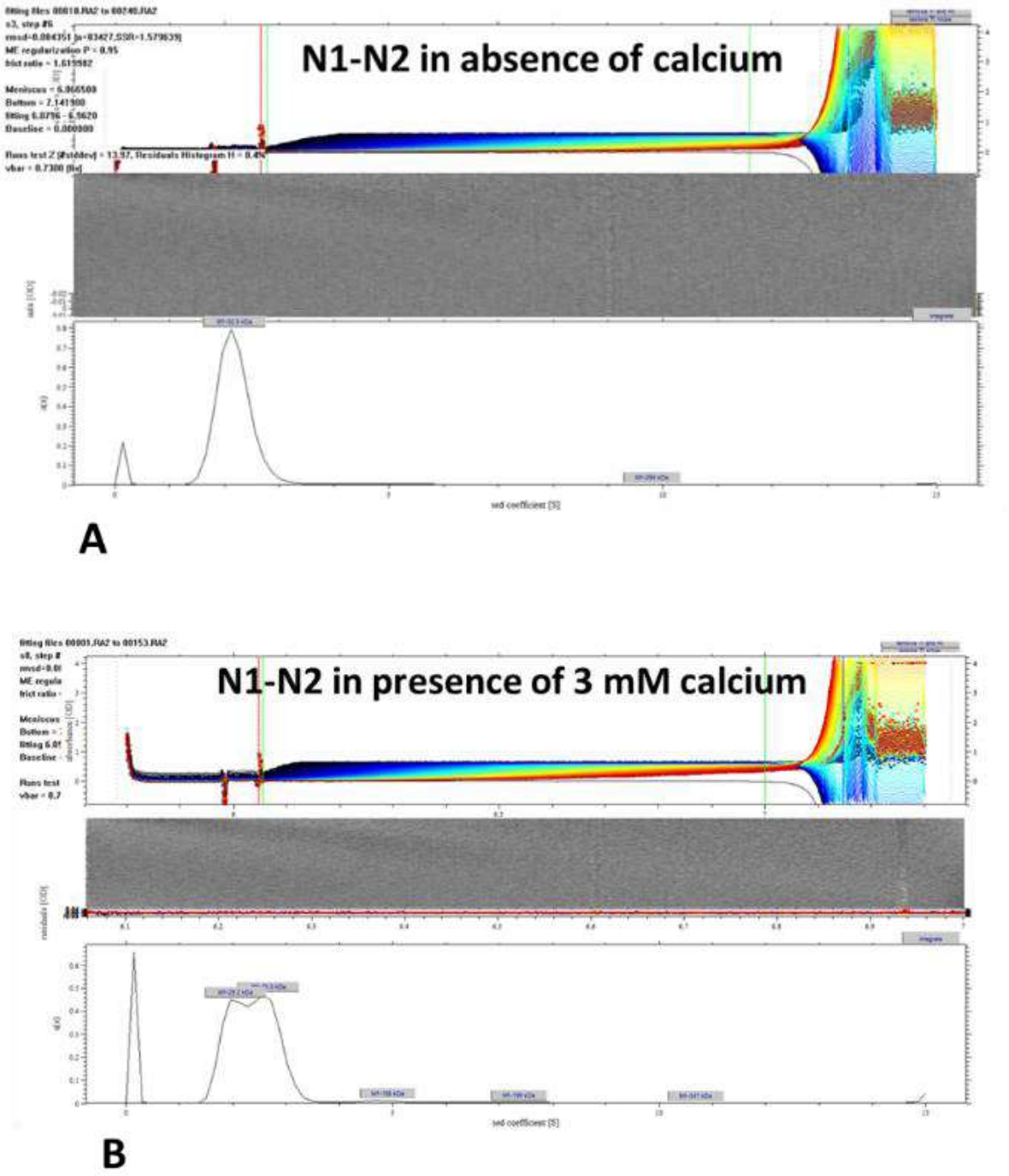


Figure 5.14. Analysis of oligomeric status of N1-N2 using AUC, *Panel A*. AUC profile showed only monomeric species in the absence of calcium, *Panel B*. AUC profile showed an equilibrium exists between monomers and dimers in the presence of calcium.

5.2.5 Evaluation of structural stability of EC domains of E- and N-cadherins by differential scanning calorimetric (DSC)

Differential scanning calorimetry (DSC) measures the heat capacity and the additional heat associated with transitions during unfolding-refolding events. It can also be used to determine the folding mechanisms and relative thermodynamic stabilities of proteins and protein domains. In this section, we examine the melting temperatures and unfolding-refolding behaviour of cadherin domain constructs, in the absence and presence of calcium.

As can be seen in Figures 16-18, The addition of 3 mM calcium stabilized the individual domains by less than 5 ºC but had a significant effect on the behaviour of the fusion domains. We anticipated that this difference would be seen because calcium binds to the linker regions, and it is conceivable that such binding both occur more efficiently when the linker is flanked by domains on both sides and is also effective in increasing the stabilities and/or rigidities of the two flanking domains. In contrast, while the individual domains do contain the linker regions at their tails, the absence of the other flanking domain in a single domain construct could be expected to affect the efficiency of the binding (i.e., extent of saturation of binding) and also result in a less noticeable effect on the structure or stability of the single domain, even when some calcium-binding does occur.

Notably, for individual E1, E2, N1 and N2 domains, we saw a ~5 ºC increase in Tm in the presence of 3 mM calcium chloride, even though no structural change could be observed in the structure by circular dichroism. E1 and N1 constructs were found to be the most thermally stable domains amongst all the extracellular domains of the E- and N-cadherins, with Tm values of 61 and 60.77 ºC, respectively. In the presence of calcium, somewhat surprisingly, E1 showed a decrease in Tm. The other three domains, i.e., E2, N1 and N2, showed increased thermal stability, in contrast, in the presence of calcium. E1 showed the ability to refold without calcium, whereas N1 did not show any ability to refold. The possible reason could be that calcium provides some rigidity and that this decreased the plasticity of the domain. This possibility is also supported by the fact there is some aggregation in the presence of calcium. Furthermore, the stability of E1 could be because of the two cysteine residues which are bonded into a disulfide bond.

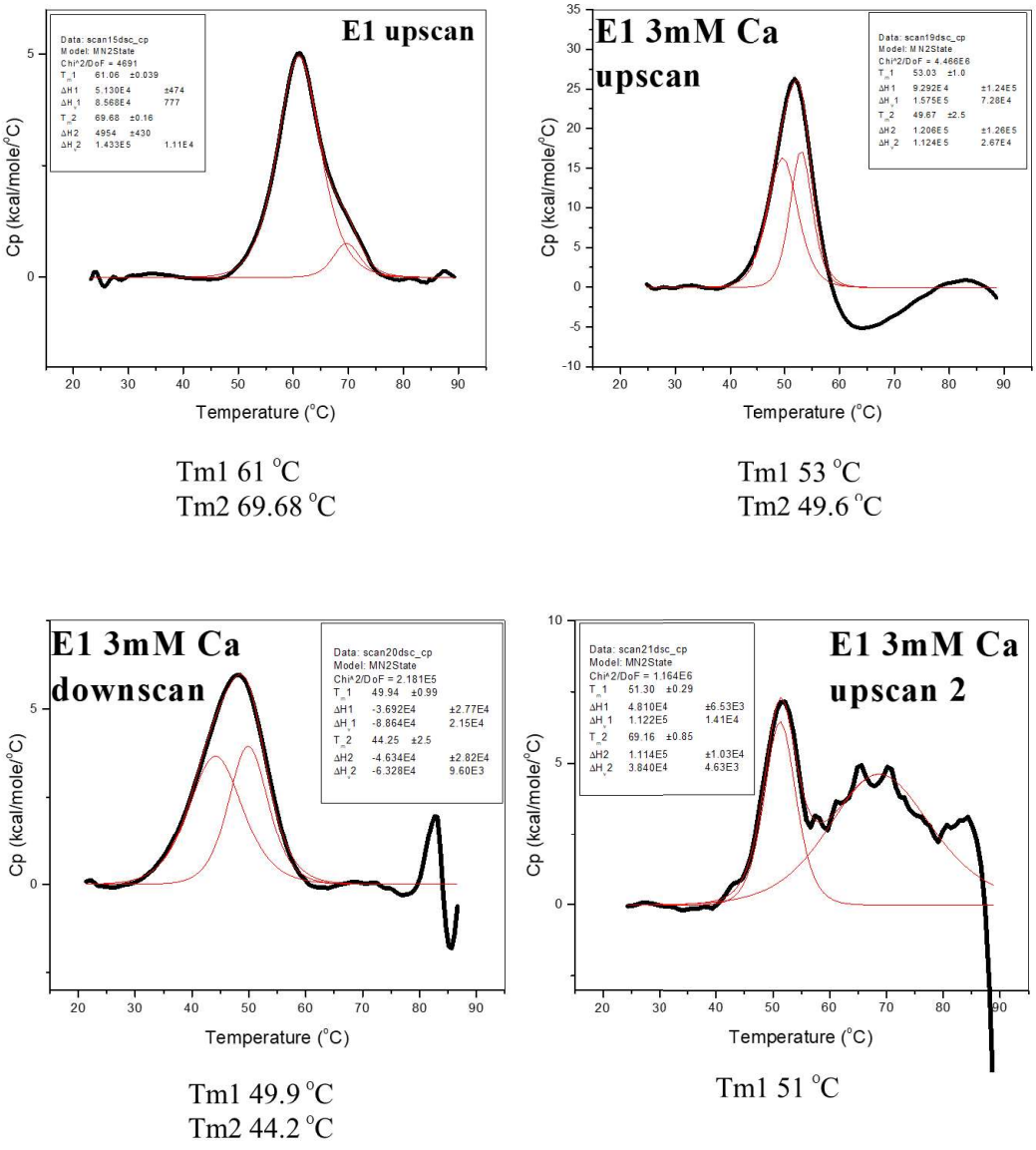


Figure 5.15. DSC profiles of the E1 domain with, and without, calcium. The second cycle of denaturation and refolding has also been reported.

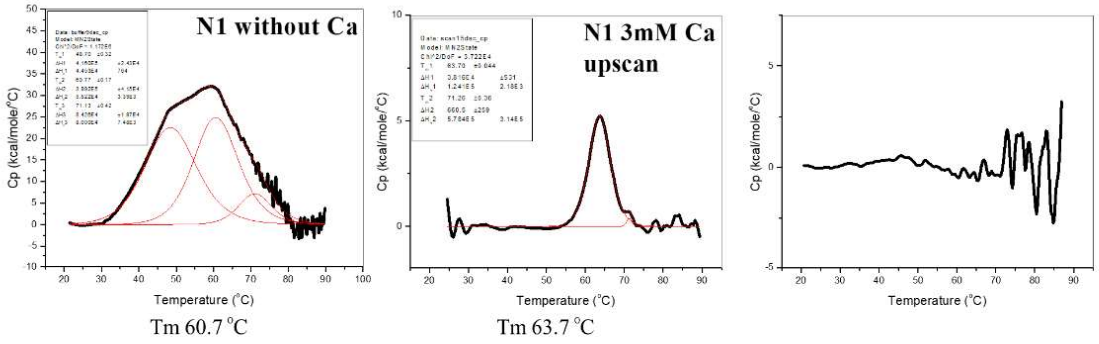


Figure 5.16. DSC profiles of the N1 domain in the absence of calcium. In the second cycle, the protein was not able to refold and aggregated.

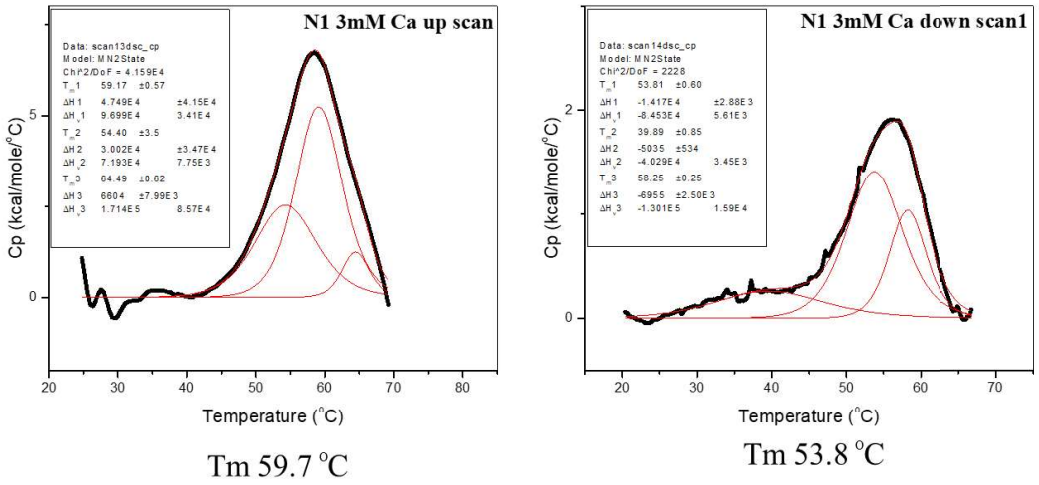


Figure 5.17. DSC profiles of the N1 domain in the presence of calcium. The upper-temperature limit was reduced to 70 ºC, and it showed refolding.

