Single Molecule Force Spectroscopy of CNGA1

Sourav Maity

(2010-2014)

Supervisor : **Prof. Vincent Torre**

Thesis submitted for the partial fulfillment
of Doctor of philosophy in Neurobiology.

Scuola Internazionale Superiore di Studi Avanzati (SISSA)

SISSA – Via Bonomea 265- 34136 TRIESTE - ITALY

To my family

Acknowledgments

I would like to sincerely thank my advisor, professor Vincent Torre, he has been a mentor to me. His guidance helped me to find my path and to be the scientist I am now. If it weren't for him, who pushed and supported me all these years, I would never had the opportunity to achieve this much. From the very first day he was honest to me, I still remember the first interaction with him, after he gave me a brief description of the project, he told me "Sourav, this is a very difficult project, but if you work your best you will be in heaven. Are you ready to take this challenge?" I am glad of that day when I choose to take the risk and begin the journey. Thank you very much Vincent for giving me this opportunity, you are all the reasons I am writing this today! I am also grateful to Manuela for all her help during past four years. She is being most kind and most helpful person to me, official work or needed a help to check the English or needed some advise, she was always there.

Thanks to Monica Mazzolini who helped me a lot during the transformation from being just a student to a researcher. Her guidence for all these years has been really a great for me to become a biologiest from a physics student. I am also grateful to Alejandro Valbuena who taught me a lot about AFM. I am thankful to Arin Merchesi who had help me to understand many biological concepts. A hearty thank to Marco Lazzarino for all his valuable discussions, I feel more and more enthusiastic whenever I speak with him.

I am really thankful to Paolo Fabris for being my friends at work and building all those programs for Matlab. Without his expertise, I might not be able to make this work so superior- as he always says – "everything is possible". I also thank to Manuel Arcangeletti for being a lovely friend and for doing all those favours of electrophysiology experiments without asking anything in return. I am really grateful to have Andrea as my lab mate, all his understanding and cooperation in the lab made my life easier and enjoyable at work. Also I need to thank him for all those coffee and cheat chat in the balcony.

I would like also to thank all the members of the Torre lab. I will memorise all the fun those we shared in or out of the lab. And thank you also for being so nice persons and making the lab like a family working place.

I thank all my teachers from the day I learned to write 'A, B, C...', they have been and will be with me as long as I live. Without their blessing it could have been a different story. A special thanks to Dr. Gauri Kulkarni for showing me the path to where I am now.

I must be thankful to Wasim Bhai, Sheetal Di and Swathi Di for being my friends and family, and for looking after me those days when I actually needed them. I can not express in words my gratitude towards them, I am lucky to have you all in my life. I also would like to thank Priya for her moral

supports and for being a sweet sister.

I am also thankful to all my friends in Trieste and long back home, whose smiling face bring the joy of my life.

Of course no acknowledgments would be complete without giving thanks to my parents and my sister and her family. I always feel special when ever I think of them. They've taught me about hard work and self-respect and about how to be independent. Mom, especially, was a great role model of resilience, strength and character. They have always expressed how proud they are of me and how much they love me. I too am proud of them and love them very much. And big thanks to my sister, growing with her makes my childhood so special. I am also grateful to my father in law for all his supports and trusts that helped me to go on with my life.

And the last but certainly not the least, I must acknowledge with tremendous and deep thanks to my best friend and my better half, Monali. She has taken care of whatever needed tending to without complaining, just so I could focus on completing my work. She has patiently endured many, many long hours alone while I worked in the Lab. At the same time, she has also given me so many happy and beautiful memories throughout this journey. We have laughed and cried together, played and planned together, and whenever faced any problem, we have learned together. I could not have completed this journey without Monali by my side. Her presence brings all the beautiful thinks in my life, she is my true inspiration.

Table of Contents

Abstract:

Single molecule force spectroscopy (SMFS) is known to be one of the most powerful tool to investigate the relation between structure and function in molecules and proteins. The possibility to work in aqueous conditions at a single molecular level opens up an extraordinary perspective to investigate rare events at a molecular level of biological systems. Over the past years Atomic Force Microscopy (AFM) based on SMFS has provided us information, that is either difficult or impossible to get from any other method. In spite of its advancements, SMFS has not been applied to many molecules of biological relevance for several reasons, such as problems with the biological samples, data analysis and other technical issues. Indeed, the development and improvement of SMFS is becoming is very important to study biological molecules and proteins in their natural environment.

The work carried out during my PhD aimed to investigate the structure and function of CNGA1 channels in their natural environment by using *Xenopus Laevis* oocyte plasma-membrane. In my thesis I have worked on three main topics and I have obtained results on all of them. These results are in the process of being prepared for submission to scientific journals.

i) Firstly: I have combined techniques of mutagenesis, bio-informatics and single molecule force spectroscopy to investigate CNGA1 channels expressed in *Xenopus laevis* oocytes.

ii) Secondly: I have shown that the CNGA1 channel shows different unfolding pathways under perturbation.

iii) Thirdly: I have successfully implemented a methodology to characterize with SMFS rhodopsin and CNG channels in situ, i.e. in rod outer segment (ROS) from vertebrate photoreceptors.

Introduction

Ion channels are membrane proteins and act as a door for the movement of ions across the cell membranes, enabling in this way cells to communicate with the extracellular world or with other cells. They play essential roles in the physiology and pathophysiology of all cells. There are different types of channels, which differ in their ion selectivity and activation mechanism. Some channels are highly selective for certain types of ions (e.g. Na^+, K^+, or Ca^{2+}), while others are non-selective, i.e. they do not differentiate between different types of ion and transport them equally. Mutation on these proteins may cause several abnormalities in the system, for example, cystic fibrosis and retinitis pigmentosa (Sanders et al., 2000, 2004). 'channelopathy' describing those diseases that result from defects in the ion channel functions (Ashcroft F. M., Ion Channels and Diseases). To understand how a mutation in ion channels causes a disease requires a primary understanding of the molecular structure, physiological and biophysical properties of those ion channels. In my PhD project I have focused on the structural and functional properties of Cyclic Nucleotide Gated (CNG) ion channels, specifically CNGA1 channels and - to some extent - of the CNGB1 subunit, using Atomic Force Microscopy (AFM) based on single molecule manipulation. In the next section, I will first illustrate what is known about protein structure and then I will go through the structural details of CNGA1 channel from the information available in literature.

Cell membranes are composed of double layer of lipid molecules and - as a result - they are relatively impermeable to ions. Consequently, ion transportation across the cell membrane or between different intracellular compartments, is mediated by membrane proteins called ion channels, transporters and pumps. Ion channels are considered to be containing pore (gate) whose opening and closing may be intrinsic or regulatory. Ion channels are can be found in all animals, plants and bacterial cells and play important roles in diverse biological processes – such as nerve and muscle excitation, sensory transaction, hormonal secretion, learning and memory, salt-water balance control etc. So, one can imagine the depth of the topic of ion channel is immense in all fields of biology.

Physical interactions that determine the Properties of proteins

To determine the chemical properties of a molecule it is important to understand what are the physical interactions that can be activated in different sites of the molecule. I will introduce those interactions which can help to understand some of the mechanical and chemical properties of proteins.

Covalent Interaction

A covalent bond is a chemical bond that involves the sharing of electron pairs between two atoms. When two atoms have a comparable electronegativity, they tend to form covalent bonding. In a covalent bond there is a stable balance of attractive and repulsive forces between atoms, thus for a stable assembly of atoms, in most of the cases each atom attained the equivalent of the full outer shell, corresponding to a stable electronic configuration.

A peptide bond is a covalent chemical bond formed by two molecules when a carboxyl group of one molecule reacts with the amino group of the other molecule through a condensation reaction. Peptide bonds are responsible for connecting between different amino acids to form the amino acid chain.

Non-covalent Interaction

Although the covalent structure is adequate to characterize the chemical properties of a small molecule but they fail to explain the properties of protein. Long polypeptide chain of protein allow them to fold back among themselves and form a well defined three-dimensional structure. This complex three-dimensional structure of protein provides the unique environment and orientation of the functional groups that give proteins their many special properties. The biological activity of proteins are also mediated by their interactions with their environment : water, salt, membranes or other proteins, nucleic acids and all other molecules that are present in the living system. All those interactions arise from diverse set of non-covalent interactions. So, to understand the behavior of proteins it is up most necessary to understand the physical basis of these interactions.

Physical nature of non-covalent interactions are well established for individual molecules is vacuum or in solid, but not in liquid. On the other hand proteins show their folded conformation

only in liquid-water environment or in membrane. Though water is one of the most important biological mater, it is not the best understood molecule in biology. The limitation of understanding the interacting properties of water leads to the limitation of understanding the behavior of proteins in natural inhabitant.

Let us see some of the most important non-covalent interactions, those may take place on a polypeptide chain in liquid environment.

Short-Range interaction

Energetically and structurally most important interaction between atoms or molecules is repulsive interaction. When two atoms come near to each other so that their electron begins to overlap, they experience strong repulsive force. According to exclusion principle two electron cannot posses the same energy state. The repulsive energy said to increase with the inverse of the 12^{th} power of the distance between the centers of the two atoms. Since the force increases so rapidly, each atom make a impenetrable volume. The individual atoms are usually modeled as spheres, defined by the van der Waals radius. Van der Waals radius can be used to estimate approximately the volume of a molecule.

Electrostatic Force

Point charge: The most fundamental non-covalent attraction is electrostatic interaction. Coulomb's law states, when two point charge x and y, separated by distance r_{xy} in vacuum, the energy of electrostatic interaction is

$$\Delta E = Z_x \cdot Z_y \cdot \varepsilon^2 / r_{xy} \qquad (1)$$

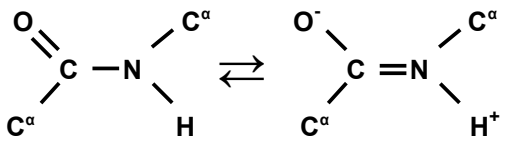
where, ε is the charge of an electron, and Z is the number of such charge in each atom. Now Coulomb's law is true in vacuum and for point charge. For other environment, the electrostatic interaction is modulated also by the contribution of the environment. In a homogeneous medium, the electrostatic interaction decreased by the dielectric constant D of the environment:

$$\Delta E = Z_x \cdot Z_y \cdot \varepsilon^2 / D \cdot r_{xy} \qquad (2)$$

Now, the value of D is always greater than unity. So, the electrostatic interaction in a media is always lower than that of a vacuum. In case of biological environment the estimation of electrostatic energy is less straight forward, since the biological environment is often inhomogeneous in nature. Moreover, special problem arises at interfaces between regions with very different bulk dielectric constants, such as between an aqueous solvent and a folded protein. Hence the pH is very significant parameter for protein functioning.