During a DSC experiment, temperature ranges of upscan (denaturation) and downscan (renaturation) are often set to go 20 to 90 ºC, and 90 to 20 ºC, respectively, to complete all unfolding and refolding transitions, especially with proteins from mesophile organisms which are completely unfolded at 90 ºC and which also manage to remain folded, or manage to refold from an unfolded state, at a temperature of 20 ºC, when they manage to refold at all. As already mentioned, many of the cadherin domains showed both higher stability in the presence of calcium, as well as some tend to aggregate. Their ability to refold was not comparable in all cases. Since cadherins never get exposed to a temperature as high as 90 ºC, which could cause unfolding of even residual kernels of structure that could be required

for refolding to occur, we decided to decrease the upper limit of temperature during the upscans. Thus, we repeated the earlier experiments by setting upper limits which are much lower, i.e., much closer to the (already determined) Tm values of these protein domains, to obtain clean denaturation and renaturation data. In the case of N1, when we performed the experiment between 20 and 90 ºC, at the higher temperatures the protein tended to aggregate and could not be refolded back, but when we set the temperature between 20 and 80 ºC, N1 showed good refolding even in the presence of calcium. This indicated to us that during these DSC experiments we should set the upper-temperature limits above the temperature of the predominant melting transition(s) but not very much higher since otherwise, we could lose subtle kernels of formed structure which are critical for refolding ability.

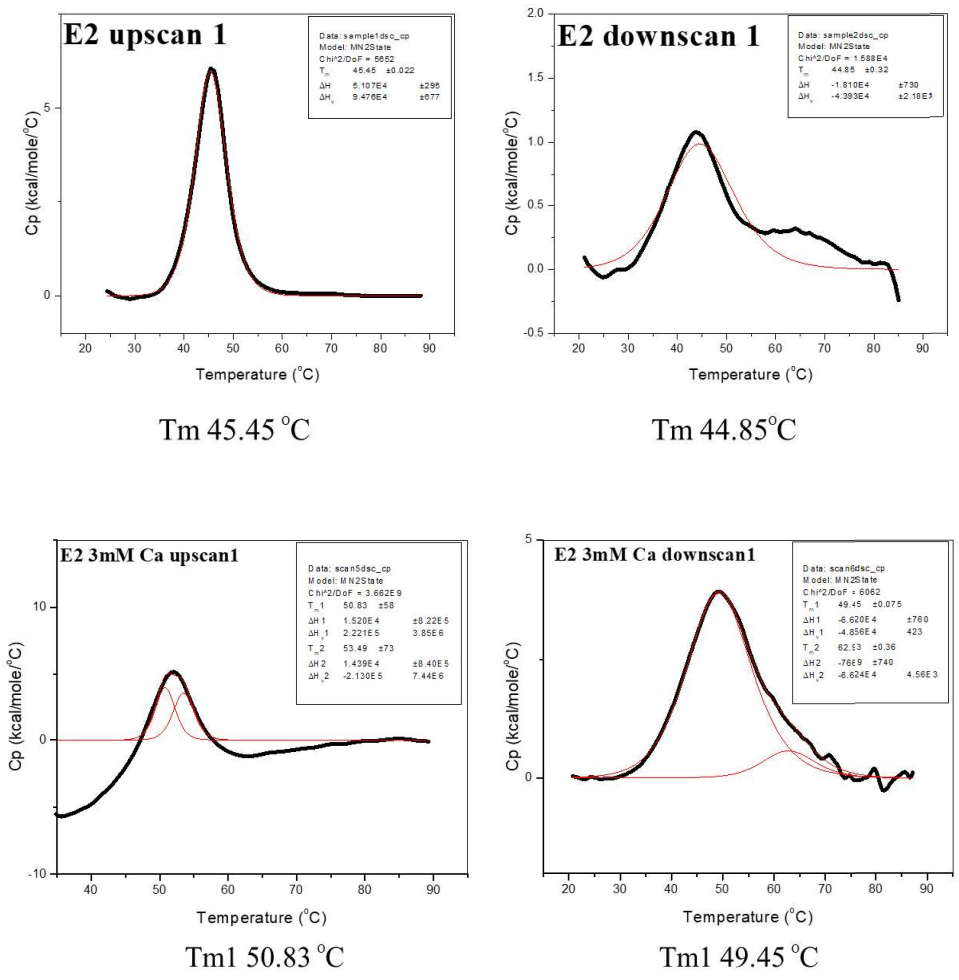


Figure 5.18. DSC profiles of the E2 domain in with and without the presence of calcium.

The Tm of the E2 domain was 45 ºC, suggesting that the domain is less stable than E1. However, it showed a peculiar ability to refold up to two cycles of upscan and downscan. Similarly, N2 also showed the refolding ability after melting at 55 ºC. The notable thing here was that unfolding and refolding were both very efficient. There was no change and no hysteresis in the upscan and downscan DSC patterns even over 22 cycles of unfolding-refolding. Therefore, the N2 domain possesses an extraordinary property of being able to refold after unfolding. This could have some physiological relevance as well. Interestingly, in each cycle, the value of heat capacity was to be almost unchanged, which shows that refolding occurred to completion in every cycle. In the presence of calcium, the Tm value was higher than in the absence of calcium, as with other domains; however, with calcium, the tendency to undergo aggregation was seen and refolding was also significantly poorer.

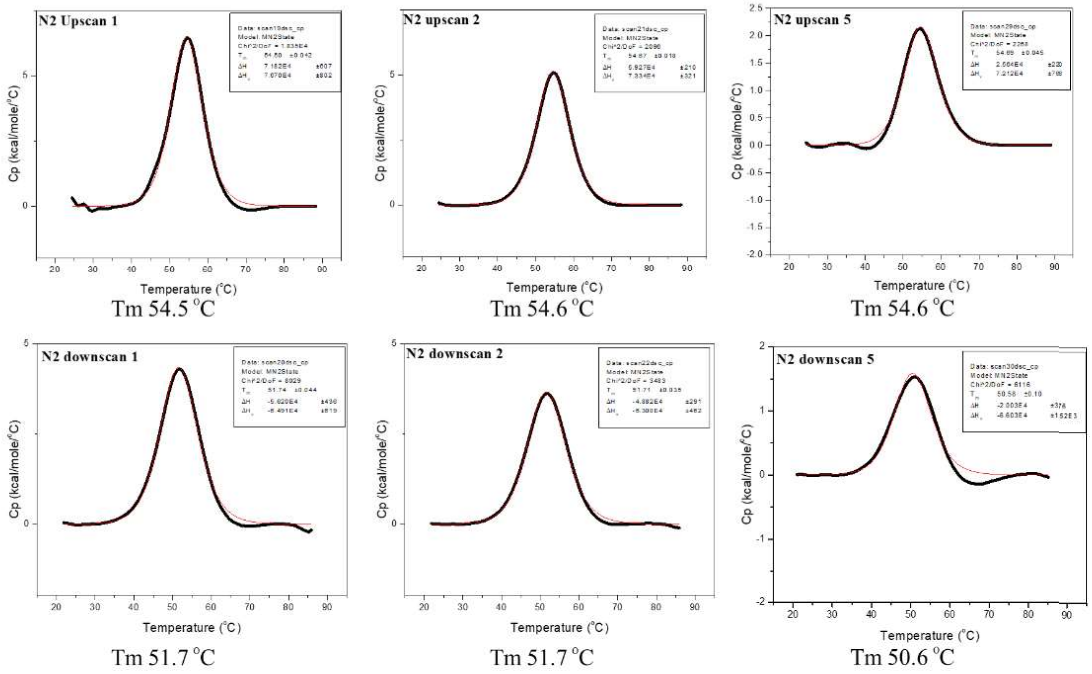


Figure 5.19. DSC profiles of the N2 domain in the absence of calcium. The domain was able to refold up to infinite times without any significant change in enthalpy.

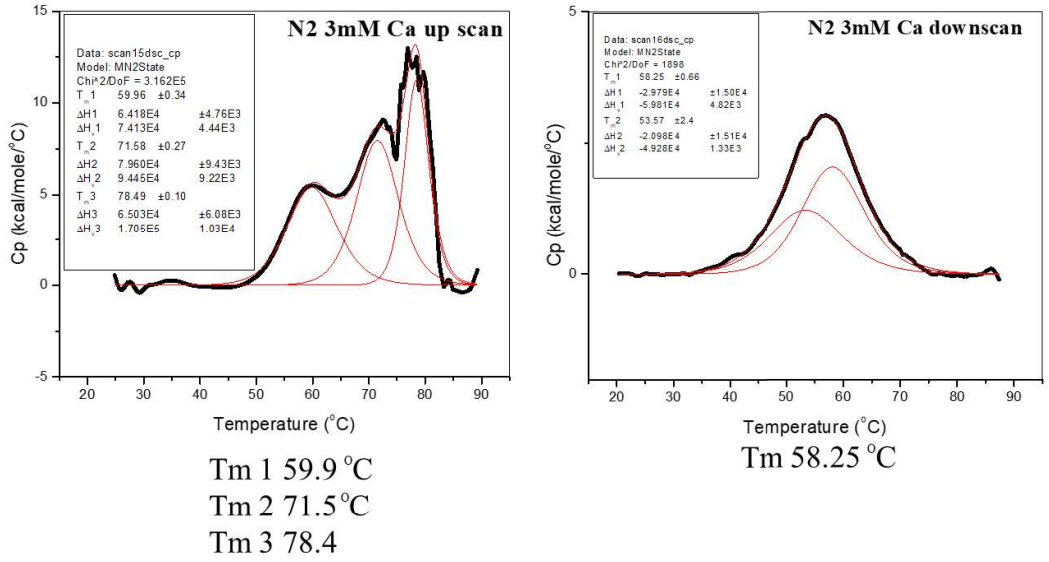


Figure 5.20. DSC profiles of the N2 domain in the presence of calcium. The refolding ability was still present but not as efficient as was in the absence of calcium. However, the stability got increased.

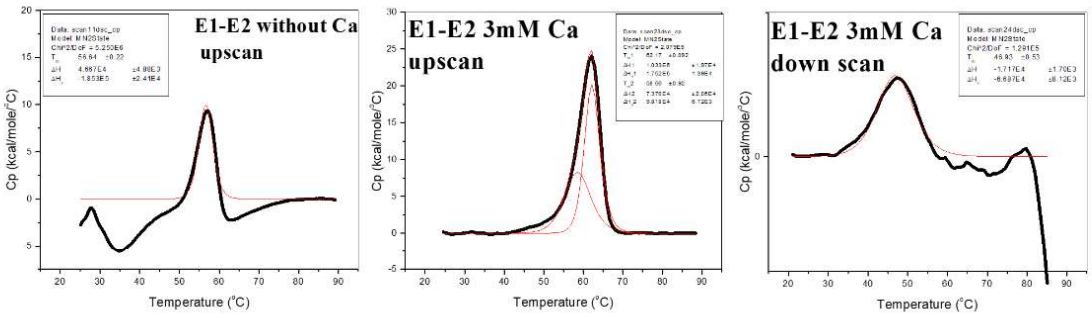
Therefore, N2 is very efficient in refolding whereas E2 is not. The physiological relevance of this can be understood by explaining the domain origins. E2 and N2, respectively, are both the second extracellular domain of epithelial, and neuronal, cadherins. We know that epithelial cells interact much more permanently than neuronal cells which keep on making and breaking contacts with other neuronal cells. During the making and breaking of contacts, these N-cadherins could undergo denaturation through mechanical forces, and the ability to refold could reduce the need for replenishment of N-cadherins at the surfaces of neuronal cells, especially given their terminally differentiated states. Now the question is why this property is present only in the second domain but not in the first domain, given that the first domain is the key player in cadherin dimerization.

There is ample evidence that the second domain (E2 or N2) is much more responsive to inter-domain interactions than the first domain (E1 or N1), in particular, in relation to the formation of X-dimers. Even if it is not involved in the direct interaction, or in domain swapping interactions, during cadherin dimerization accompanying cell-cell interactions, the second domain has a vital role in communicating between the first domain and the remaining domains. Furthermore, the second domain is an important site for both lateral and X-dimer

interactions this domain has mechanistic implications in cadherin dynamics. (Harrison, Bahna et al. 2010, Hong, Troyanovsky et al. 2011). A comparison of the behaviour of different domains is given in Table 2.

	Without Ca Up scan	Down scan	With Ca Up scan	Down scan	Without Ca Refolding	Without Ca	Without Ca Aggregation	With Ca
E1	61, 54	-	53, 51.7	49, 49	-	Yes	-	No
E2	45.5	44.5	51.6	49.8	Yes	Yes	No	No
E1-E2	56.64	-	62.1	46.9	No	Yes	No	Yes
N1	60.77	-	63.7	-	No	No	Yes	Yes
N2	54.5	51.7	59.9	58.2	Yes	Yes	No	Yes
N1-N2	56.3	-	72.3	-	No	No	Yes	Yes
E2-E3	33, 45		49, 52	48	Yes	Yes	No	No

Table 2. Comparison of melting temperatures (Tm) and refolding abilities of cadherin domains.



Tm1 56.6 °C

Tm1 62.1 °C
Tm2 58.5 °C

Tm1 46.93 °C

Figure 5.21. DSC profiles of the E1-E2 domain with and without calcium. The refolding ability was present, but aggregation was also seen at a higher temperature.

We next became interested in finding out the effect of the linker regions and how the neighbouring domains play roles in determining thermal stability as well as binding of calcium. We performed DSC experiments with the fusion domains E1-E2, N1-N2 and E2-E3 (only these three fusion domains were expressed and purified in soluble form). E1-E2 showed a single transition indicative of co-operative unfolding and no sign of refolding behaviour. The presence of calcium resulted in increased Tm but refolding ability was compromised and also induced protein aggregation at higher temperatures.

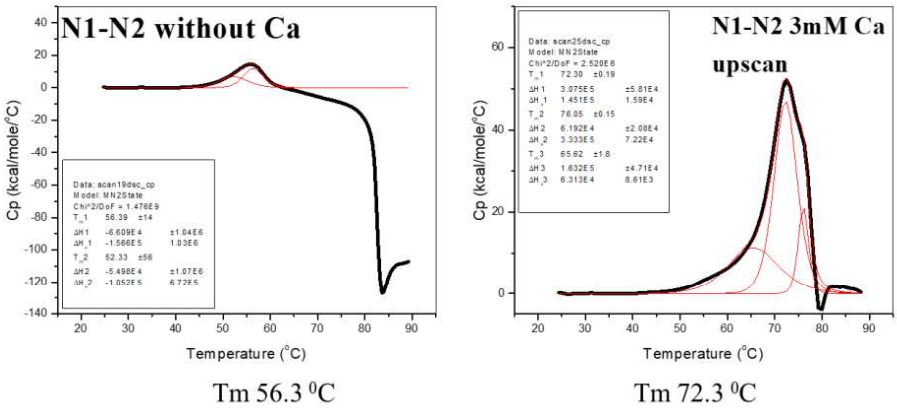


Figure 5.22. DSC profiles of the N1-N2 domain with and without calcium. The refolding ability was diminished when the linkers are present and became prone to aggregation.

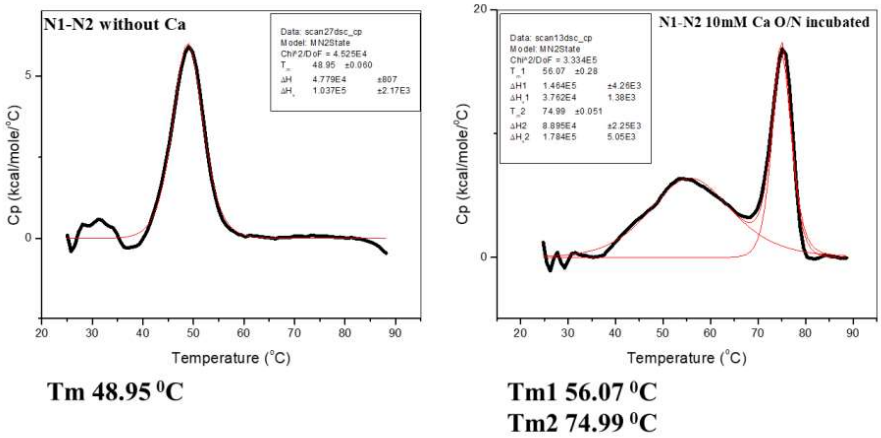


Figure 5.23. DSC profiles of the N1-N2 domain with and without calcium. In presence 10 mM calcium both the domains underwent unfolding at different temperatures.

The N1-N2 fusion domain also showed similar cooperative unfolding behaviour at 56 ºC. With calcium, the Tm increased to 72 ºC. The data suggest that the calcium has a significant role in the stability of N1-N2 because, in this case, the Tm increased by 16 ºC and this increase was also confirmed by CD experiments. A considerable amount of aggregation was seen at higher temperatures. Overall, when DSC was done with fusion domains, the thermal stability was seen to have decreased. This could be the effect of the linker regions, as we know that the calcium binding sites are present at the liker regions and these sites are dominated by aspartate and glutamate amino acid residues. These amino acids can increase negative charge repulsion which could result in a decrease of Tm. The fusion domains did not show any significant ability to refold. That could be because of linker regions and also because calcium binding caused them to become stiff. The indication would be that refolding requires the absence of calcium.

5.2.6 Evaluation of structural stabilities by CD analysis of heated samples

The change in secondary structure of heated samples was also observed by CD spectroscopy. Samples were heated in a cuvette with a path length of mm, and standard parameters were used to collect spectra.

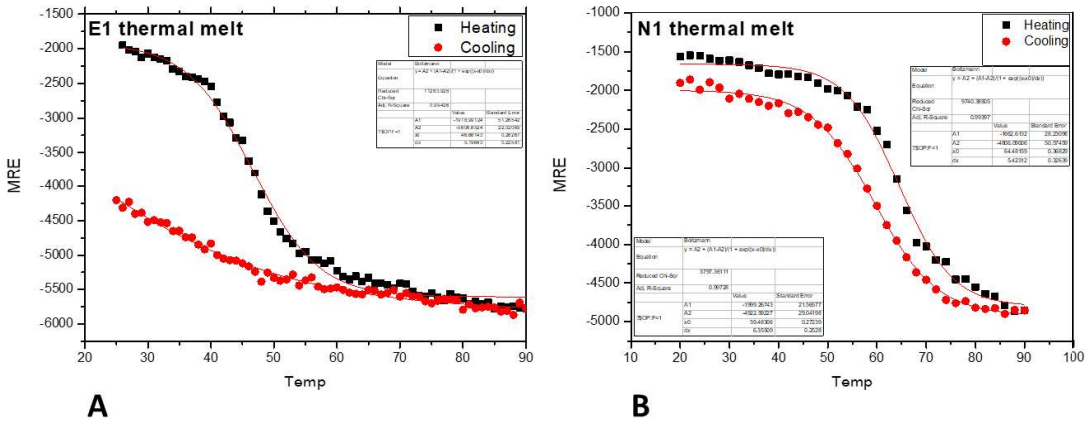


Figure 5.24. CD analysis by thermal denaturation of *Panel A.* CD spectra of E1 collected at 218 nm, black showing upscan and red, downscan, *Panel B.* CD spectra of N1 collected at 218 nm, black showing upscan and red, downscan

Figure 25 depicts the CD spectral data plotted as mean residue ellipticity (MRE) collected an as a function of temperature at 218 nm. E1 and N1 both showed Tm at 45 ºC and 55 ºC respectively, which indicated that N1 is thermally more stable than E1. This corroborates the DSC data. Again, the CD spectra collected during downscan indicate that N1 is able to refold back to its original structure at a physiological temperature which has also been verified by DSC.

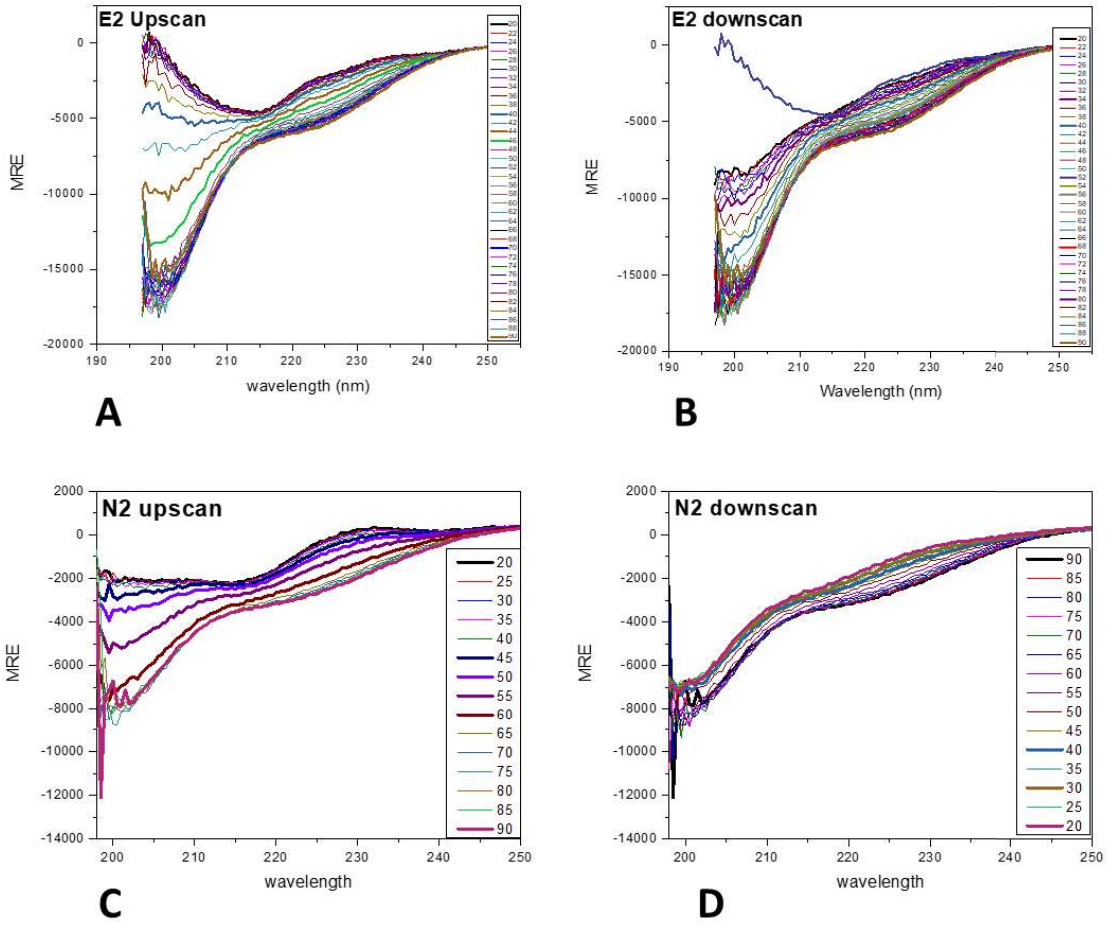


Figure 5.25. CD analysis by thermal denaturation, *Panel A and B.* upscan and downscan profiles by CD spectra of E2, *Panel C and D.* upscan and downscan profiles by CD spectra of N2

The thermal denaturation profile of E2 shows that, upon heating, the unfolding starts at ~40 ºC, while N2 denaturation starts at the higher temperature of ~50 ºC. This suggests that N2 is more stable than E2, which was also confirmed by DSC studies. The downscan CD profile

indicates that E2 has some ability to refold, but the protein was not able to refold completely. N2 also showed evidence of refolding, but the profile was not very clear because of scattering at lower wavelengths.