Another problem of dealing with Coulomb's law is the finite size of an atom. Coulomb's law is valid for point charges and in a scale of distance significantly greater than atomic dimension. For atoms the charge is separated between nucleus and the different electron cloud. So it cannot be taken as point charge at short distance. In case of protein this is even more complex, since some of the amino acid (for example Lus, Arg, His, Asp etc), the charge distribution of ionized group of the side chain is separated in between two or more hydrogen or oxygen atoms. Consequently, interaction between two oppositely charged group in a protein, known as salt bridges, usually consist not only electrostatic interaction but also at some extent of hydrogen bonding.

Dipoles: Molecules that does not have net charge also participate in electrostatic interaction. Because, even if the molecule is neutral it can have atoms with different electronegativity. Now inside the molecule the highest electronegative atom will have excess negative charge while the lowest electronegative atom will have excess of positive charge. The electronegativities of the atoms in proteins are 0, 3.45; N, 2.98; C, 2.55; S, 253; and H, 2.13. For example, peptide bond which has practically double-bonded character due to resonance with a form in which the more electronegative oxygen atom acquires a net partial negative charge, whereas the -NH- group is somewhat positively charged.



(3)

The separation of charge in a molecule determines its dipole moment μ_D , which is given by the product of the magnitude of the separated excess charge Z and the distance d by which it is separated

$$\mu_D = Z \cdot d \qquad (4)$$

A dipole can interact with a point charge or a dipole or even more complex charge separation known as quadrupole, octupole and so on. Interaction involving dipoles also modify the nature of the dipole charge distribution in the interacting molecules. Inhomogeneity of the medium and irregular charge distribution in the molecules leads protein environment most unfavorable to determine a direct calculation of the electrostatic energy.

Electrostatic interactions such as ion pairs and salt bridges in proteins have been an area of active research (Kumar and Nussinov 2002a,b). unlike some other non-covalent interactions like hydrogen bonds or hydrophobic forces, electrostatic interactions are highly specific and therefore play an important role in specifying the protein folding , protein flexibility and functions. Also there are some evidences that salt bridge enhance the thermal stability of protein from thermophilic organisms (Szilagyi and Zavodszky 2000, Li et al 2005; Razvi and Scholtz 2006).

Van der Waals Interaction

Every atoms and molecules attract each other, even if there is no charged group. These ubiquitous attractions known as van der Waals interaction. These interactions are weak and close range, which varies as the sixth power of the distance between them, $d^{-6.}$ Then arise from three types of interactions; those between two permanent dipoles (Keesom force), those between a permanent and induced dipoles (Debye force), and those between two mutually induced dipoles, known as London or dispersion forces.

Van der Waals interaction can be described as an energy potential, function of distance d that repulsive forces at close range. The most well known mode is the Lennard-Jones potential-

$$E(d) = 4\epsilon \left[\frac{C_n}{d^n} - \frac{C_6}{d^6} \right] \quad (n > 6) \qquad (5)$$

where the first term represent the repulsive contribution and the 2nd term gives the attractive contribution. C_n and C_6 are constants. A van der Waals interaction generally considered to be independent of the interactive molecule, but this is not always true. In case of thermal fluctuation of the van der Waals force can be repulsive or attractive relative to the mutual interaction of the molecule. A graphic demonstration of equation 5 will be given later.

Hydrogen bond

A hydrogen bond is the electromagnetic attractive interaction between two polar molecules, where two electronegative atoms compete for the same hydrogen atom. The hydrogen atom is covalently bonded to the donor molecule but it interact favorably with the acceptor molecule. For, not to be confused with covalent bond, hydrogen bond is not a true bond but effectively strong dipole-dipole attraction. Hydrogen bond can occur in both inter-molecule and intra-molecule and it has energy (5-30 KJ/mole) which is stronger than van der Waals interaction, but weaker than covalent or ionic bonds. The lengths and strengths of hydrogen bonds depend on the electronegativities of the acceptor and donor ; the greater their electronegativities , the shorter the distance and stronger the bond. Hydrogen bonds in proteins most frequently involves the C=0 and N-H groups of the peptide backbone. In these type of hydrogen bond the length is normally 1.9 – 2.0 Å (Pace et al., 1998). in the liquid environment of the proteins, hydrogen bond is one of the most important factors to be considered .

The Hydrophobic Interaction

As I have discussed in the previous section , in spite of having all the theoretical possibility to estimate the non-covalent energies of a molecule, the protein plays unlikely to follow those theories, and that is because of the liquid water environment of proteins. The interactive nature of water to polar molecules in the protein chain makes the fundamental laws of protein stability and activity. On the other hand non-polar molecules does not interact with water molecules and they does not participate on making hydrogen bond, which tend to be the most important phenomena in aqueous medium. As a result those non-polar molecules favor themselves, a non-polar environment rather than an aqueous environment. This preference of the non-polar atoms for non-aqueous environment is known as hydrophobic interaction (Muller N., 1990). It is a major contributor to the stability of proteins, nucleic acids and membranes. The hydrophobic effect is usually viewed as a combined effect of hydration (an entropic effect) and van der Waals interactions between solute molecules (an enthalpic effect) (Makhatadze and Privalov, 1995) . it is therefore entropic at low temperature and enthalpic at high temperatures, which result in a complex temperature dependence of its internal energy (Schellman, 1997).

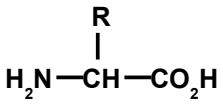
Apart from all these definite interactions, Protein-conformation itself opens a variety of possible interacting sites which can be inter-molecular or intra-molecular. In the next chapter I will

discuss about how these interaction can be used to explain the hierarchy of protein structure.

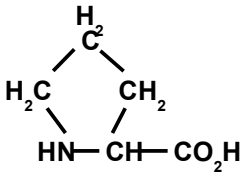
Protein Structure

Protein are more complex than most of the linear polymers in sense that they can assembled by 20 different monomers (amino acids) instead of only one and two monomers as it is in polymers. On the other hand proteins are structurally less complex. Polymers generally synthesized by chemically polymerizing of a mixture of monomers, this leads to a random sequence of monomers with a distribution of a chain length. But, natural proteins, however are linear and unbranched with a precise lengths exact sequence of amino acids. This sequence of amino acid differentiates between two different proteins.

Out of 20 amino acids found in proteins, 19 of them have general structure-



with different in the side chain, the 20th natural amino acid (proline) have different in structure.

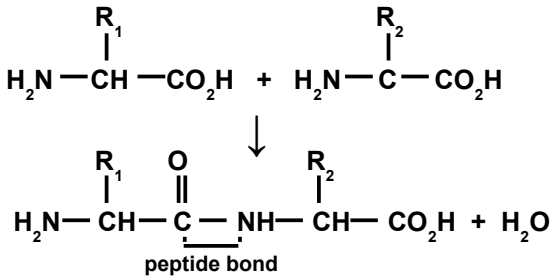


Where its side chain is bounded to the nitrogen atom.

The conformation of protein may be described at several levels of complexity. The structure can be described as its primary structure, secondary structure and tertiary structure. Some proteins made up of several subunits (tertiary structure) to form quaternary structure. The structure of protein is mainly determined by its amino acid sequence. Theoretically, therefore it should be possible to determine the three-dimensional structure of the protein from its primary sequence. But, in practice, this goal has not yet been achieved for any protein because the laws governing protein folding have not yet been elucidated.

Primary Structure

To understand the structure of protein at its up most complex level, it is important to have basic understanding of it's simpler form. The 20 amino acids are linked into positions by the peptide bond, through condensation of two amino acids ;



(6)

Generally proteins can have 50 to 25000 such amino acids linked through the peptide bonding. The simplest form of polymerized amino acids is – 'peptide', a short chain of residues with a defined sequence. The physical properties of peptide are those as expected from the sum of its amino acid residues. They do not have define three-dimensional structure. Proteins on the other hand have a definite sequence and three-dimensional structure under physiological conditions. Volume of similar proteins is conserved under similar biological environment. And most importantly the physical properties of proteins are way different than the sum of its individual residues (Creighton, 1989). For a protein this sequence of amino acid chain is referred to the primary structure of the protein. One can have basic information about a protein from its sequence analysis and comparing the sequence between different proteins can help to find the common ancestor of those proteins which leads to have the information regarding molecular evolution between their different families.

Though the physical properties of a protein are much different than its residue's, but primary sequence is mainly responsible for some of its structural properties. Amino acids can be positively charged (acidic), negatively charged (basic), polar but uncharged or non-polar (see table 1). Non-polar residues are hydrophobic therefore tend to be found in the trans-membrane domains of ion channel proteins, whereas charged and polar residues are more likely to be located in extracellular or intracellular regions of the proteins. Proteins have a rigid ring structure and introduce a bend in the polypeptide chain. It tends to be found in the end of an α-Helix or in the loop between two trans-membrane domain. Many type of ion channels are modulated by phosphorylation. Serine, tyrosine and threonine undergoes phosphorylation where a negatively charged phosphate group bins to the residue, thus alter the electrostatic properties of protein. Glycosylation, the attachment of a sugar residue to the side chain of the amino acid may also alter the structure of the protein. There are also

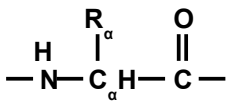
many other biological test that help to predict the structural properties of a protein for example – lipid attachment, sulfation, hydroxylation, disulphide bond promotion etc. (Creightion, 1989).