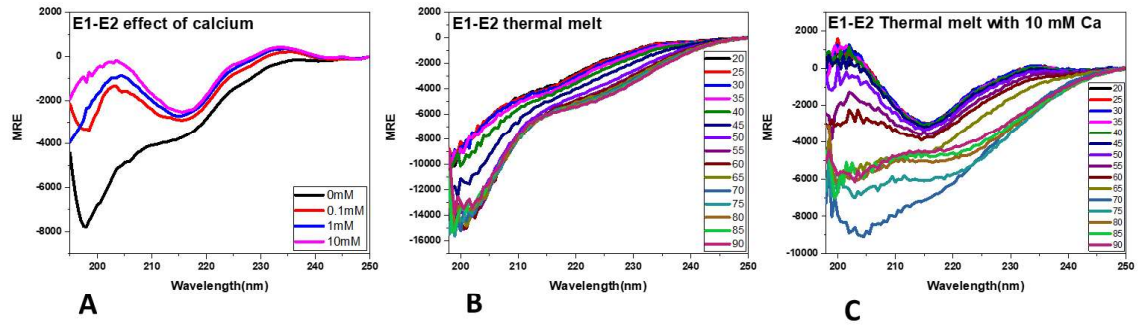


Figure 5.26. CD spectra of heated E1-E2 samples, *Panel A.* effect of calcium incubation, *Panel B.* CD spectra of heated samples in the absence of calcium, *Panel C.* CD spectra of heated samples in the presence of 3 mM calcium.

E1-E2 showed a significant change in secondary structure when incubated with calcium chloride. In terms of thermal stability, without calcium, the protein showed a transition at ~40 ºC while in the presence of calcium it showed a transition at ~50 ºC. This indicates that calcium causes an increase in structure as well as in stability. Although, CD spectra show the more negative value of ellipticity and, in this case, the increase in random coil contribution makes the spectral shape more negative (rather shifting towards zero).

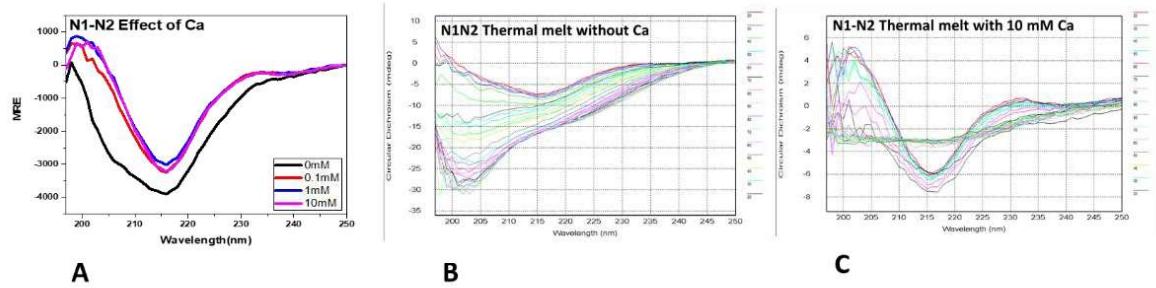


Figure 5.27. CD spectra of heated N1-N2 samples, *Panel A*. effect of calcium incubation, *Panel B*. CD spectra of heated samples in the absence of calcium, *Panel C*. CD spectra of heated samples in the presence of 3 mM calcium.

Similar was the case with N1-N2, which showed an enhanced β-sheet content in CD structure upon addition of calcium. When the sample was heated, the denaturation started at ~40 ºC, but in the presence of calcium, the same occurred at ~55 ºC. The spectral shape was approaching towards zero at 216-219 nm when denaturation occurred at high temperature, i.e., in other domains it has been seen that when unfolding starts (thermal as well as chemical), the CD spectra at 216-219 nm was becoming more negative which is unusual but in case of N1-N2 with calcium it became like a straight line with no structure.

5.2.7 Evaluation of structural stability by circular dichroism analysis of chemical denaturation

Chemical denaturation studies were performed to measure stability in the presence of increasing concentrations of denaturants like urea and guanidium chloride. Circular dichroism was used to monitor loss of structure under equilibrium conditions after overnight incubation at room temperature. By making detailed comparisons, we wished to see how these domains relate to each other, especially in terms of their immediately neighbouring domains, in terms of stability to chemical denaturation. Figure 29 below shows CD spectra collected for different concentrations of urea (0.1 M to 8 M urea) and guanidium hydrochloride (0.1 M to 6 M GdmCl). The spectra in black show the CD spectra collected without any denaturant present, and other spectra shown in various colours are for different concentrations of urea or GdmCl. From the figure, it is evident that, with urea, the domain E1 shows denaturation even at a very low concentration (C_m of 1.8 M) of denaturant whereas N1 shows denaturation at higher concentrations of denaturant (C_m of 3.6 M). On the other hand, with GdmCl, both E1 and N1 are seen to be denatured by very low concentrations of 0.009 M, and 0.67 M, respectively. This reveals that N1 is significantly more stable than E1 in respect of chemical denaturation which, it will be recalled, was also the outcome seen with thermal denaturation experiments.

It is well established that urea destabilizes hydrogen bonds in proteins as well as hydrogen bonds in the bulk water, whereas GdmCl, being ionic, destabilizes electrostatic interactions in addition to destabilizing hydrogen bonds. Therefore, with GdmCl, the concentration required to unfold the half of the protein population (C_m) is expected to be lower than that applicable to urea. It is commonly assumed as a 'rule of thumb' that the concentration of GdmCl required to cause unfolding of half of a protein population is approximately half of the concentration of urea required to achieve the same effect. However, in the case of E1, this rule of thumb no longer holds true. We find that 1.8 M urea is required to unfold E1, whereas only 0.009 M GdmCl is sufficient to unfold the same protein domain (approximately 200 times lower than the urea concentration required). This unusual finding suggests that the structure of E1 is so significantly stabilized by electrostatic interactions (in comparison with all other non-covalent interactions including hydrogen bonding, hydrophobic interactions and other van der Waals interactions) that this causes E1 to unfold at very low concentrations of GdmCl. This excess dependence on electrostatic interactions is also quite obviously then the reason for the very high anomalous mobility seen during SDS-PAGE electrophoresis of E1, which is not seen with N1. However, in the case of N1 too, the C_m with GdmCl was only 0.67 M in comparison with the C_m of 3.6 M which was seen with urea (approximately five times lower, rather than half the concentration). This indicates that there is a preponderance of electrostatic interactions in the case of N1 also; only not to the same extent as seen in the case of E1, because of which there is no electrophoretic anomaly seen with N1.

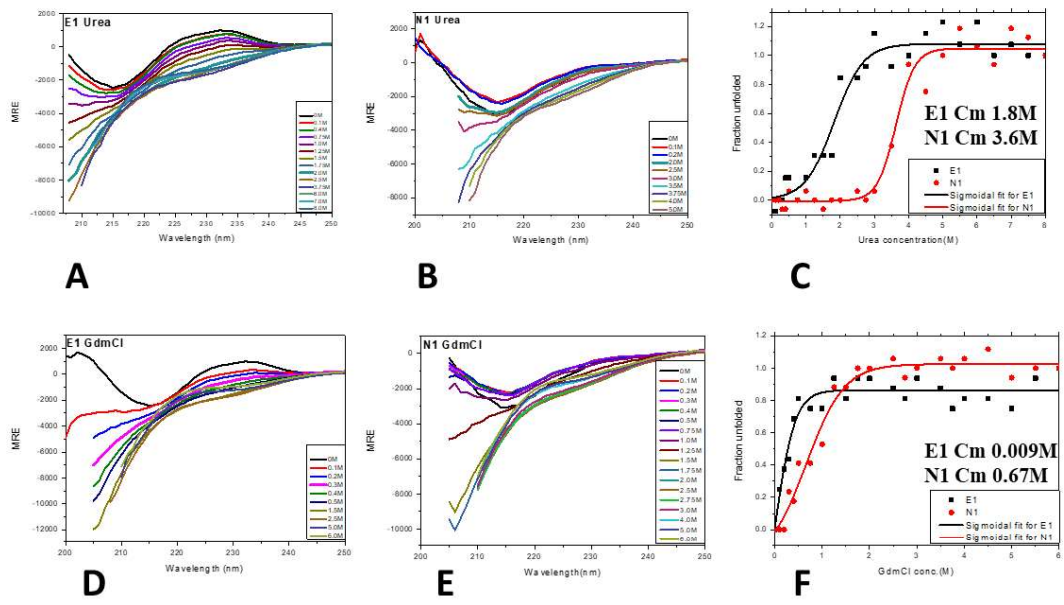


Figure 5.28. Chemical denaturation monitored by CD *Panel A.* E1 with urea, *Panel B.* N1 with urea, *Panel C.* Sigmoidal plot for E1 and N1 to compare their C_m, *Panel D.* E1 with GdmCl, *Panel E.* N1 with GdmCl, *Panel F.* Sigmoidal plot for E1 and N1 to compare their C_m.

In all cases of chemical and thermal denaturation monitored by CD, we observed a peculiar shape of the CD spectrum while effecting unfolding by denaturants or heat. When any protein undergoes unfolding, the negative mean residue ellipticity (MRE) value generally decreases and approaches towards zero as the protein becomes fully unfolded at most wavelengths above 210 nm. Below 210 nm, the negative MRE value generally increases further owing to the increase in the content of random coil content. In case of the cadherin domains, at later stages of unfolding, the spectral shape at 205 nm become more negative in keeping with the increase in random coil content. Interestingly, the negative MRE in the range of 216-219 nm also becomes more negative, instead of moving toward lower values of negative MRE. Therefore, to calculate C_m, the fitting in the sigmoidal equation could not be done properly.

5.2.8 Evaluation of structural stability by fluorescence spectroscopic analysis of chemical denaturation

As discussed earlier, from CD experiments sigmoidal fitting could not be done properly. Therefore, we additionally performed fluorescence spectroscopic analysis of thermal denaturation of the domains, in particular, the fusion domains, E1-E2 and N1-N2 by exciting them with light of 275 nm wavelength and recording emission spectra from the domains (principally owing to the fluorescence of tryptophan) at different temperatures, to record the change in the wavelength of maximal emission owing to the unfolding and the exposure of the tryptophan residues. The wavelengths of maximal emission were then plotted as a function of temperature and C_m values were determined from such plots. Figure 30 below shows that the C_m value of E1-E2 with urea is 4.48 M, and 0.75 M with GdmCl. Once again it is seen that unlike the usual ratio of C_m between urea and GdmCl of 2:1, where the ratio is 6:1. This shows that the three-dimensional structure of E1-E2 is also stabilized mainly by electrostatic interactions, as already seen with the domain E1 through CD studies.

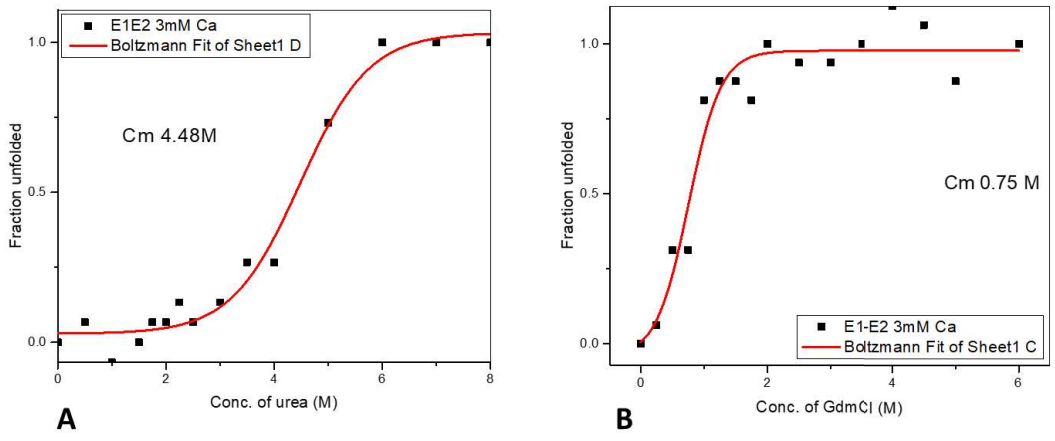


Figure 5.29. Chemical denaturation analysis of E1-E2 by fluorescence spectroscopy *Panel A*. Boltzmann fit to calculate the C_m in the presence of urea; *Panel B*. Boltzmann fit to estimate in the presence of GdmCl.

Similarly, as seen in Figure 31, N1-N2 showed a C_m of 5 M with urea, and of 1.5 M with GdmCl, i.e., a ratio of 3:1 in comparison with the ratio of 5:1 seen with N1 alone, indicating that N2 is stabilized by electrostatic interactions to an even lower extent than N1. Also, it may be noted that N1-N2 is more stable than E1-E2 to chemical denaturation by both urea and GdmCl.

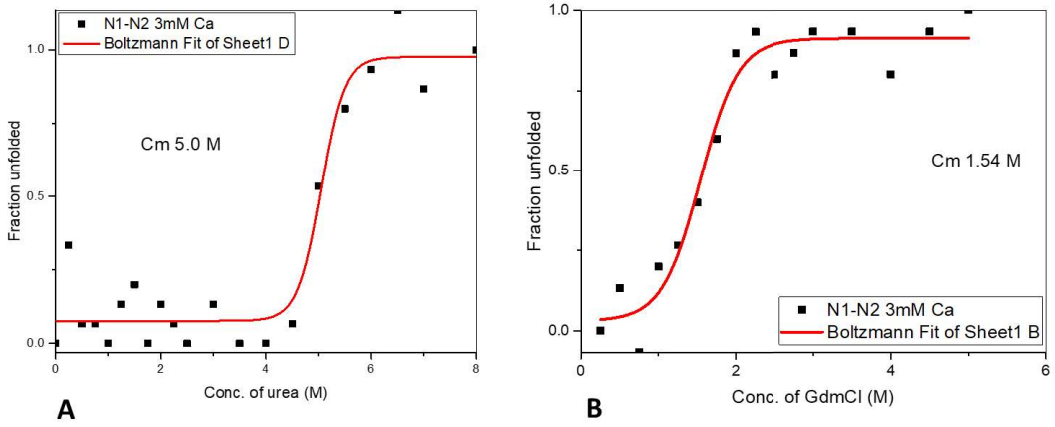


Figure 5.30. Chemical denaturation analysis of N1-N2 by fluorescence spectroscopy *Panel A*. Boltzmann fit to calculate the C_m in the presence of urea; *Panel B*. Boltzmann fit to estimate in the presence of GdmCl.

5.3 Evaluation of structural content, stability and calcium binding of the E5 domain

Next, the biophysical characterization of E5 was done. This domain is directly anchored to the membrane. Therefore, its properties can probably be used as relevant information in understanding how the cadherin behaves at its base. The E5 domain has a vital role during T-cell adhesion to epithelial cells via the αEβ7 integrins present on the surfaces of T cells, where there appears to be a critical role of the fifth domain of E-cadherin for heterophilic adhesions with αEβ7, but not for homophilic adhesion. Thus, the EC5 protein domain may be a good candidate for blocking T-cell adhesion to epithelial cells during inflammation and is therefore worthy of study. Also, we wanted to observe the effect of the two disulfide bonds known to be present in EC5, to see how they affect its biophysical properties.

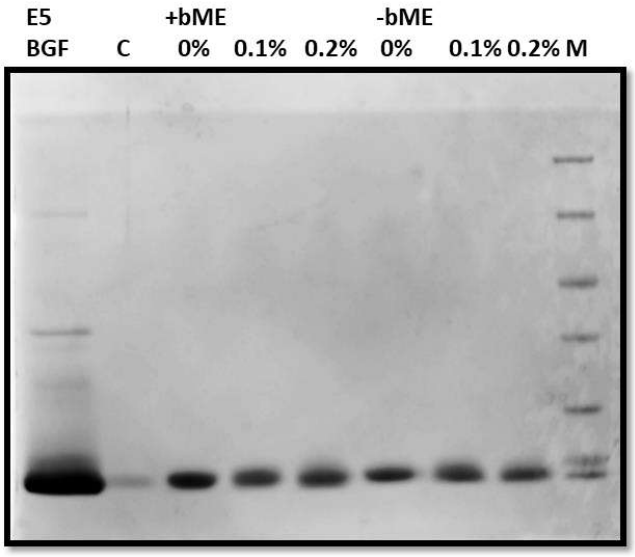


Figure 5.31. SDS-PAGE profile showing effect of β-mercaptoethanol on E5 oligomer formation

The SDS-PAGE in Figure 32 shows crosslinking experiments done with E5 in the presence of the reducing agent, β-mercaptoethanol. The idea behind this experiment was to observe whether disulphide bonds play any role in forming higher oligomers, through misfolding and intermolecular formation of disulfide-based crosslinks using the participating cysteine residues, but we did not observe any oligomeric species in this experiment.

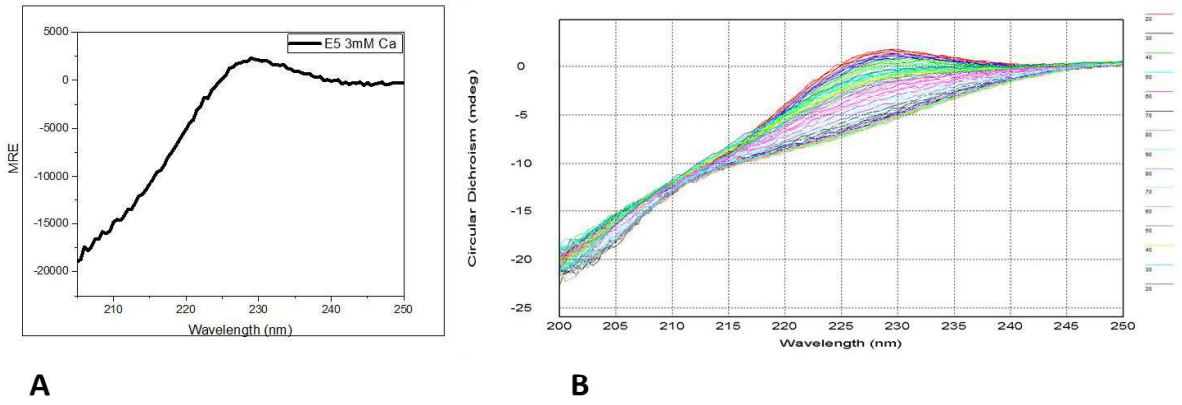


Figure 5.32. CD spectra of E5. *Panel A.* CD spectra in the presence of 3 mM calcium, *Panel B.* thermal melting of E5 analyzed by CD

The CD spectrum of domain E5 in panel A of Figure 33 shows a positive peak at 230 nm, which could be a contribution from aromatic residues and the disulphide bonds present in E5. Interestingly, there was no sign of β-sheet structure in the spectrum. However, the protein's native conformation is dominated by β-sheets and few random coils. This suggests that the protein domain is not natively folded, but the high solubility of the protein in the cytoplasm also indicates that the protein domain is not aggregated. This suggests that domain E5 has a tendency towards formation of 'intrinsically disordered' structure when it is by itself and that it only forms beta sheet-based structure when it is the presence of its neighbouring domains. Panel B of Figure 33 shows thermal melting which again gradually shows a movement towards more negative MREs at a higher temperature like the other cadherin domains, in the regions of the spectrum in which the negative MRE should actually move towards zero, indicative of conversion of some disordered structure to ordered structure. At the same time, at wavelengths below 210 nm, there was an increase in the negative MRE owing to random coil content, possibly owing to the conversion of the ordered structure into the disordered structure in other parts of the protein domain.

5.3.1 Evaluation of structural stability of E5 by fluorescence spectroscopy

To determine the T_m of the E5 domain, since the CD data could not be fitted into a sigmoidal plot because the spectra showed increased MRE values at 215-218 nm with increasing temperature, once again we resorted to the use of fluorescence spectroscopy. The fluorescence emission spectra recorded during the thermal melting experiment was plotted against temperature. As seen in Figure 34, The T_m obtained was 61.6 ºC which is slightly higher than that of other domains. Probably, disulfide bonds are responsible for this apparently higher stability which has also been described by other researchers (Zheng, Laurence et al. 2009).

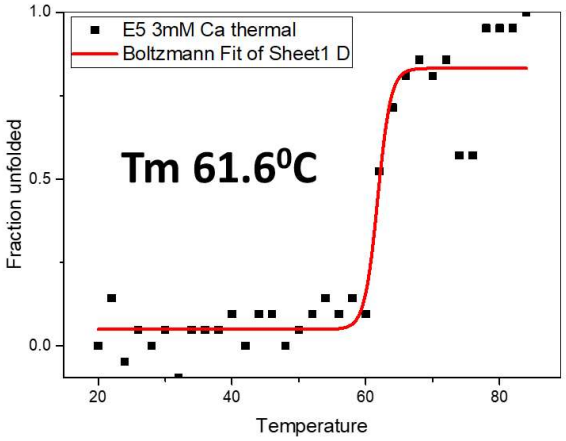


Figure 5.33. Boltzmann fit obtained by fluorescence spectroscopy to calculate Tm of E5 domain construct

5.3.2 Evaluation of structural stability of E5 by differential scanning calorimetry analysis

Thermal melting was also done by DSC for domain E5, to determine the thermodynamic parameters, through two cyles of unfolding (upscan) and refolding (downscan). Interestingly, as seen in Figure 35, the T_m obtained by DSC in the first upscan was 64.5 ºC which is very close to the T_m obtained by CD. The melting showed a cooperative unfolding in which a single transition was observed for melting of the E5 domain, as can be seen in Figure 35. E5 domains also showed refolding ability after denaturation, like E2, but in this case in the very next downscan cycle, the T_m reduced to 61.9 ºC from 64.5 ºC, indicating that all aspects of the domain's structure had not undergone refolding after the first cycle, as was the case with E2, but not the case with N2 (in which unfolding and refolding produced the same T_m through 22 cycles of unfolding and refolding, with no hysteresis). This indicates that refolding ability is also possessed by E5 domain, which is comparable to that seen with E2, but not as good as that seen with N2.

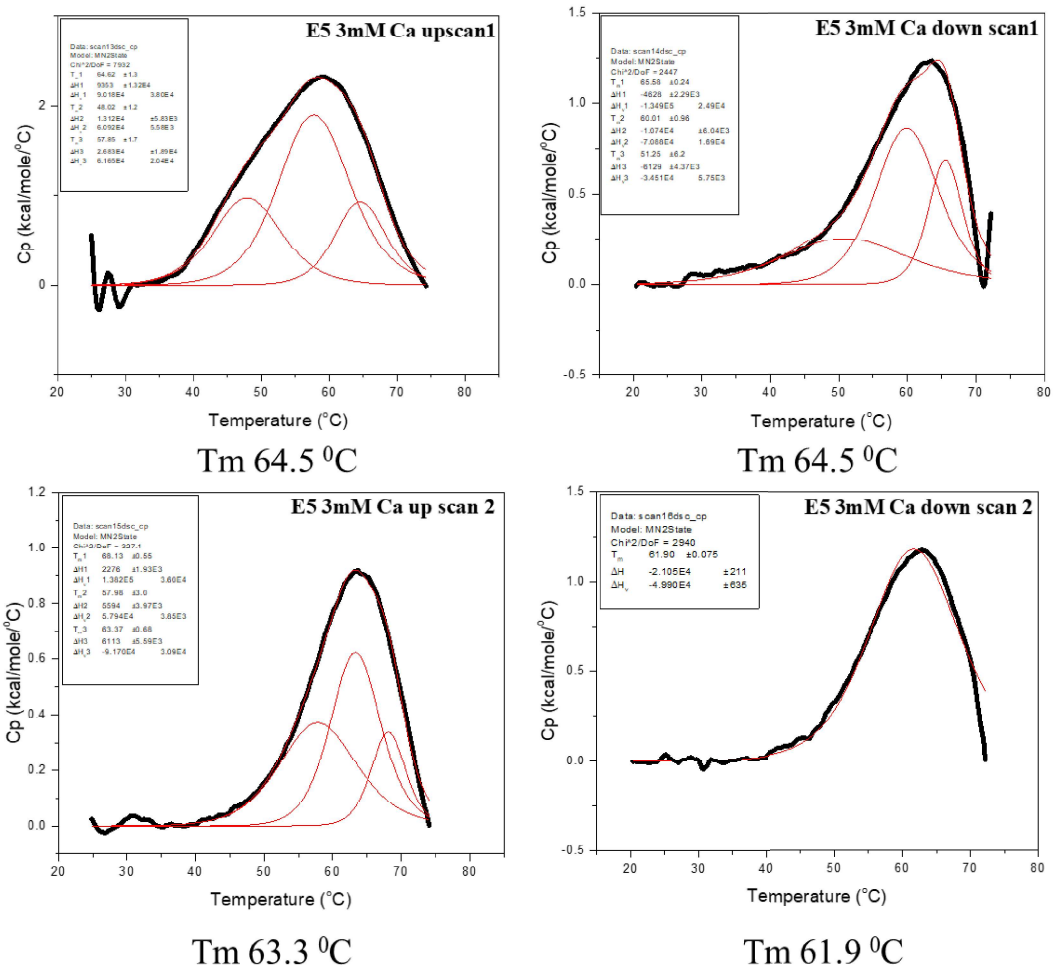


Figure 5.34. DSC profile of E5 in the presence of 3 mM calcium. The data were recorded between 20 and 80 ºC.