Side chain polarity	Amino acid	Abbreviations	Side chain charge (PH 7.4)	Hydropathy index
Nonpolar	Isoleucine	Ile (I)	Neutral	4.5
	Valine	Val (V)	Neutral	4.2
	Leucine	Leu (L)	Neutral	3.8
	Phenylalanine	Phe (F)	Neutral	2.5
	Methionine	Met (M)	Neutral	1.9
	Alanine	Ala (A)	Neutral	1.8
	Tryptophan	Trp (W)	Neutral	-0.9
	Proline	Pro (P)	Neutral	-1.6
Polar	Cysteine	Cys (C)	Neutral	2.5
	Glycine	Gly (G)	Neutral	-0.4
	Threonine	Thr (T)	Neutral	-0.7
	Serine	Ser (S)	Neutral	-0.8
	Tyrosine	Tyr (Y)	Neutral	-1.3
	Histidine	His (H)	Positive (10%) Neutral (90%)	-3.2
	Glutamine	Gln (Q)	Neutral	-3.5
	Asparagine	Asn (N)	Neutral	-3.5
Acidic	Aspartic acid	Asp (D)	Negative	-3.5
	Glutamic acid	Glu (E)	Negative	-3.5
Basic	Lysine	Lys (K)	Positive	-3.9
	Arginine	Arg (R)	Positive	-4.5

Table1: *list of the 20 amino acids with their corresponding electrostatic and hydropathic properties. Information has been taken from IUPAC-IUB Joint Commission on biochemical Nomenclature, 1985.*

Secondary structure

Two most commonly found secondary structures in ion channel proteins are α-Helix and β-sheet . The backbone of the linear polypeptide chain consist of repeated sequence of three atoms – the amide N, the C^{α} and the carbonyl C :



both the structure is a result of H-Bond between C=O group of one residue to NH group of another residue. Crystal structure analysis determines that α-helix are formed by the formation of H-bond between C_i=0 residue to the N_{i+4}-H residue. Thus forming a turn between 3.6 residues and the

distance for each turn is ~5.44 Å, thus leading to ~1.5 Å rise per residue. An α-helix of 20 amino acids should therefore be ~ 30 Å long, which is sufficient to span the membrane. It is believed that many of the trans-membrane domains of ion channels are formed from α-helices. The β-sheet is formed by hydrogen bonding between the peptide bonds of amino acids in different polypeptide chain parallel strands run in same direction and anti-parallel strands run in opposite direction (fig.1).

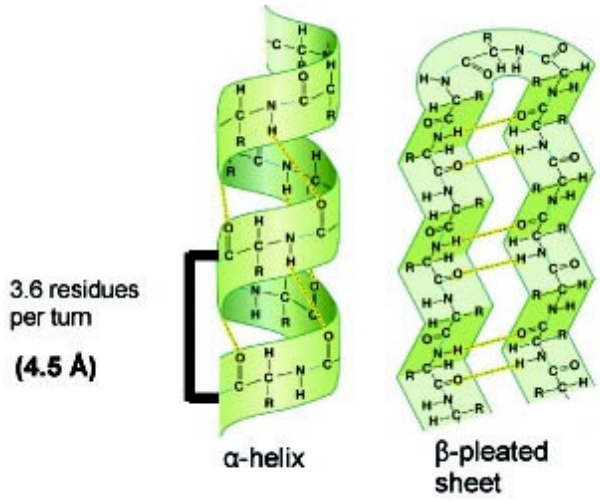


Figure 1: *cartoon representation of alpha helix and beta pleated sheet.*

The stereochemistrie's of the strands in the two cases are slightly different. Respective to their amino acid composition, anti parallel sheets are thought to be intrinsically more stable than parallel sheets.

β-sheets are significantly different from α-helices. Helices are virtually one-dimensional, where the interactions take place between residues which are very closed in primary structure. On the other hand for the β-sheet the interaction takes place between residues from different strands and are distant in the primary structure or that are from different molecules, depending whether the β-sheet in intra or inter molecular.

Apart from those two fundamental secondary structure that are observed in proteins, there are two different kind of secondary structures that can take place on a polypeptide chain. These structures are mainly slightly diverse configuration from α-helices, can be explained as more tightly or more loosely coiled- α-helices. If a hydrogen bond formed to the residues i+3 and i+5 , are designated the 3_{10} -helix and the π-helix respectively. The backbone of 3_{10} -helices are somewhere

packed tightly which inhibit the formation of a nonlinear hydrogen bonding. 3_{10}-helices are formed under special circumstances (e.g with the unnatural amino acid α-aminoisobutyric acid) or at the end of α-helices , where the final turn may have this conformation. It is thought to be mostly found in the membrane proteins or ion channels (Barlow & Thornton, 1988; Vieira-Pires et al., 2010). On the other hand π-helices are where the turn less lightly packed, so that there is a lack of contact between the backbone atoms. π-helices has usually not been observed in proteins.

Tertiary structure

Proteins can be found as diverse as in length of 50 – 25000 in nature. But average proteins have length more than 200 residues. Generally these proteins appear to consist of two, three or mote structural units, usually referred to as domains. These domains in a protein interact themselves with a active force less than what they have inside the structural element of each domain. Different types of non-covalent interactions such as ionic bonds, hydrogen bonding, hydrophobic interactions and van der Waals force stabilizer these domains in a confined three-dimensional structure known as tertiary structure of the protein. The tertiary structure of a protein defines most of its properties. Now when a protein situated in its biological environment i.e. in aqueous medium the interaction of protein and water molecules are completely inseparable. So, physical properties of the media i.e temperature, pH and pressure, are serving as really important quantity for the protein structure and stability. Theoretically there could be infinite number of possible tertiary structure that a protein can have in their natural environment. However, proteins follows a finite number of energetically steady state configuration (i.e minimum energy conformation) and usually classifies as individual conformations. For example, each amino acid residue may have eight different conformations. So, if there is a protein of 100 residues, then it may have 8^{100} different conformations. Some of these theoretical conformations are not possible, because they might overlap atoms in volume space. Unfortunately there is no calculation of total number of conformations yet being possible, though there can be a rough estimation.

Quaternary Structure

Many proteins formed an aggregation of two or more polypeptide chains, called subunits. Different subunits can be similar (homologous) or different (heterologous). This kind of aggregation can be dimers, tetramers or even higher order of quaternary configurations of identical polypeptide chain. Each subunit of a protein usually folded as independent globular conformation (monomers) –

which, then interact with other subunits through different kind of non-covalent interactions. There are two possible scenario for this assembly has been observed, isologous and heterologous. Isologous association is generally form dimer with two fold rotational symmetry. Normally this type of association are formed by two identical monomers which shares similar space and are have a complementary sites which are mirror image to each other. It is possible to form tetramer with isologous association, in those cases each dimers have one or more complementary site which interact with the second dimer and compile to force fold symmetry.

Heterologous associated results when two non-equivalent monomers have complementary site for each other. Heterologous interaction forms dimers of no axial symmetry and often a result of indefinite polymerization. Some elongated structural proteins, such a action are formed by indefinite heterologous polymerization of globular monomers. Each protein in the living organism serves a particular purpose. The tertiary structure of a protein gives the protein to fulfill their duty in a rectified manner i.e. even each protein serves one or more binding site they does not interact to every molecule or protein. For those interactions associated to two molecules requires only spatial and physical complementarity of the interacting surfaces, and this is the main property why each protein serve an unique biophysical properties. None the less this uniqueness of proteins is also responsible for the protein related diseases. This is why a major number of scientists have dedicated themselves to find a way to understand the properties of proteins.

Proteins in solution and in membranes

As I have discussed in the previous two sections, protein forms a compact folded conformation which validate it's chemical nature. Existence and stability of a protein in biological environment depends how they interact to their environment. On the basis of this concept we can explain the categories of natural protein; those exist in solution or water soluble proteins and membrane proteins.

Proteins in solution

Proteins gain their chemical, physical and biological properties from their folded conformation, unlikely to the unfolded polypeptides where their properties depends mainly on the assemble of their residues. The interactions of a protein molecule with solvent or with other molecules are determined primarily by its surface. The most favorable interactions with aqueous solvent are

provided by charged and polar groups of the hydrophilic side chains. The surface of most water soluble globular proteins is covered uniformly with charged and polar groups. Most proteins in aqueous solution are surrounded by a tightly bound hydration layer that has properties that are different from those of bulk water. This hydration layer contains ~ 0.3g of water per gram of protein, which is equivalent to nearly 2 H_2O molecule per amino acid residue. The solubility of proteins in water generally increase with pH values farther away from its isoelectric point, which is the pH at which the protein has zero net charge. Thus, the higher the net charge on the protein molecule, the greater the electrostatic repulsion between molecules, which tends to keep them in solution. But one cannot be misunderstood the solubility of protein is effective when the protein maintain its folded conformation, i.e. at certain pH, protein tend to unfold, which may open many non-polar site to the solvent, and a degradation result. Many proteins can also be dissolved in detergent or denaturants solution, but in those environment proteins are generally unfolded.

Proteins in membrane

Insulating nature of membranes forms a physical and chemical barrier between the cell interior to its environment. Membrane composed of two layers lipids molecules, where each lipid molecule composed of polar head and non-polar tails. In a biological environment the non-polar tails aggregate side by side to avoid contact with water. The interaction of water layer to the lipid bi-layer occurs mainly through the polar, usually charged head groups of the lipids. Proteins typically composed 25 – 75% of the mass of natural membranes. The membrane proteins mediate various important functions of the membrane such as transport of ions into or out of the cell, transducing chemical signals in the cell environment. Membrane proteins have amphipathic structure that reflects the membrane in which they have been embedded. They have polar surfaces which interact with the aqueous solution and polar lipid head and also non-polar surfaces that interact with the non-polar environment inside the bilayer. This dual nature of those trans-membrane domains makes them unfavorable to be stable in both aqueous and nonpolar environment. They have stability only when they are embedded into the membrane. This is one of the main reasons why most of the membrane proteins fails to crystallize. There are many experimental evidence of artificial lipid bilayers that have been used to study membrane proteins and ion channels (Goc et al., 2008; Shen et al., 2013).

Although the water-soluble proteins have mostly polar surface, the surface of membrane proteins, specially the interacting region to the bilayer are extremely non-polar. These surfaces are primarily covered by the side chain of Leu, Ile, Val and Phe residues (Fasman & Gilbert, 1990;

Heijne & Manoil, 1990).

Some physical properties of proteins

So far I have discussed that how a protein can be assembled in a biological space and what are the interacting forces responsible for a protein to be stable in it's natural environment. Proteins as it is complex in structure it shows also some complexity in their behavior. Here I discuss some of their basic behavior that often been observed even when they are in a biological environment.

Stability of proteins

Theoretically a molecule in space can be assumed to be stable when they are in their minimum energy state. Proteins, on the other hand play a bit tricky on this statement. The stability of protein depends mainly how they interact with their environment. As I have discussed before, structurally a protein (reasonably large) can have infinite number of conformations, not all of them can be stable, regardless to their energy minima. The stability of a protein is marginal at it's best conditions, a change in the environment conditions like pH, temperature, pressure or even in presence of certain molecules can disrupt it's folded conformation. This phenomenon is called 'Denaturation'. Denaturation of a protein often be reversible, exception, when denaturation caused by breaking of some covalent bonds.

Flexibility of protein structure

Crystallographic analysis have demonstrated that a given protein can have varying degree of structural flexibility. This flexibility can be due to small scale vibrations of bond length or angle. Also there can exist multiple conformation with similar energy and each conformation can be equally favorable. In those case the protein can shows higher order of flexibility to attain any of those conformations. There are also many experimental evidence of protein showing flexibility though hydrogen exchange with environments (Perrin, 1989), or could be because of the rotations of side chains in the protein surface (McCammon et al., 1983).