5.3.3 Evaluation of structural stability of E5 by chemical denaturation analysis

Chemical denaturation studies were performed for E5 with urea and GdmCl. As can be seen in Figure 36, the C_m values obtained with urea, and GdmCl, respectively, were 2.5 M and 1.9 M. This clearly indicates that the folding of E5 is stabilized mainly by hydrogen bonding and hydrophobic interactions. The involvement of electrostatic interactions is considerably low, and this conclusion is also supported by the fact that the NC: PC ratio of E5 is only 1.5

(please refer to the electrophoretic anomaly experiments in which domain E5 did not show any anomalous mobility during denaturing and reducing electrophoresis, or AM-DRE).

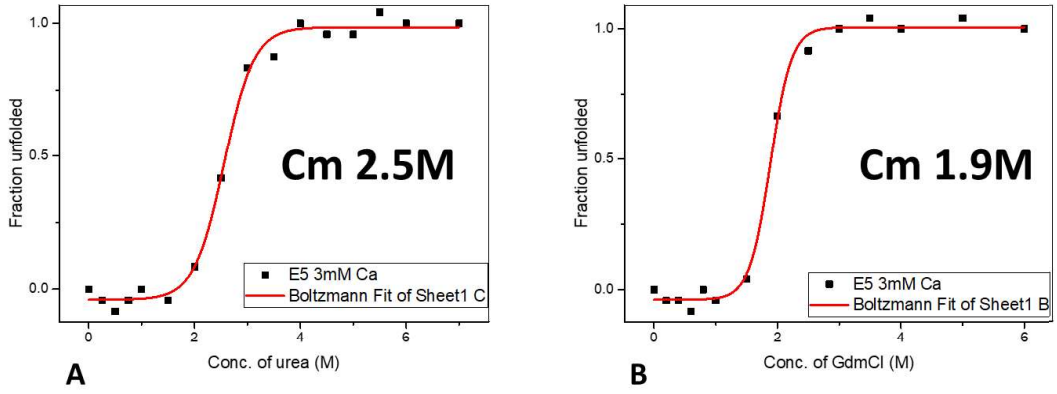


Figure 5.35. Boltzmann (sigmoidal) fit for chemical denaturation measured by fluorescence spectroscopy.

5.4 Evaluation of structural contents, stability and calcium binding of the E2-E3 fusion domain:

5.4.1 Improving solubility of an insoluble domain in the presence of a soluble domain

The E2-E3 fusion domain is an interesting example of the effects of fusing a domain that is soluble with a domain that is insoluble when the two domains are produced in the *E.coli* cytoplasm. It will be recalled that after successful purification and characterization of the individual domains of E- and N-cadherins, we had observed that a found few domains are highly soluble, and a few are insoluble or poorly soluble. E2 was expressed in soluble form, but E3 turned out to be a mixture of soluble aggregates and insoluble aggregates. Surprisingly, the fusion of the two domains to create the fusion domain, E2-E3, resulted in the formation of soluble and well-folded protein. This could be an example of how an equilibrium between soluble and insoluble aggregated populations of a domain like E3 can completely shift towards soluble form when it is fusion with a soluble domain like E2. Later, we also tried fusion with other soluble domains with insoluble domains.

5.4.2 Evaluation of the effects of calcium binding upon secondary structure of E2-E3 by circular dichroism

It will be recalled that both E2 and E3 did not show any change in secondary structure upon addition of calcium. However, as can be seen in Figure 37 panel A, when we checked the effect of calcium upon the fusion domain, E2-E3, there was a clear shift of the random coil into a β-sheet which means that in the presence of the neighbouring domain, the linker region responds to calcium, and this results in enhanced β-sheet formation. The inference is that both the domains are folded, and the resulting spectrum shows a pure β-sheet structure, native of cadherin EC domains, and this is either because of the effect of the linker region which binds to calcium which precedes the E3 domain, or because of the presence of the E2 domain located further upstream. This question can only be answered by creating a construct consisting of E3 alone, with the linker preceding it and not the linker succeeding it.

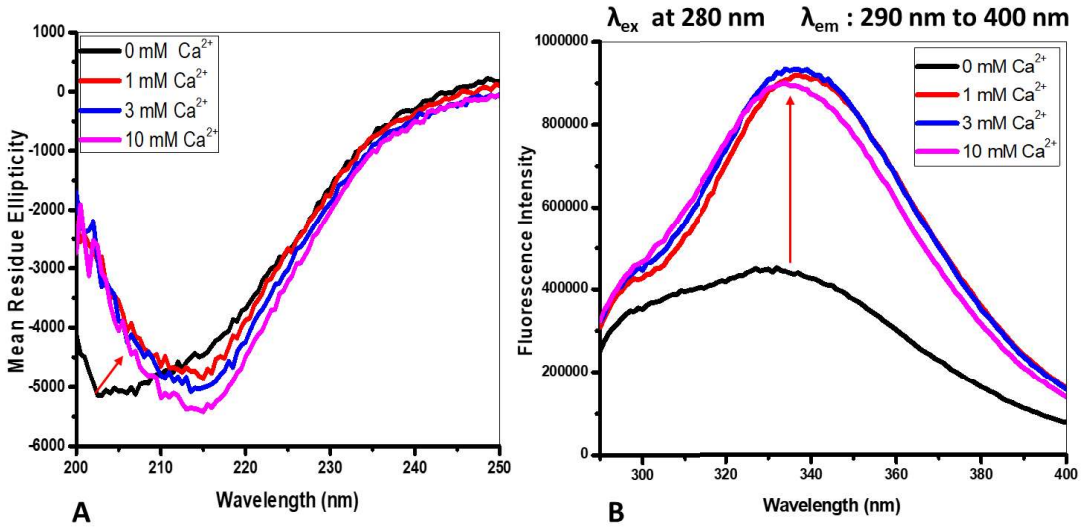


Figure 5.36. E2-E3 effect of calcium by *Panel A*. CD spectra in the presence of different concentration of calcium showing an enhanced β-sheet formation in the presence of calcium, *Panel B*. Fluorescence emission spectra showing FRET between E2 and E3 domain in the presence of calcium.

5.4.3 Evaluation of the effects of calcium binding upon tertiary structure (fluorescence emission)

The same samples used for CD data collection were also used to record fluorescence emission spectra. In the panel B of Figure 37, the emission spectra of tryptophan have been plotted against fluorescence intensity (FI). E2 contains no tryptophan while E3 contains only one tryptophan residue. This inherent property was used to study the effect of the addition of calcium. The black curve represents an emission spectrum collected with excitation at 275 nm in the absence of calcium, and this emission consists of two distinct peaks at 307 nm and 335 nm, corresponding to the emission from tyrosine in the E2 and E3 domains, and tryptophans in the E3 domain. When calcium is added, there is a sudden increase (almost two-fold) in FI, owing to fluorescence resonance energy transfer (FRET). Is the FRET phenomenon occurring between the two domains in an intra-molecular fashion or intramolecular fashion? It seems that the excitation energy from tyrosines (after excitation at 275 nm) being transferred to the single tryptophan present only in the E3 domain through radiationless energy transfer and that this tryptophan is, in turn, emitting energy resulting in an increase in FI up to two-fold. This event is correlated with the cadherin dimerization. To explain this in a better way, let us think that in the presence of calcium a dimer of E2-E3 forms and that E2 of one molecule is interacting with E3 of the other molecule and *vice-versa* because the linkers present between the domains are too short to make a loop so that the two domains can come together and interact in intramolecular fashion. Therefore, possibly two molecules are being flipped, and energy transfer is taking place which results in increased FI. Is this the case, or is it that in the dimer of E2-E3, the tyrosines of the E2 are closer to the tryptophans of the E3 because of the change in the structure of the intervening linker (through calcium binding)? We do not know.

5.4.4 Evaluation of quaternary structure by gel filtration chromatography

The E2-E3 protein domain fusion was chromatographed on a pre-equilibrated Superdex-200 gel filtration column in a buffer of 50 mM Tris, pH 7.5, with and without calcium. Both the profiles are overlaid and shown below for comparison, in Figure 38. It is evident that calcium

causes a drastic change in the quaternary structure which suggests that, in the absence of calcium, largely the population remains as soluble aggregates but when calcium is added a new population arises at an elution volume of 16.5 ml, which corresponds to a molecular weight of ~35 kDa. Like other experiments, this also confirms that calcium causes the soluble aggregates to become folded and leads to dimer formation.

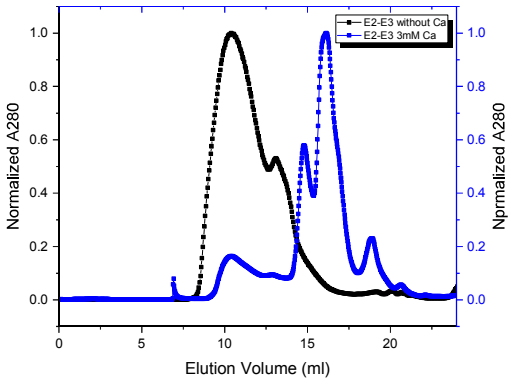
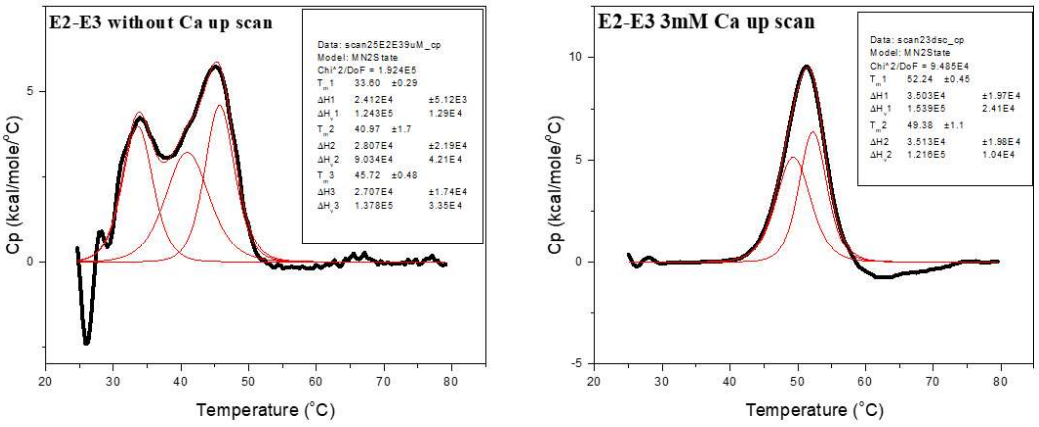


Figure 5.37. Gel filtration profile of E2-E3 showing the effect of calcium on oligomerization

5.4.5 Evaluation of structural stability by differential scanning calorimetry analysis

The DSC data suggests that in the absence of calcium, the unfolding of the two domains in the fusion of E2-E3 occurs at two different temperatures. This is an example of a typical non-cooperative unfolding where presumably the 33 ºC peak corresponds to the unfolding of E3 and the unfolding of E2 corresponds to the 45 ºC peak. The T_m of E2 is in the same range as reported in the DSC result obtained from the E2 domain produced without any fusion. In the presence of calcium, there is a drastic change which results in a cooperative unfolding at a temperature close to 50 ºC, suggesting that the two domain now functions like one unit, although a few more experiments need to be done, to prove that there is a conformational change happening upon addition of calcium.



Tm 1 33.6 ^{0}C
Tm 2 40.9 ^{0}C
Tm 3 45.7 ^{0}C

Tm 1 52.2 ^{0}C
Tm 2 49.3 ^{0}C

Figure 5.38. comparison of DSC profiles of E2-E3 with and without calcium

Chapter 6: Conclusion

Cell adhesion is an essential process in which a variety of cell types are involved. For example, epithelial cells require tissue integrity while neuronal cells do not need the same level of tissue integrity. Several cell adhesion molecules (CAMs) are known to play various roles in the process of cell adhesion, and out of all these CAMs, the cadherins play a vital role. In this thesis, we have described studies of two major types of classical cadherins, known as epithelial (E) and Neuronal (N) cadherins. Sequence and structural alignments involving the domains of human classical cadherins (described in Chapter 3) revealed that the domains are all quite similar and that E- and N-cadherin too are quite similar, despite being reported to perform different functions, even when they are present on the same cell. E-cadherin is involved in epithelial integrity while N-cadherin can regulate cell-cell adhesions during invasion (Tiwari et al. 2012; Shih & Yamada 2012), presumably because quick 'make-and-break' types of contacts are required during tissue invasion, similar to the quick 'make-and-break' types of contacts engaged in by neuronal cells in neural tissue. The fact that the cadherins perform similar functions with dissimilarities, and possess similar domains (also displaying dissimilarities of presentation and function) intrigued us, and caused us to explore possible reasons underlying this behaviour from a biophysical-chemical standpoint supported by protein engineering techniques. We report extensive biophysical and biochemical studies of the extracellular domains of E-and N- cadherins and discuss the associated physiological implications.

It is well understood by various research groups that during the process of cell-cell interaction, cadherins undergo dimerization through their extracellular domains in the intercellular space. The transmembrane domains help in anchoring, and the cytoplasmic domains perform the signalling function via catenins and eventually through the microtubules. The fact that only the first extracellular domain appears to be involved in the dimerization process and that the other four domains do not have any specifically assigned roles, caused us to wonder about why these domains exist since nature does not create anything without reason. If these domains perform only the function of removing the first extracellular domain away from the cell surface, it would be possible for such a trivial function to be performed by many types of proteins, and domains, with no necessity for

conservation of sequence and structure over long evolutionary timescales. We decided to study the properties of the domains, as well as domain functions, partly by exploring their behaviour under different physico-chemical situations (e.g., in the presence or absence of calcium, or by examining their stability to physical, thermal or chemical denaturation).

The bioinformatics analyses revealed that there are significant similarities (>57 %) between the first two EC domains (EC1 and EC2) in case of both E- and N-cadherin's EC domains, which indicates that they might have evolved together. Structural alignment data showed that the two domains have similar folds and are superimposable. The similarities amongst the remaining three domains are considerably lower and least in case of EC5, and the fifth domain could have a very different role from the other domains in both E- and N-cadherin. We concluded that there could be some crucial amino acid side chains or that the tertiary/quaternary structures of all the domains are different from each other to different extents. To explore this further, we cloned, expressed and purified all the EC domains of E- and N-cadherins in all possible combinations of individual domain constructs and domain-fusion constructs, to study their behaviour. The individual domains were investigated first.

The protein purification of the domains was done under non-denaturing and denaturing conditions. While working on the individual domains, we accidentally came across an important observation which we explored in great detail; The SDS-PAGE profile of the individual domain constructs showed an unusual behaviour which we termed as anomalous behaviour during denaturing and reducing electrophoresis (AM-DRE). Briefly, the apparent molecular weight of the domains was not the same as their calculated molecular weight, i.e., the electrophoretic mobility of the constructs was not according to their molecular weights. To understand this, we confirmed the identities of the constructs as well as the integrity of the polypeptide chain with the help of mass spectrometry. Then we looked into the amino acid sequence and found that there is an abundance of high negative charge which appears to be responsible for this anomalous mobility by causing insufficient binding of SDS molecules to the peptide chain during sample preparation for SDS-PAGE. We concluded that all constructs display AM-DRE if the ratio of negative charge to positive charge (NC: PC) is greater than 1.50, and when the ratio is smaller, the proteins show normal mobility during denaturing and reducing electrophoresis (NM-DRE).

Protein expression and purification data showed that when expressed in *E.coli* E1, E2, E5, N1 and N2 are soluble in the cytoplasm, but other domains showed aggregation and formed inclusion bodies. Therefore, our focus was mainly on the soluble constructs and the attempt was also made to refold the domains which were not soluble by using various refolding techniques. Structural analysis with CD revealed that the secondary structure of E1, E2, N1 and N2 is dominated largely by β-sheets and there was no effect of calcium on their secondary structure. We explored the possibility of fusing the neighbouring domains in order to check whether the folding of the EC domains occurs independently or other domains play any role. E1-E2 and N1-N2 fusion domains showed a very well formed β-sheet structure, and in the presence of calcium, the secondary structural content was enhanced in terms of sheet content. It is evident from the Trp emission spectra that domains E1, E2, N1 and N2 are well-folded by themselves, and their tryptophan residues are buried within their hydrophobic cores. E2 showed a typical Tyr emission spectrum with a peak at 307 nm; this domain is devoid of any Trp residue. The quaternary structural analyses of the individual domains by size exclusion chromatography showed that E3, E4, N3, N4 and N5 form soluble aggregates and are eluted at void volume during gel filtration, suggestive of the formation of soluble multimers. E1, E2, N1 and N2 formed tetramers while E5 formed dimers. The crosslinking data suggested that there is an overestimation of oligomeric status as these domains form a 'cigar or rod' like shape and because of that the hydrodynamic radius becomes increased. A similar case was also seen with the fusion constructs. To overcome this problem, we performed analytical ultracentrifugation (AUC) experiments which indicated that E1-E2 and N1-N2 formed dimers only in the presence of calcium.

A comparison of thermodynamic stability and melting temperature (T_m) data obtained by using DSC suggested that the individual domains have a T_m in the range of 45-60 ºC and presence of calcium there was an increase of T_m by at least 5 ºC in most domains. N2 showed an interesting refolding behaviour without any considerable change in enthalpy and T_m which could correlate with a fundamental property for N-cadherins, i.e., the need to quickly associate and dissociate from other cells, with mechanical forces presumably causing denaturation that is required to be reversed rapidly. Thus this refolding behaviour of the N2 domain might have physiological implications, as neuronal cell very often form and break connections in order to store memories where because of this refolding ability the same copy

of molecules can be utilized several times without replenishment of molecules at the cell surface. Fusion domains showed lower thermal stabilities which might be because of the presence of electrostatic repulsions owing to the presence of the linker regions.

Chemical denaturation studies showed that the secondary structures of the EC domains were stabilized mainly by electrostatic interactions, as the effect of Gdm.HCl was observed to be inordinately greater (up to 200 times greater, for some domains) than urea. The results obtained by CD showed that with an increase in denaturation, the spectral shape is gradually increasing towards more negative ellipticity which is probably due to increase in random coil content, causing shifting of the spectral shape towards more negative values of mean residue ellipticity. The comparison of denaturation studies revealed that the EC domains of N-cadherin were more stable than those of the E-cadherins. Based on denaturation studies we hypothesize that during the event separation at the cell surface, cadherin EC domains function as 'springs' where EC1, i.e., either E1 or N1 (being the most stable) remains folded while other domains (which are less stable than EC1) unfold and refold in response to mechanical stresses. The CD spectra of E5 showed a negative peak at 195 nm, and a peculiar positive peak at 228 nm, which probably owes to disulphide bonding because E1 also showed this positive peak along with β-sheet content and there is one disulphide bond in E1. DSC studies of E5 showed it is the most stable domain both in terms of thermal as well as chemical stability and also was able to refold back. Chemical denaturation data indicated that E5 is quite stable as the C_m in the presence of Gdm.HCl is 1.9 M, which is the highest in comparison to other domains. This suggests that electrostatic interactions are least responsible for secondary and tertiary structural stability in E5, and possibly also N5, amongst all domains.

Another interesting observation came from the fusion construct E2-E3 which was initially created to investigate the refolding ability of E3 in the presence of its neighbouring domain E2, which is a soluble domain, unlike E3 which is insoluble. The E2-E3 construct was expressed in soluble and folded form with an excellent β-sheet structure, and the structure was even enhanced in the presence of calcium. Interestingly, there is a single tryptophan in E3 and E2 contains only tyrosine. The presence of this inherent fluorophore was used to investigate the mode of dimerization. In the presence of calcium, the E2 domain appears to

interact with the E3 domain of another molecule and *vice versa.* This was confirmed by intermolecular Förster resonance energy transfer (FRET) involving the tyrosines of E2 and E3 and the single Trp residue present in E3.

To investigate the interaction of the constructs E1-E2 and N1-N2 on mammalian cell lines, fusion constructs have been made. Since the N-terminal of EC-domains participates in dimerization, therefore, we have fused the fluorescent proteins at the C-terminus of the EC domains. The work with these constructs is in progress.

In conclusion, while we have found many interesting features of the individual domains and domain constructs, suggestive of correlations with the differential functions of the E- and N-cadherins, perhaps the most exciting feature is the clear difference between the EC1 and EC2 domains which are soluble and which dimerize, and the EC3 and EC4 domains which form soluble multimers under many different types of conditions, leaving aside EC5 for the moment (which is once soluble like the EC1 and EC2 domains). Our current view is that the EC3 and EC4 domains help in *cis* interactions between cadherins when the EC1 and EC2 domains are involved in *trans* interactions with cadherins on juxtaposed cells. That is, the cadherins interact 'at the waist'. Below, EC5 which is close to the cell surface remains autonomous and above, EC1 and EC2 also remain autonomous. This could help clusters of cadherins on the cell surface to engage in multitudinous trans interactions at cell-cell interfaces quickly.

Chapter 7: Bibliography

Abedin, M. and N. King (2008). "The premetazoan ancestry of cadherins." Science **319**(5865): 946-948.

Al-Amoudi, A., D. Castaño-Diez, D. P. Devos, R. B. Russell, G. T. Johnson and A. S. Frangakis (2011). "The three-dimensional molecular structure of the desmosomal plaque." Proceedings of the National Academy of Sciences **108**(16): 6480-6485.

Astick, M., K. Tubby, W. M. Mubarak, S. Guthrie and S. R. Price (2014). "Central topography of cranial motor nuclei controlled by differential cadherin expression." Current Biology **24**(21): 2541-2547.

Bennett, M. J., M. P. Schlunegger and D. Eisenberg (1995). "3D domain swapping: a mechanism for oligomer assembly." Protein Science **4**(12): 2455-2468.

Boggon, T. J., J. Murray, S. Chappuis-Flament, E. Wong, B. M. Gumbiner and L. Shapiro (2002). "C-cadherin ectodomain structure and implications for cell adhesion mechanisms." Science **296**(5571): 1308-1313.

Brasch, J., O. J. Harrison, B. Honig and L. Shapiro (2012). "Thinking outside the cell: how cadherins drive adhesion." Trends in cell biology **22**(6): 299-310.

Buxton, R., P. Cowin, W. Franke, D. Garrod, K. Green, I. King, P. Koch, A. Magee, D. Rees and J. Stanley (1993). "Nomenclature of the desmosomal cadherins." The Journal of Cell Biology **121**(3): 481-483.

Carmeliet, P., M.-G. Lampugnani, L. Moons, F. Breviario, V. Compernolle, F. Bono, G. Balconi, R. Spagnuolo, B. Oosthuyse and M. Dewerchin (1999). "Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis." Cell **98**(2): 147-157.

Clevers, H. and R. Nusse (2012). "Wnt/β-catenin signaling and disease." Cell **149**(6): 1192-1205.

Curto, M., B. K. Cole, D. Lallemand, C.-H. Liu and A. I. McClatchey (2007). "Contact-dependent inhibition of EGFR signaling by Nf2/Merlin." J Cell Biol **177**(5): 893-903.

Davis, M. A., R. C. Ireton and A. B. Reynolds (2003). "A core function for p120-catenin in cadherin turnover." J Cell Biol **163**(3): 525-534.

Dejana, E., E. Tournier-Lasserve and B. M. Weinstein (2009). "The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications." Developmental cell **16**(2): 209-221.

Delva, E., D. K. Tucker and A. P. Kowalczyk (2009). "The desmosome." Cold Spring Harbor perspectives in biology **1**(2): a002543.

Gallin, W. J. (1998). "Evolution of the" classical" cadherin family of cell adhesion molecules in vertebrates." Molecular biology and evolution **15**(9): 1099-1107.