Protein Dynamic

Protein in homogeneous solution moves following the Brownian motion. On the other hand the

motion of membrane proteins are more specific. Membrane protein shows rapid diffusion (10^{-10} cm^2/s) into a membrane if there is not active interaction present in either side of the membrane. Proteins in membrane usually retain their vertical nature and does not flip their side. High density and bigger size of proteins in the membrane makes them really interactive to each other. This could be a reason, many proteins in membranes are oligomeric.

CNG Channels

CNG channels are allosteric proteins that are activated by direct binding of guanosine 3', 5'-cyclic monophosphate (cGMP) (Fesenko, 1985, Kcupp UB, Seifert R.2002, Matuef K, Zagotta WN, 2003), were first discovered in the plasma membrane of the other segment of vertebrate rod photoreceptor (Fesenko, 1985). CNG channels are identified to play important role in both visual (Yau and Baylor 1989) and olfactory (Zufall et al 1994) signal transduction. Over the past several years, CNG channels have been found in a varity of other nonsensory tissues including hippocampus, pancreas, kidney, testis and heart, but their detail functions have not been well understood (Bielet, 1993,94,96; Kaupp U.B., 1991; Yau, K.W 1994). CNG channels are members of the voltage gated family of channels, like voltage gated k^+ channels (Zagotta and Siegelbaum 1996). They are non selective cation channels composed of four subunits, (Gordon and Zogotta 1995, Liu et al, 1996; Varnum and Zogotta 1996). In vertebrates six members of the CNG channels have been identified and they are grouped in two subtypes – CNGA ($CNGA_1$, $CNGA_2$, $CNGA_3$, $CNGA_4$) and CNGB ($CNGB_1$, $CNGB_3$) (Bradley et al 2001). CNGA subtypes can form functional homomeric channel in heterologous expression systems, while the CNGB subtypes fail to form a functional homomeric channel, but they co-assemble with CNGA types subunits modulating their properties. In nature most or all of the CNG channels are thought to be assembled as a combination of different subunit types. For example, the CNG channels in rod photoreceptor contains 3 CNGA1 subunit and 1 CNGB1 subunit (Weitz D. et al 2002, Zheng J. Et al 2002, Zhong H. et al 2002 , Noah G. Shuaet et al 2011). Those of cone photoreceptor are thought to be composed of 2:2 subunit stoichiometry of CNGA3 and CNGB3 (Peng C. 2004).

CNG Channels in photo-transduction

Two types of photoreceptor cells are found in the retina of vertebrates: rods and cones (fig.2 A). These cells are organized in three principal parts: the outer segment, the inner segment and a synaptic ending. The absorption of light and transduction to electrical signaling occurred in the outer segment of both cell types (Palczewski et al. 1996). The outer segment is an elaborate, modified cilium that contains the biochemical machinery needed for visual transduction. The components of the photo-transduction cascade are packed into stacks of flattened, membranous vesicles (“discs”) that in rods are enclosed by the plasma membrane of the outer segment (Fig.2 A). Cones have a similar structure, but the disk membranes are an extension of the plasma membrane, arranged into a series of infoldings. The visual pigment is a G-protein-coupled receptor (GPCR)

that consist of a protein, opsin, covalently attached to a vitamin A-derived chromophore, 11-cis retinal (Hargrave P. A., 2001). The outer segment also houses the cGMP-gated channels and the Na^+/Ca^{2+}-K^+ exchangers. The inner segment contains the nucleus as well as mitochondria and the endoplasmic reticulum. In addition to these cellular organelles, the inner segment includes Na^+/K^+ pumps in its membrane. The photoreceptor cells, with the synaptic terminals, are connected to bipolar and horizontal cells (Watanabe & Matthews, 1988; Rieke & Schwartz, 1994; Savchenko *et al.*, 1997). In darkness, the cGMP-gated channels in the outer segment are open and allow a steady "dark-current" to enter the photoreceptor cells (Hagins *et al.*, 1970; Baylor *et al.*, 1979): Ca^{2+} and Na^+ ions are entering the cells through the cGMP-gated channels and Ca^{2+} ions are leaving the outer segment via the Na^+/Ca^{2+}-K^+ exchangers. Upon absorption of a photon, the chromophore is isomerized from 11-cis to all-trans (Emeis *et al.*, 1982). The activated visual pigment molecule triggers the transduction-cascade that ultimately results the closer of non-selective CNG channels, thus causes a decrease in the intracellular Ca^{2+} concentration (Yau & Nakatani, 1984).

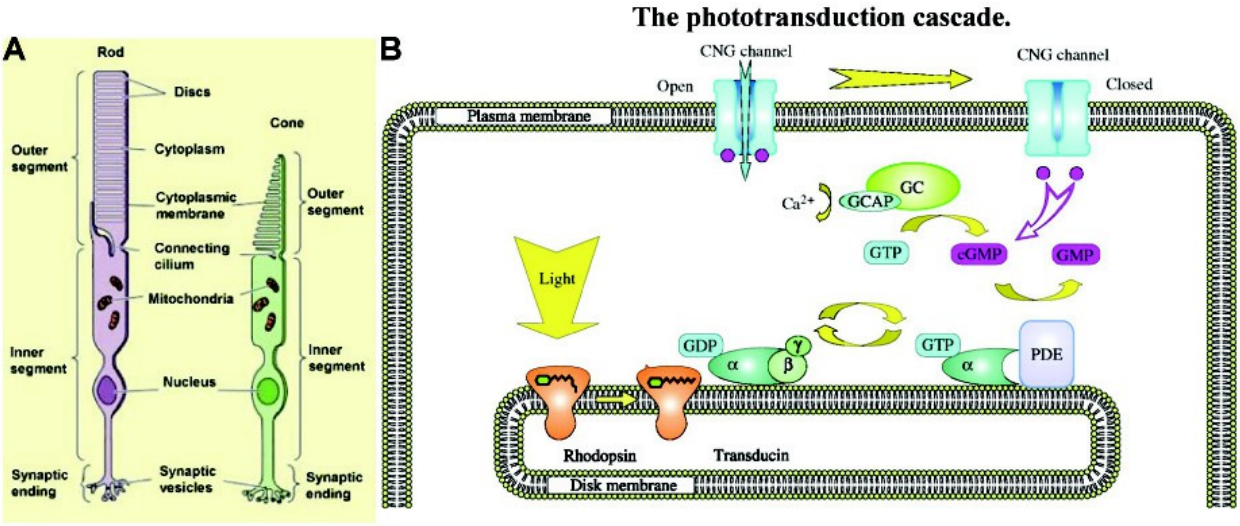


Figure2: *A-Cartoon of Rod and cone cells marking for their different celluler elements. B- photo-transduction cascade in rod photoreceptors. Phosphodiesterase (PDE), guanylyl cyclase-activating protein (GCAP) (Kramer & Molokanova, 2001).*

Structure of CNGA1

As a member of family of voltage gated k^+ channels (Jan and Jan 1990) , CNG channels are composed of four subunits around a water-filled pore (Gordon and Zagotta 1995) . Each subunit composed of six trans-membrane domain (S1 -S6), and intracellular amino-terminal and carboxy terminal region. In between S5 and S6 domain their is a small membrane embedded coiled structure known as P-loop (Kaupp et al 1991). The carboxy terminal regions composed of a cyclic nucleotide segment that connect between CNBD and S6 , called the C-linker (Matulef and Zagotta 2003) . All these structural segments are common to all subtypes of CNG channels (CNGA and CNGB), though their functions are thought to be different for different types of subunits. The amino-terminal of the channel have different architect for different subunits, e.g. the N-terminal of CNGB1 are longer (~ 656 aa) and have some specific functional and structural site, e.g. for CNGB1 the amino terminal contain a specific domain with a chain of Glutamic acid (GARP), which play important role in channel modulation. Here I will discuss in some details each segment of CNG channels.

Cyclic Nucleotide Bonding Domains (CNBD)

The CNG channels show a 20% sequence similarity with other cyclic nucleotide -bonding proteins e.g. cGMP and cAMP dependent protein kinase: PKG and PKA respectively- although the sequence similarity between these channels are low but they have a conserved residues that make important contacts with the bound cAMP. The crystal structure of CAP and PKA have shown identity (Su et al 1995; Weber et al 1987) considered to be a role model to represent the ligand binding domain of whole family of these channels. The CNBD of CAP contains an eight stranded anti-parallel β roll, followed by two α-helices (B and C). The cAMP binds to CAP in the anti configuration between the β roll and the c-helix (Altenhofen et al 1991; Weber et al 1989). Experimental studies shows that CNBD has at least two ligand specific sites: first is a threomine T560(CNGA) in the β roll and the second is in the C-helix, the D604 for CNGA1 shows a high effect on ligand specificity (Varnum et al 1995; Mazzolini et al., 2010) . Further experimental studies shows that cysteine residues in the c-helix produce an inter-subunit disulphide bond primarily when the channel is closed (Matulef and Zagotta 2002; Mazzolini et al 2002). They suggested that the c-helices might be nearer to each other or more flexible in the closed state of the channel and separate upon opening.

The CNBD of retinal and olfactory CNG channels shows 80% of sequence similarity, but they perform very differently to the cyclic nucleotide selectivity. The CNG channels in visual

sensory system shows high performance for CGMP, then cAMP , while the CNG in olfaction shows similar selectivity to cGMP and cAMP (Varnum et al 1995).

The channel efficiency is highly depends on the nucleotide concentration. The single channel experiment from Ruiz et al. (1997) state that CNG channels open more than they would spontaneously when two ligands were bound (approximately 1% of the maximum current), increases to 33% when three ligand is bond and maximal when all four subunits are bounded. These observations were further validated by Liu et al. (1998).

The C-Linker

The CNB domain and the last trans-membrane domain of the channel is connected by approximately 90 amino acid long chains, known as C-linker. The C-linker appears to be important for the allosteric opening of the channel. Some experimental observation shows that the c-linker is responsible for modulation of CNG channels by transition metal including Ni^{2+} and Zn^{2+} , (Karpen et al 1993; Gordon and Zagotta 1995). In CNGA1 channels, H420 residue coordinated Ni^{2+} between neighboring subunits and had a higher affinity for the open state then for the closed state(Gordon and Zagotta 1995) . Similar effect had been observed when histidine had been introduced in position 409, 413 and 417. This similarity in every 4^{th} amino acid support strongly with secondary structure prediction that this region is α-helices. In contrast to what observed in crystal structure of hyperpolarized activated cyclic nucleotide gated (HCN) channel, it can be proposed that the c-linker is composed of six small α-helices (A' – F') . where F'-helix is suggested to have conformational transition upon gating.

The Pore

In voltage gated ion channels and in the homologous CNG channels , the loop between the S_5 and S_6 trans-membrane segment (p region) is thought to form the lining of the pore. The basic architectural plan for the pore of this family of channels was revealed by the crystal structure of a **KcsA**, a bacterial k^+ channel (Doyle et al ,1998). An open channel conformation was revealed in 2002 by the crystal structure of another ion channel, the **MthK** channel (Jiang et al, 2002). The pore region of the channel controls both the single channel conductance and the pore diameter of the channel (Goulding E.H. et al ,1993).

The determination of how ions permeate the channel and how the ionic selectivity occur is of great

physiological importance because ion permeation is responsible for generating membrane depolarization, which is critical for electrical signaling; but CNG channels are permeable to Ca^{2+} , which is an important element in the activation of intracellular targets. Balasubrmaniam et al (1997) demonstrated that the permeation properties of olfactory CNG channels are significantly different than those in photoreceptor. They also displayed a lack of voltage dependence of in asymmetric solution, which suggested that ion binding sites were situated midway along the channel pore.

The Trans-membrane domain

Each subtypes of CNG channels are composed of six trans-membrane domains, and it is thought to have a conserved structural arrangement of those trans-membrane domains to all sub types. As similar to voltage gated k^+ channels, CNG channels have a voltage sensitive (S_4) domain. But the function of this domain in a ligand gated channel was not well understood. In presence of a saturating agonist CNG channels were thought to be voltage independent, even though it contains the voltage sensor. But recently Marchesi and Mazzolini et al (2012) have shown that CNGA1 shoes high voltage-dependency in presence of Rb^+, Cs^+ and organic cations. Some recent experimental evidence suggest a possible bending of S6 occurred in the opening of the channel (Flynn and Zagotta, 2001,2003; Nair et al 2009) though they do not act as a primary gate of CNG channels which have been located in the selectivity filter (Fodor et al, 1997; Becchetti et al, 1999; Giorgetti et al, 2005; Mazzolini et al, 2009). In spite of many attempts the crystal structure of CNG channel has not been achieved yet, which leads us to a lack in understanding the overall conformation of CNG channels.

Apart from all these main structural segments that are thought to be the main domains for functioning of the channel, CNG channels have two other important segments that have evidence to participate in the allosteric transition of the channel opening – the amino terminal domain (N-terminal) and the post- CNBD region (Fodor et al, 1997; Gordon and Zagotta 1995; Goulding et al, 1994; Liu et al, 1994) . Though their effects are not so strong for CNGA1 subtypes, but they have higher influence for CNGA2 and CNGB1 (Chen and Yan 1994; Liu et al 1994). The autoexcitatory region of the CNGA2 amino-terminal domain (Ca^{2+} /CaM binding domain) inhibits the channel's opening and plays a significant role in olfactory adaptation (Kurahashi and Manini 1997). There are also evidence that these autoexcitatory domains are directly responsible for inter-subunit interactions (Varnum and Zagotta, 1996). Gordon et al. (1997) have shown that the N-terminal and the C-terminal regions of each channel subunit interact. This interaction involved precise residues

and the formation of as disulphide bond between Cys 35 (N-terminal) and Cys 481 (C-linker). A recent finding shows that in retinal cell the glutamic acid- rich (GARP) domain in the CNGB1 N-terminal direct interaction with the disc rim specific scaffold protein , peripherin-2 (Becirovic et al., 2014).

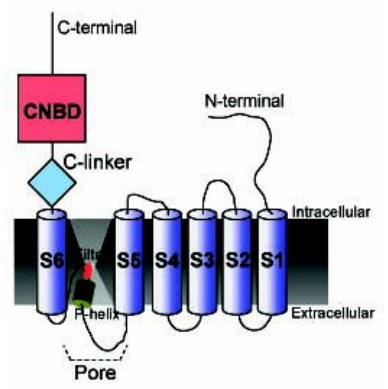


Figure3: *cartoon representation of CNGA1 single subunit.*

Structure determination and Atomic force microscopy

Determination of amino acid sequence (primary structure) is the first step to characterize a protein molecule. The second step is to find the secondary and tertiary structure. There are many advanced techniques playing their extraordinary role in structural biology. Here I quote some of those techniques those are being used in a large scale.

X-ray diffraction, analyzing the diffraction pattern of a crystallized protein molecule, then obtaining crystal structure and electron density of the protein can provide us the structure of the protein. For this technique the protein has to form a well defined crystal structure and should have a resistance to X-ray radiation. Also it is necessary for the protein to maintain their tertiary structure when crystallized. The technique is extraordinary in many sense, but it has some limitations too, e.g. many proteins, specially membrane proteins are difficult to be crystallized, also the information we obtain from crystallography is an ensemble averaging, so we may loose some information in a real time scale.

Electron microscopy, is a form of transmission electron microscopy (EM) where the sample is studied at cryogenic temperatures (generally liquid nitrogen temperatures). This technique often used when x-ray diffraction can not, mainly because the protein structure change if crystallized, for example most of the membrane proteins and all radiation sensitive proteins. In practice, the resolution of electron microscopy maps is not high enough to allow for unambiguous model construction as in x-ray diffraction. This technique has been successfully used for small proteins.

Nuclear magnetic resonance spectroscopy(NMR), is used to obtain information about the structure and dynamics of proteins, and also nucleic acids, and their complexes. NMR measures the interaction between a strong magnetic field and spin ½ nuclei by sending radio frequency signals through the sample, and measuring the absorption of those signals. Depending on the environment of atoms within the protein, the nuclei of individual atoms will absorb different frequencies of radio signals. Advantage of this technique is that, the measurement can be done in aqueous solution and at ambient temperature, thus protein can stay in their physiological environment. This technique is not suitable for large protein molecule since there will be many overlapping of absorption peaks, which then can not be evaluated easily.

Atomic Force Microscope

Atomic force microscopy (Binning et al., 1986) is part of the family of scanning probe microscopy and was invented by Binnig, Quate and Gerber in 1986. Last three decades, AFM has increasingly become a powerful tool for studying molecular processes at single molecule level (Bustamante et al 2000; Engel and Muller 2000; Fisher et al 2000). Two kinds of measurements are commonly made with such an instrument , (I) an imaging mode is used where the cantilever tip is scanned on the sample surface, direct interaction of of the AFM tip to the surface can be interpreted as the surface topography. (II) the second mode is a pulling mode where unfolding or unbinding forces and dynamics can be measured, providing detailed information about the folding or binding energy landscapes of biomolecules and complexes.

In the beginning of it's invention, AFM was mostly used for imaging (Drake et al 1989; Kasas et al 1997; Muller et al 1999), till date AFM has been proved itself to be a success in the field of structural biology. A continuous technical development and improvement of the sample preparation make AFM to reach a lateral resolution of ~0.5nm and a vertical resolution of ~0.1nm (Muller and Engel 1999, 2007; Muller et al 1999).

The mechanical unraveling of proteins using atomic force microscopy (AFM)(Bining et al 1986) is a powerful tool for studying protein structure. Until few years before, the only way to measure the stability of a protein was to change its physical (e.g, temperature, pressure) or chemical environment (denaturant, pH etc.) and monitor the loss of protein conformation using spectroscopic techniques from which the transition energy can be obtained. AFM in force spectroscopic mode, is capable of measuring a force at a range of piconewtons and can resolve force changes caused by the displacement of its probe by a fraction of a nanometer. Thus this technique open a vast era to investigate physical and chemical properties of molecules, atoms or many microscopic objects (Carrion-Vazquez, M and Fernandez et al., 2000; Fisher and Fernandez et al., 1999). Perhaps the most powerful aspect of AFM is that the measurement can be done in an aqueous environment, thus become one of the most suitable instrument to study of biological material under a bio-environment, with controlled conditions such as substrate, temperature, pressure, pH etc. AFM allows the manipulation at a single molecular level and therefore can uncover rare events that are not observed by using traditional bulk methods in which measurements are done by ensemble averaging. In contrast to other structural techniques (i.e crystallography), AFM works in a dynamic model and in real time. Since AFM can be operated in aqueous solution, the crystal form of the sample is not

required.

Instrumentation of AFM

The architecture of AFM is way simpler than it's applications, a conventional bio-AFM is consist of three major parts, i) an inverted microscope, ii) the head of AFM and iii) the electronic controller. A simple schematic diagram of a commercial JPK AFM is shown in figure 4. The head of AFM carry all important mechanical parts, the AFM cantilever is attached through a holder in the piezo electric controller which has a sub-nanometric sensitivity in all x, y and z direction. Top of the cantilever, the laser source play the role to detect the cantilever movements. The laser reflects from the edge of the cantilever to a highly efficient photo sensitive detector (PSD). A deflection in the cantilever will cause a change in reflecting angle for the laser, thus the deflection can be monitored by the PSD. The camera is mounted on the inverted microscope.