Garrod, D. R., A. J. Merritt and Z. Nie (2002). "Desmosomal cadherins." Current opinion in cell biology **14**(5): 537-545.

Gottardi, C. J., E. Wong and B. M. Gumbiner (2001). "E-cadherin suppresses cellular transformation by inhibiting β-catenin signaling in an adhesion-independent manner." The Journal of cell biology **153**(5): 1049-1060.

Green, K. J. and C. L. Simpson (2007). "Desmosomes: new perspectives on a classic." Journal of Investigative Dermatology **127**(11): 2499-2515.

Gumbiner, B. M. (2005). "Regulation of cadherin-mediated adhesion in morphogenesis." Nature reviews Molecular cell biology **6**(8): 622.

Haas, A. R. and R. S. Tuan (1999). "Chondrogenic differentiation of murine C3H10T1/2 multipotential mesenchymal cells: II. Stimulation by bone morphogenetic protein-2 requires modulation of N-cadherin expression and function." Differentiation **64**(2): 77-89.

Halbleib, J. M. and W. J. Nelson (2006). "Cadherins in development: cell adhesion, sorting, and tissue morphogenesis." Genes & development **20**(23): 3199-3214.

Harrison, O. J., F. Bahna, P. S. Katsamba, X. Jin, J. Brasch, J. Vendome, G. Ahlsen, K. J. Carroll, S. R. Price and B. Honig (2010). "Two-step adhesive binding by classical cadherins." Nature structural & molecular biology **17**(3): 348.

Harrison, O. J., X. Jin, S. Hong, F. Bahna, G. Ahlsen, J. Brasch, Y. Wu, J. Vendome, K. Felsovalyi and C. M. Hampton (2011). "The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins." Structure **19**(2): 244-256.

Hatta, K., A. Nose, A. Nagafuchi and M. Takeichi (1988). "Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule: its identity in the cadherin gene family." The Journal of Cell Biology **106**(3): 873-881.

Hatta, K. and M. Takeichi (1986). "Expression of N-cadherin adhesion molecules associated with early morphogenetic events in chick development." Nature **320**(6061): 447-449.

Hatzfeld, M. (2007). "Plakophilins: Multifunctional proteins or just regulators of desmosomal adhesion?" Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1773**(1): 69-77.

Häussinger, D., T. Ahrens, T. Aberle, J. Engel, J. Stetefeld and S. Grzesiek (2004). "Proteolytic E-cadherin activation followed by solution NMR and X-ray crystallography." The EMBO journal **23**(8): 1699-1708.

Hong, S., R. B. Troyanovsky and S. M. Troyanovsky (2011). "Cadherin exits the junction by switching its adhesive bond." The Journal of cell biology **192**(6): 1073-1083.

Hulpiau, P. and F. Van Roy (2009). "Molecular evolution of the cadherin superfamily." The international journal of biochemistry & cell biology **41**(2): 349-369.

Hulpiau, P. and F. Van Roy (2010). "New insights into the evolution of metazoan cadherins." Molecular biology and evolution **28**(1): 647-657.

Hyafil, F., C. Babinet and F. Jacob (1981). "Cell-cell interactions in early embryogenesis: a molecular approach to the role of calcium." Cell **26**(3): 447-454.

Kang, H.-G., J. M. Jenabi, J. Zhang, N. Keshelava, H. Shimada, W. A. May, T. Ng, C. P. Reynolds, T. J. Triche and P. H. Sorensen (2007). "E-cadherin cell-cell adhesion in ewing tumor cells mediates suppression of anoikis through activation of the ErbB4 tyrosine kinase." Cancer research **67**(7): 3094-3105.

Kister, A. E., M. A. Roytberg, C. Chothia, J. M. Vasiliev and I. M. Gelfand (2001). "The sequence determinants of cadherin molecules." Protein Science **10**(9): 1801-1810.

Koch, A., K. Manzur and W. Shan (2004). "Structure-based models of cadherin-mediated cell adhesion: the evolution continues." Cellular and molecular life sciences **61**(15): 1884-1895.

Leckband, D. and A. Prakasam (2006). "Mechanism and dynamics of cadherin adhesion." Annu. Rev. Biomed. Eng. **8**: 259-287.

Luo, M., Z. Li, W. Wang, Y. Zeng, Z. Liu and J. Qiu (2013). "Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression." Cancer letters **333**(2): 213-221.

Mahoney, P. A., U. Weber, P. Onofrechuk, H. Biessmann, P. J. Bryant and C. S. Goodman (1991). "The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily." Cell **67**(5): 853-868.

McCrea, P. D., M. T. Maher and C. J. Gottardi (2015). Nuclear signaling from cadherin adhesion complexes. Current topics in developmental biology, Elsevier. **112:** 129-196.

McCrea, P. D., C. W. Turck and B. Gumbiner (1991). "A homolog of the armadillo protein in Drosophila (plakoglobin) associated with E-cadherin." Science **254**(5036): 1359-1361.

McLachlan, R. W., A. Kraemer, F. M. Helwani, E. M. Kovacs and A. S. Yap (2007). "E-cadherin adhesion activates c-Src signaling at cell–cell contacts." Molecular biology of the cell **18**(8): 3214-3223.

Murray-Stewart, T., P. M. Woster and R. A. Casero (2014). "The re-expression of the epigenetically silenced e-cadherin gene by a polyamine analogue lysine-specific demethylase-1 (LSD1) inhibitor in human acute myeloid leukemia cell lines." Amino acids **46**(3): 585-594.

Nakayama, T., H. Gardner, L. K. Berg and J. L. Christian (1998). "Smad6 functions as an intracellular antagonist of some TGF-β family members during Xenopus embryogenesis." Genes to Cells **3**(6): 387-394.

Nichols, M., N. Townsend, R. Luengo-Fernandez, J. Leal, A. Gray, P. Scarborough and M. Rayner (2012). "European cardiovascular disease statistics 2012."

Nollet, F., P. Kools and F. Van Roy (2000). "Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members." Journal of molecular biology **299**(3): 551-572.

Overduin, M., T. S. Harvey, S. Bagby, K. I. Tong, P. Yau, M. Takeichi and M. Ikura (1995). "Solution structure of the epithelial cadherin domain responsible for selective cell adhesion." Science **267**(5196): 386-389.

Ozawa, M., H. Baribault and R. Kemler (1989). "The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species." The EMBO journal **8**(6): 1711-1717.

Peifer, M., P. D. McCrea, K. J. Green, E. Wieschaus and B. M. Gumbiner (1992). "The vertebrate adhesive junction proteins beta-catenin and plakoglobin and the Drosophila segment polarity gene armadillo form a multigene family with similar properties." The Journal of cell biology **118**(3): 681-691.

Perrais, M., X. Chen, M. Perez-Moreno and B. M. Gumbiner (2007). "E-cadherin homophilic ligation inhibits cell growth and epidermal growth factor receptor signaling independently of other cell interactions." Molecular biology of the cell **18**(6): 2013-2025.

Qian, X., T. Karpova, A. M. Sheppard, J. McNally and D. R. Lowy (2004). "E-cadherin-mediated adhesion inhibits ligand-dependent activation of diverse receptor tyrosine kinases." The EMBO journal **23**(8): 1739-1784.

Rudini, N., A. Felici, C. Giampietro, M. Lampugnani, M. Corada, K. Swirsding, M. Garrè, S. Liebner, M. Letarte and P. Ten Dijke (2008). "VE-cadherin is a critical endothelial regulator of TGF-β signalling." The EMBO journal **27**(7): 993-1004.

Shapiro, L., A. M. Fannon, P. D. Kwong, A. Thompson, M. S. Lehmann, G. Grübel, J.-F. Legrand, J. Als-Nielsen, D. R. Colman and W. A. Hendrickson (1995). "Structural basis of cell-cell adhesion by cadherins." Nature **374**(6520): 327.

Shapiro, L. and W. I. Weis (2009). "Structure and biochemistry of cadherins and catenins." Cold Spring Harbor perspectives in biology: a003053.

Siemens, J., C. Lillo, R. A. Dumont, A. Reynolds, D. S. Williams, P. G. Gillespie and U. Müller (2004). "Cadherin 23 is a component of the tip link in hair-cell stereocilia." Nature **428**(6986): 950.

Simcha, I., C. Kirkpatrick, E. Sadot, M. Shtutman, G. Polevoy, B. Geiger, M. Peifer and A. Ben-Ze'ev (2001). "Cadherin sequences that inhibit β-catenin signaling: a study in yeast and mammalian cells." Molecular Biology of the Cell **12**(4): 1177-1188.

Sivasankar, S. (2013). "Tuning the kinetics of cadherin adhesion." Journal of Investigative Dermatology **133**(10): 2318-2323.

Suyama, K., I. Shapiro, M. Guttman and R. B. Hazan (2002). "A signaling pathway leading to metastasis is controlled by N-cadherin and the FGF receptor." Cancer cell **2**(4): 301-314.

Suzuki, K., Y. Tsunekawa, R. Hernandez-Benitez, J. Wu, J. Zhu, E. J. Kim, F. Hatanaka, M. Yamamoto, T. Araoka and Z. Li (2016). "In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration." Nature **540**(7631): 144.

Tanihara, H., K. Sano, R. L. Heimark, T. St. John and S. Suzuki (1994). "Cloning of five human cadherins clarifies characteristic features of cadherin extracellular domain and provides further evidence for two structurally different types of cadherin." Cell adhesion and communication **2**(1): 15-26.

Tanoue, T. and M. Takeichi (2004). "Mammalian Fat1 cadherin regulates actin dynamics and cell–cell contact." J Cell Biol **165**(4): 517-528.

Thiery, J. P., H. Acloque, R. Y. Huang and M. A. Nieto (2009). "Epithelial-mesenchymal transitions in development and disease." cell **139**(5): 871-890.

Thomason, H. A., A. Scothern, S. McHarg and D. R. Garrod (2010). "Desmosomes: adhesive strength and signalling in health and disease." Biochemical Journal **429**(3): 419-433.

Townes, P. L. and J. Holtfreter (1955). "Directed movements and selective adhesion of embryonic amphibian cells." Journal of experimental zoology **128**(1): 53-120.

Troyanovsky, S. (2012). Adherens junction assembly. Adherens Junctions: from Molecular Mechanisms to Tissue Development and Disease, Springer**:** 89-108.

Uchida, N., S. Sivaraman, N. J. Amoroso, W. R. Wagner, A. Nishiguchi, M. Matsusaki, M. Akashi and J. Nagatomi (2016). "Nanometer-sized extracellular matrix coating on polymer-based scaffold for tissue engineering applications." Journal of Biomedical Materials Research Part A **104**(1): 94-103.

Urushihara, H. and M. Takeichi (1980). "Cell-cell adhesion molecule: identification of a glycoprotein relevant to the Ca2+-independent aggregation of Chinese hamster fibroblasts." Cell **20**(2): 363-371.

Vanhalst, K., P. Kools, K. Staes, F. Van Roy and C. Redies (2005). "δ-Protocadherins: a gene family expressed differentially in the mouse brain." Cellular and Molecular Life Sciences CMLS **62**(11): 1247-1259.

Vunnam, N. and S. Pedigo (2011). "Prolines in βA-sheet of neural cadherin act as a switch to control the dynamics of the equilibrium between monomer and dimer." Biochemistry **50**(32): 6959-6965.

Wheelock, M. J. and K. R. Johnson (2003). "Cadherins as modulators of cellular phenotype." Annual review of cell and developmental biology **19**(1): 207-235.

Wheelock, M. J., Y. Shintani, M. Maeda, Y. Fukumoto and K. R. Johnson (2008). "Cadherin switching." J Cell Sci **121**(6): 727-735.

Xiao, K., J. Garner, K. M. Buckley, P. A. Vincent, C. M. Chiasson, E. Dejana, V. Faundez and A. P. Kowalczyk (2005). "p120-Catenin regulates clathrin-dependent endocytosis of VE-cadherin." Molecular biology of the cell **16**(11): 5141-5151.

Yang, J. and R. A. Weinberg (2008). "Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis." Developmental cell **14**(6): 818-829.

Zheng, K., J. S. Laurence, K. Kuczera, G. Verkhivker, C. Russell Middaugh and T. J. Siahaan (2009). "Characterization of Multiple Stable Conformers of the EC5 Domain of E-cadherin and the Interaction of EC5 with E-cadherin Peptides." Chemical biology & drug design **73**(6): 584-598.