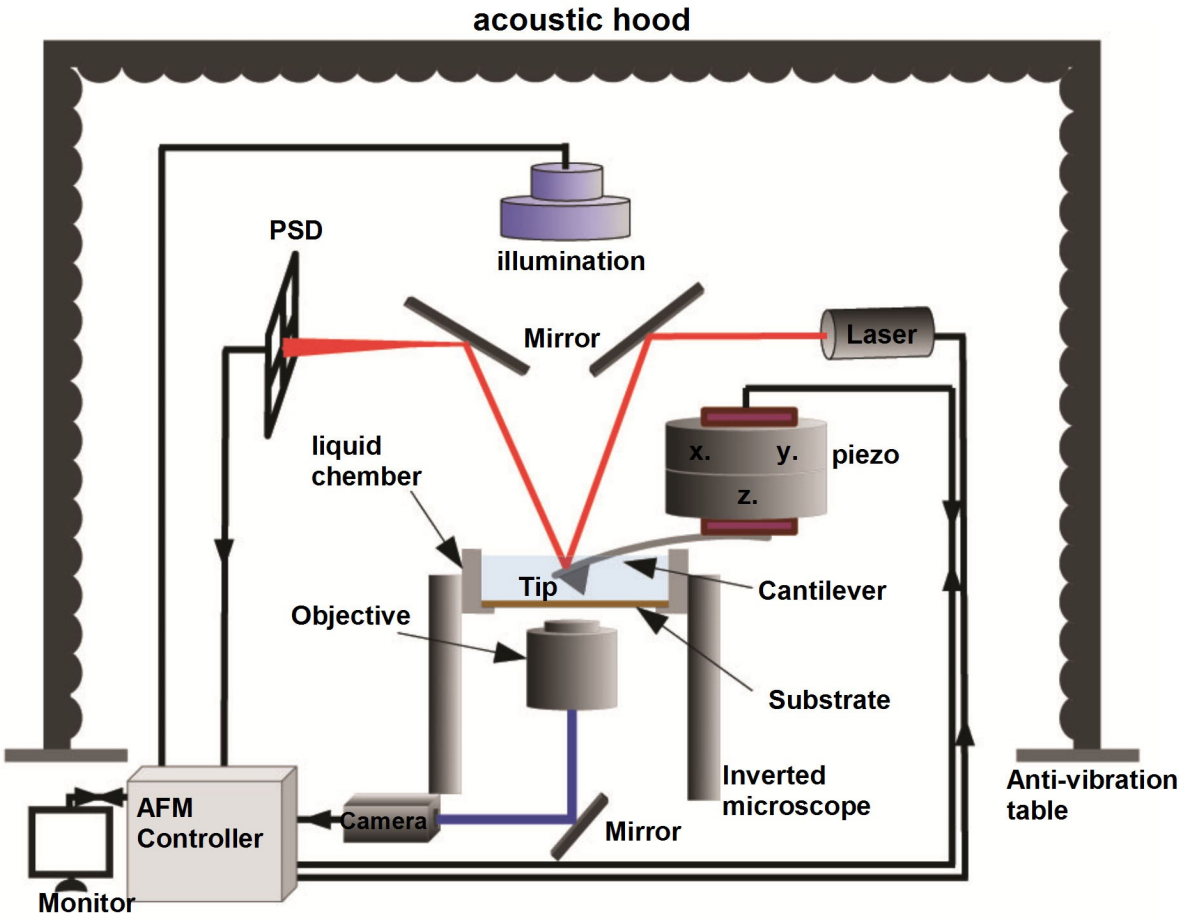


Figure 4: *a schematic diagram of JPK Nanowizard-3 AFM mounted on an inverted microscope. The setup is mounted on a anti-vibration table and isolated by a acoustic hood.*

The efficiency of AFM is highly depends on two major factor i) the cantilever: shape, sharpness and elasticity the cantilever highly effect on the results. For high resolution imaging one can expect to have a sharp tip, while in case for a soft sample the cantilever need to be soft too. In case of imaging in air, normally need comparatively hard but sharp cantilever. ii) the sample: both imaging or force spectroscopy by AFM need a very clean sample preparation. Any contamination in the sample surface or in the tip may cause artifacts in the results.

AFM imaging

The scanning of AFM cantilever can be operated in three modes; i) contact mode, ii) tapping mode and iii) non contact mode.

Contact mode

The contact mode can be run in two different conditions which are the constant height or the constant force. The laser beam measures the deflection of the cantilever, and imaging is collected on the basis of feedback to the piezoelectric scanner that keeps the force constant. Low spring constant AFM probes are used for contact mode imaging. The force applied between the tip and a surface in contact is given by Hooke's law.

$$\mathbf{F = -k_c \, . \, d} \qquad (7)$$

Where F is the force, k_c is the force constant of the cantilever and d the deflection distance. Forces between the tip and samples can be studied by using different tip materials and surfaces. The hardness / elasticity of the surface can be analyzed by varying the force at each point, which is commonly measured for biological samples in air or in liquid. Contact mode is used for high resolution imaging. However, this mode has some limitations: 1.Tip contamination caused by attachment of particles on a surface; 2 it is easy to damage the surface when a strong force is applied to scanning of soft samples.

Tapping mode

Tapping mode is dynamic and intermittent contact imaging in which the tip senses the surface with only minimal contact to the surface at a given amplitude. The tip oscillates with an amplitude of several nm. The typical frequency of the tip is 10-400 kHz. A resonant oscillation frequency is

given by the equation:

$$\omega = \sqrt{\frac{k_{eff}}{m}} \qquad (8)$$

Where k_{eff} is the effective force constant and m the effective cantilever mass.

Non-contact mode

Similar to the tapping mode, the non-contact mode brings the cantilever into oscillation slightly higher than the resonance frequency (usually 10-30% of resonance frequency) but the tip does not actually touch the sample surface, remaining 5-10 nm from the surface. The tip oscillates with the amplitude of several nm. The typical frequency of the tip is 10-400 kHz, depending upon the cantilever shape, dimension and material. This operation mode is suitable for 'soft' biological materials like bacteria, proteins, DNA or polypeptides. The spatial resolution in this mode is less than that one can achieve from contact mode.

All these three modes can be explained by the Lennard-Jones potential that describe the interaction between two atoms with respect to the inter-atomic distance (equation 5). A typical graphic representation of equation 5 is presented below.

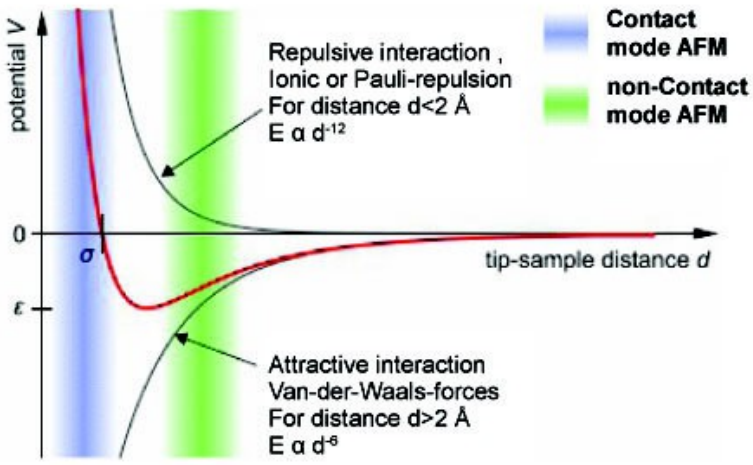


Figure 5: *a graphical representation of Lennard-Jones potential. Figure adapted from Steve Pawlizak, 2009.*

Single Molecule Force Spectroscopy

The term 'spectroscopy' is bit tricky, since there is no matter radiation interaction taking part in this kind of experiments, although, this is the term that usually used by scientific community. Mechanical unfolding of a molecule that attached specifically or non-specifically can be described by a single force-distance (F-D) curve.

Force distance curves

A typical force-extension or force-distance (F-D) curve is illustrated in Fig.6. Force curves offer a wealth of information on the mechanical properties, surface forces, adhesion energies and viscoelasticity resulting from the interactions between the tip and the sample surface. On a quasi-infinitely hard substrate (e.g. glass cover slip, mica), a characteristic graph is produced, which is discussed in the following.

Trace: First the cantilever approach towards the surface in a force free manner, so, no deflection but a declining distance is measured (a). When the cantilever reaches very close to the surface the cantilever can be suddenly attracted by the sample due to adhesion forces (e.g. electrostatic interaction), i.e. the cantilever flicks down the remaining distance and gives a small downward deflection (b). When the cantilever is moved further down, it bend upward since the surface impenetrable, this bending depends on the property of the cantilever and the surface (c). The slope of this section can be used to calculate the sensitivity of the cantilever.

Retrace: As soon as a defined setpoint of deflection is reached, the cantilever start to retract. The cantilever gets more and more unbent, while moving upwards again (d).The decrease in the bending of the cantilever is displayed by the linear decrease in the deflection signal. The deflection sensitivity (rate of signal strength correlated to extension in V/nm) can now be determined by fitting a line to this contact regime. At the contact point to the surface further retraction caused the cantilever bent downward due to nonspecific adhesion forces (e), until it suddenly loses contact and flicks up into its initial position. Now, at this stage if the cantilever tip attached (specifically or non-specifically) to any molecule in the surface a further retraction will cause unbinding of the attached molecule resulting as a sawtooth like pattern (f), where each energetically isolated segment of the molecule can be unfolded in each step. After the tip reach at a certain distance where the sample molecule is detached from the tip or the molecule detach from the surface, results the cantilever to

return to it's initial state (g), where no vertical deflection will be observed.

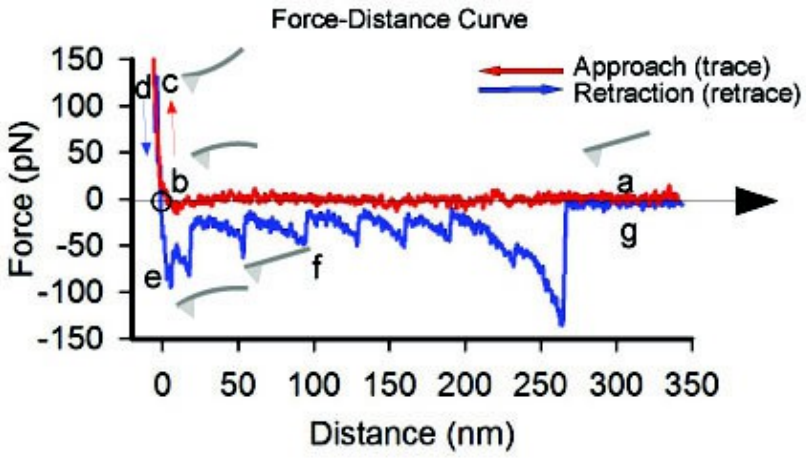


Figure 6: *a typical Force-Distance (F-D) curve showing trace and retrace of the AFM cantilever with respect to the surface.*

When the cantilever is calibrated, i.e. its sensitivity *s* and spring constant *k* are know, it is possible to calculate the applied force *F* which is proportional to the vertical cantilever deflection *x*.

$$F = -k \cdot x \qquad (9)$$

The force value usually lies in the range of piconewton (pN) to nanonewton (nN). The spring constant of a cantilever is principally affected by its geometrical properties, such as length, width and thickness. Additionally, the material of a cantilever (Young's modulus) further affects the spring constant. The theoretical spring constant of a rectangular cantilever can be derived from-

$$k = \frac{Et^3w}{4l^3} \qquad (10)$$

where, E is the Young's modulus, t is the thickness, w is the width and l is the length of the cantilever (Yeh et al., 2006). Typically, for biological AFM applications, rectangular or triangular cantilevers are used.

Measurement of spring constant

There are several way to calculate the spring constant of an AFM cantilever (Burnham et al., 2003). The most frequently used method for the calibration of AFM cantilevers is the so-called "thermal tuning" that was introduced by Hutter and Bechhoefer (Hutter et al., 1993). The cantilever needs to be kept in recording solution, at a fixed position at least 50 µm away from the surface. Brownian motions induce minor free oscillations of the cantilever. After Fourier transformation of the free

oscillation frequencies of the cantilever recorded over several seconds, a probability-oscillation frequency plot is generated. The resulting curve can be fitted with a power law, which allows determination of the actual spring constant of the respective cantilever. Measured spring constants can deviate by 50% compared to nominal spring constants provided by the manufacturer, this suggest one to calibrate the cantilever for each time of experiments.

Polymer extension models

In order to describe the stretching of a flexible polymer chains the freely jointed chain model was developed (Flory, 1969). From this model, the worm-like chain (WLC) model emerged (Bustamante et al., 1994). The WLC model describes the extension of semi-flexible polymer chains (Bustamante et al., 1994, Marko et al., 1995). The force F required to stretch the polymer chain is given by -

$$\mathbf{F(D)=k_bT/L_p[0.25(1-D/L_c)^{-2} + D/L_c - 0.25]} \qquad (11)$$

where $\mathbf{k_b}$ is the Boltzmann constant and T is the absolute temperature. L_p is the persistence length of the polymer and describes its rigidity. Values for L_c are specific for each polymer and must be experimentally determined (Bouchiat et al., 1999). A persistence length of 4 Å has been successfully applied to describe the extension at forces higher than 45 pN (Rief et al., 1997; Oesterhelt et al., 2000). F-D curves recorded upon unfolding of single polypeptides can be fitted using the WLC model (Puchner et al., 2009). The contour length Lc obtained from fitting a force peak using the WLC model describes the length of the polypeptide that had been unfolded and stretched.

Bibliography

Altenhofen W, Ludwig J, Eismann E, Kraus W, Bönigk W, Kaupp UB. Control of ligand specificity in cyclic nucleotide-gated channels from rod photoreceptors and olfactory epithelium.Proc NatlAcadSci U S A. 1991 Nov 1;88(21):9868-72.

Ashcroft F. M., Ion Channels and Diseases, academic press; 2000.

Balasubramanian S, Lynch JW, Barry PH. Concentration dependence of sodium permeation and sodium ion interactions in the cyclic AMP-gated channels of mammalian olfactory receptor neurons.J Membr Biol.1997 Sep 1;159(1):41-52.

Balasubramaniam V. Magnetic resonance image-guided stereotactic cingulotomy for intractable psychiatric disease.Neurosurgery. 1997 May;40(5):107-8.

Boesze-Battaglia K, Fliesler SJ, Albert AD.Relationship of cholesterol content to spatial distribution and age of disc membranes in retinal rod outer segments.J BiolChem. 1990 Nov 5;265(31):18867-70.

Barlow DJ, Thornton JM.Helix geometry in proteins.JMol Biol. 1988 Jun 5;201(3):601-19.

Baylor DA, Lamb TD &Yau KW (1979). The membrane current of single rod outer segments.J Physiol (Lond) 288, 589–611.

Becchetti A, Gamel K & Torre V (1999).Cyclic nucleotide-gated channels. Pore topology studied through the accessibility of reporter cysteines. J Gen Physiol114, 377–392.

Becirovic E, Nguyen ON, Paparizos C, Butz ES, Stern-Schneider G, Wolfrum U, Hauck SM, Ueffing M, Wahl-Schott C, Michalakis S, Biel M. Peripherin-2 couples rhodopsin to the CNG channel in outer segments of rod photoreceptors.HumMol Genet. 2014 Jun 24.pii: ddu323.

Biel M, Zong X, Distler M, Bosse E, Klugbauer N, Murakami M, Flockerzi V, Hofmann F. Another member of the cyclic nucleotide-gated channel family, expressed in testis, kidney, and heart.ProcNatlAcadSci U S A. 1994 Apr 26;91(9):3505-9.

Biel M, Zong X, Hofmann F.Cyclic nucleotide-gated cation channels molecular diversity, structure, and cellular functions. Trends Cardiovasc Med. 1996 Nov;6(8):274-80.

Biel M, Zong X, Ludwig A, Sautter A & Hofmann F (1999).Structure and function of cyclic nucleotide-gated channels.Rev PhysiolBiochemPharmacol135, 151–171.

Biel M, Altenhofen W, Hullin R, Ludwig J, Freichel M, Flockerzi V, Dascal N, Kaupp UB,Hofmann F. Primary structure and functional expression of a cyclic nucleotide-gated channel from rabbit aorta.FEBS Lett. 1993 Aug 23;329(1-2):134-8.

Binnig G., Quate C. F., Gerber C. (1986): Atomic Force Microscope, Phys. Rev. Lett. 56(9):930-933 .

Bouchiat, C., M.D.Wang, J. Allemand, T. Strick, S. M. Block, and V. Croquette (1999).Estimating the persistence length of a worm-like chain molecule from force-extension measurements. Biophys. J. 76:409-413.

Bradley J, Frings S, Yau KW & Reed R (2001).Nomenclature for ion channel subunits.Science 294, 2095–2096.

Burnham, N. A., X. Chen, C. S. Hodges, G. A. Matei, E. J. Thoreson, C. J. Roberts, M. C. Davies, and S. J. B. Tendler (2003). Comparison of calibration methods for atomic-force microscopy cantilevers. Nanotechnology 14:1-6.

Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith (1994).Entropic Elasticity of Lambda-Phage DNA.Science 265:1599-1600.

Bustamante C, Smith SB, Liphardt J, Smith D. Single-molecule studies of DNA mechanics. Curr OpinStruct Biol. 2000 Jun;10(3):279-85.

Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering.ProgBiophysMol Biol. 2000;74(1-2):63-91. Review.

Creighton, ed. (1989) Protein structure : IRL press oxford.

Drake B, Prater CB, Weisenhorn AL, Gould SA, Albrecht TR, Quate CF, Cannell DS, Hansma HG, Hansma PK. Imaging crystals, polymers, and processes in water with the atomic force microscope. Science. 1989 Mar 24;243(4898):1586-9.

Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT & MacKinnon R (1998). The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 280, 69–77.

Emeis D, Kühn H, Reichert J & Hofmann KP (1982). Complex formation between metarhodopsin II and GTP-binding protein in bovine photoreceptor membranes leads to a shift of the photoproduct equilibrium. FEBS Lett143, 29–34.

Engel A, Müller DJ. Observing single biomolecules at work with the atomic force microscope.Nat Struct Biol. 2000 Sep;7(9):715-8.

Fasman G.D. And Gilbert W.A. (1990). The prediction of transmembrane proteins sequence and their conformation: an evaluation. Trends Biochem Sci. 15:89-92.

Fesenko EE, Kolesnikov SS, Lyubarsky AL. Induction by cyclic GMP of cationic conductance in plasma membrane of retinal rod outer segment.Nature. 1985 Jan 24-30;313(6000):310-3.

Fisher TE, Marszalek PE, Fernandez JM. Stretching single molecules into novel conformations using the atomic force microscope.NatStruct Biol. 2000 Sep;7(9):719-24. Review.

Fisher TE, Marszalek PE, Oberhauser AF, Carrion-Vazquez M, Fernandez JM. The micro-mechanics of single molecules studied with atomic force microscopy.J Physiol. 1999 Oct 1;520 Pt 1:5-14.

Fisher TE, Oberhauser AF, Carrion-Vazquez M, Marszalek PE, Fernandez JM.The study of protein mechanics with the atomic force microscope.TrendsBiochem Sci. 1999 Oct;24(10):379-84. Erratum in: Trends BiochemSci 2000 Jan;25(1):6.

Flory, P. J. (1969). Statistical mechanics of chain molecules Interscience Publishers, New York.

Flynn GE &Zagotta WN (2001). Conformational changes in S6 coupled to the opening of cyclic nucleotide-gated channels. Neuron 30, 689–698.

Flynn GE &Zagotta WN (2003). A cysteine scan of the inner vestibule of cyclic nucleotide-gated channels reveals architecture and rearrangement of the pore. J Gen Physiol121, 563–582.

Fodor AA, Black KD &Zagotta WN (1997). Tetracaine reports a conformational change in the pore of cyclic nucleotide-gated channels. J Gen Physiol110, 591–600.

Giorgetti A, Nair AV, Codega P, Torre V &Carloni P (2005). Structural basis of gating of CNG channels.FEBS Lett579, 1968–1972.

Goc A, Angel TE, Jastrzebska B, Wang B, Wintrode PL, Palczewski K. Different properties of the native and **reconstituted** heterotrimeric G protein transducin.Biochemistry. 2008 Nov 25;47(47):12409-19. doi:

Goulding EH, Tibbs GR, Siegelbaum SA.Molecular mechanism of cyclic-nucleotide-gated channel activation.Nature. 1994 Nov 24;372(6504):369-74.

Goulding EH, Ngai J, Kramer RH, Colicos S, Axel R, Siegelbaum SA & Chess A (1992). Molecular cloning and single-channel properties of the cyclic nucleotide-gated channel from catfish olfactory neurons.Neuron 8, 45–58.

Goulding EH, Tibbs GR, Liu D &Siegelbaum SA (1993).Role of H5 domain in determining pore diameter and ion permeation through cyclic nucleotide-gated channels.Nature 364, 61–64.

Goulding EH, Tibbs GR, Siegelbaum SA.Molecular mechanism of cyclic-nucleotide-gated channel activation.Nature. 1994 Nov 24;372(6504):369-74.

Gordon SE &Zagotta WN (1995). Localization of regions affecting an allosteric transition in cyclic nucleotide-activated channels. Neuron 14, 857–864.

Gordon, S.E., Varnum, M.D. &Zagotta, W.N. Direct interaction between amino- and carboxyl-terminal domains of cyclic nucleotide-gated channels.*Neuron***19,** 431-441 (1997).

Hagins WA, Penn RD &Yoshikami S (1970).Dark current and photocurrent in retinal rods.Biophys J 10, 380–412.

Hargrave, P. A. (2001) Invest. Ophthalmol. Vis. Sci. 42, 3–9

Heijne von G. and Manoil C., (1990). Membrane proteins: from sequence to structure. Protein Eng-4:109-112.

Hutter, J. L., and J. Bechhoefer (1993).Calibration of Atomic-Force Microscope Tips.Rev. Sci. Instrum.64:3342-3342.

Jan, L.Y. & Jan, Y.N. A superfamily of ion channels. *Nature* 345**,** 672 (1990).

Jiang Y, Lee A, Chen J, Cadene M, Chait BT & MacKinnon R (2002). Crystal structure and mechanism of a calcium-gated potassium channel. Nature 417, 515–522.

Karpen JW, Brown RL, Stryer L & Baylor DA (1993).Interactions between divalent cations and the gating machinery of cyclic GMP-activated channels in salamander retinal rods.J Gen Physiol101, 1–25.

Karpen JW, Zimmerman AL, Stryer L & Baylor DA (1988). Gating kinetics of the cyclic-GMPactivated channel of retinal rods: flash photolysis and voltage-jump studies. ProcNatl AcadSci USA 85, 1287–1291.

Kasas S, Thomson NH, Smith BL, Hansma HG, Zhu X, Guthold M, Bustamante C, Kool ET, Kashlev M, Hansma PK. Escherichia coli RNA polymerase activity observed using atomic force microscopy.Biochemistry. 1997 Jan 21;36(3):461-8.

Kaupp UB, Niidome T, Tanabe T, Terada S, Bönigk W, Stühmer W, Cook NJ, Kangawa K, Matsuo H & Hirose T (1989). Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel.Nature 342, 762–766.

Kaupp UB. The cyclic nucleotide-gated channels of vertebrate photoreceptors and olfactory epithelium.TrendsNeurosci. 1991 Apr;14(4):150-7.

Kaupp UB & Seifert R (2002).Cyclic nucleotide-gated ion channels.Physiol Rev 82, 769–824.

Kedrov, A., H. Janovjak, K. T. Sapra, and D. J. Muller (2007).Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy.Annu. Rev. Biophys. Biomol.Struct.36:233-260.

Koch KW (1992). Biochemical mechanism of light adaptation in vertebrate photoreceptors.Trends BiochemSci17, 307–311.

Koch KW &Stryer L (1988).Highly cooperative feedback control of retinal rod guanylatecyclaseby

calcium ions. Nature 334, 64–66.

Koutalos Y & Yau KW (1996).Regulation of sensitivity in vertebrate rod photoreceptors by calcium. Trends Neurosci19, 73–81.

Kramer RH & Siegelbaum SA (1992). Intracellular Ca2+ regulates the sensitivity of cyclic nucleotide-gated channels in olfactory receptor neurons. Neuron 9, 897–906.

Kumar S, Nussinov R. Close-range electrostatic interactions in proteins. Chembiochem. 2002 Jul 2;3(7):604-17.

Kumar S, Nussinov R. Relationship between ion pair geometries and electrostatic strengths in proteins.Biophys J. 2002 Sep;83(3):1595-612.

Kurahashi T &Menini A (1997).Mechanism of odorant adaptation in the olfactory receptor cell.Nature 385, 725–729.

Lagnado L & Baylor DA (1994). Calcium controls light-triggered formation of catalytically active rhodopsin. Nature 367, 273–277.

Laio A & Torre V (1999).Physical origin of selectivity in ionic channels of biological membranes. Biophys J 76, 129–148.

Li G, Xi N, Wang DH. Investigation of angiotensin II type 1 receptor by atomic force microscopy with functionalized tip.Nanomedicine.**2005** Dec;1(4):306-12.

Liu J &Siegelbaum SA (2000). Change of pore helix conformational state upon opening of cyclic nucleotide-gated channels. Neuron 28, 899–909.

Lolley RN &Racz E (1982).Calcium modulation of cyclic GMP synthesis in rat visual cells.Vision Res 22, 1481–1486.

Liu DT, Tibbs GR, Paoletti P, Siegelbaum SA. Constraining ligand-binding site stoichiometry suggests that a cyclic nucleotide-gated channel is composed of two functional dimers. Neuron. 1998 Jul;21(1):235-48.

Makhatadze GI, Privalov PL. Energetics of protein structure.Adv Protein Chem. 1995;47:307-425.

Marchesi A, Mazzolini M, Torre V. A ring of threonines in the inner vestibule of the pore of CNGA1 channels constitutes a binding site for permeating ions.J Physiol. 2012 Oct 15;590(Pt 20):5075-90. doi: 10.1113/jphysiol.2012.238352. Epub 2012 Aug 6.

Marchesi A, Mazzolini M, Torre V. Gating of cyclic nucleotide-gated channels is voltage dependent.NatCommun. 2012 Jul 24;3:973.

Mari SA, Pessoa J, Altieri S, Hensen U, Thomas L, Morais-Cabral JH & Müller DJ (2011). Gating of the MlotiK1 potassium channel involves large rearrangements of the cyclic nucleotidebinding domains. ProcNatlAcadSci USA 108, 20802–20807.

Marko, J. F., and E. D. Siggia (1995).Stretching DNA.Macromolecules 28:8759- 8770.

Matulef K, Zagotta WN. Multimerization of the ligand binding domains of cyclic nucleotide-gated channels.Neuron. 2002 Sep 26;36(1):93-103.

Matulef K, Zagotta WN. Cyclic nucleotide-gated ion channels.Annu Rev Cell Dev Biol.2003;19:23-44.

Mazzolini M, Anselmi C & Torre V (2009). The analysis of desensitizing CNGA1 channels reveals molecular interactions essential for normal gating. J Gen Physiol133, 375–386.

Mazzolini M, Marchesi A, Giorgetti A & Torre V (2010). Gating in CNGA1 channels.Pflugers Arch 459, 547–555.

Mazzolini M, Punta M, Torre V. Movement of the C-helix during the gating of cyclic nucleotide-gated channels.Biophys J. 2002 Dec;83(6):3283-95.

McCammon RL, Stoelting R, Madura JA. Reversal of fentanyl induced spasm of the sphincter of Oddi.SurgGynecol Obstet. 1983 Mar;156(3):329-34.

Menini A, Rispoli G & Torre V (1988).The ionic selectivity of the light-sensitive current in isolated rods of the tiger salamander.J Physiol (Lond) 402, 279–300.

Marchesi A, Mazzolini M, Torre V. A ring of threonines in the inner vestibule of the pore of CNGA1 channels constitutes a binding site for permeating ions.J Physiol. 2012 Oct 15;590(Pt 20):5075-90. doi: 10.1113/jphysiol.2012.238352. Epub 2012 Aug 6.

Marchesi A, Mazzolini M, Torre V. Gating of cyclic nucleotide-gated channels is voltage dependent.NatCommun. 2012 Jul 24;3:973.

Muller, D. J., H. J. Sass, S. A. Muller, G. Buldt, and A. Engel (1999). Surface structures of native bacteriorhodopsin depend on the molecular packing arrangement in the membrane. J. Mol. Biol. 285:1903-1909.

Muller, D. J., and A. Engel (2007).Atomic force microscopy and spectroscopy of native membrane proteins.Nat. Protoc. 2:2191-2197.

Muller N. (1990), Search for a realistic view of hydrophobic effects. Acc. Chem. Res-23:23-28.

Nair AV, Nguyen CHH &Mazzolini M (2009).Conformational rearrangements in the S6 domain and C-linker during gating in CNGA1 channels.EurBiophys J 38, 993–1002.

Perrin C. L., (1989). Proton exchange in amides: Surprises from simple systems. Acc. Chem. Res. 22:268-275.

Peng C, Rich ED, Varnum MD. Subunit configuration of heteromeric cone cyclic nucleotide-gated channels. Neuron. 2004 May 13;42(3):401-10.

Puchner, E. M., and H. E. Gaub (2009). Force and function: probing proteins with AFM-based force spectroscopy. Curr.Opin.Struct. Biol. 19:605-614.

Razvi A, Scholtz JM. Lessons in stability from thermophilic proteins. Protein Sci. 2006 Jul;15(7):1569-78.

Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub (1997).Reversible unfolding of individual titin immunoglobulin domains by AFM.Science 276:1109-1112.

Rieke F & Schwartz EA (1994). A cGMP-gated current can control exocytosis at cone synapses. Neuron 13, 863–873.

Ruiz ML, Karpen JW.Single cyclic nucleotide-gated channels locked in different ligand-bound states. Nature.**1997** Sep 25;389(6649):389-92.

Sanders, C. R., and J. K. Nagy (2000).Misfolding of membrane proteins in health and disease: the lady or the tiger? Curr.Opin.Struct. Biol. 10:438-442.

Sanders, C. R., and J. K. Myers (2004).Disease-related misassembly of membrane proteins.Annu. Rev. Biophys. Biomol.Struct.33:25-51.

Savchenko A, Barnes S & Kramer RH (1997). Cyclic-nucleotide-gated channels mediate synaptic feedback by nitric oxide. Nature 390, 694–698.

Schellman JA. Temperature, stability, and the hydrophobic interaction.Biophys J. 1997 Dec;73(6):2960-4.

Shen HH, Lithgow T., Martin LL., (2013). Reconstitution of Membrane proteins into Model Membranes: Seeking Better Ways to Retain Protein Activities. Int. J. Sci. 14, 1589-1607.

Szilágyi A, Závodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure. 2000 May 15;8(5):493-504.

Varnum MD, Black KD, Zagotta WN. Molecular mechanism for ligand discrimination of cyclic nucleotide-gated channels.Neuron. 1995 Sep;15(3):619-25.

Varnum MD, Zagotta WN. Subunit interactions in the activation of cyclic nucleotide-gated ion channels.Biophys J. 1996 Jun;70(6):2667-79.

Vieira-Pires RS, Morais-Cabral JH.3(10) helices in channels and other membrane proteins.J Gen Physiol.2010 Dec;136(6):585-92.

Watanabe S & Matthews G (1988). Regional distribution of cGMP-activated ion channels in the plasma membrane of the rod photoreceptor. J Neurosci8, 2334–2337.

Weitz D, Ficek N, Kremmer E, Bauer PJ, Kaupp UB. Subunit stoichiometry of the CNG channel of rod photoreceptors.Neuron. 2002 Dec 5;36(5):881-9.

Yau KW &Nakatani K (1984).Cation selectivity of light-sensitive conductance in retinal rods.Nature 309, 352–354.

Yau KW. Cyclic nucleotide-gated channels: an expanding new family of ion channels.ProcNatlAcadSci U S A. 1994 Apr 26;91(9):3481-3

Yeh, M. K., B. Y. Chen, N. H. Tai, and C. C. Chiu (2006). Force measurement by AFM cantilever with different coating layers. Key. Eng. Mat. 326-328:377-380.

Zhong H, Molday LL, Molday RS, Yau KW.The heteromeric cyclic nucleotide-gated channel adopts a 3A:1B stoichiometry.Nature. 2002 Nov 14;420(6912):193-8.

Zagotta WN1, Siegelbaum SA. Structure and function of cyclic nucleotide-gated channels.AnnuRev Neurosci. 1996;19:235-63.

Zheng J1, Trudeau MC, Zagotta WN. Rod cyclic nucleotide-gated channels have a stoichiometry of three CNGA1 subunits and one CNGB1 subunit. Neuron. 2002 Dec 5;36(5):891-6.

Zhong H, Molday LL, Molday RS, Yau KW. The heteromeric cyclic nucleotide-gated channel adopts a 3A:1B stoichiometry. Nature. 2002 Nov 14;420(6912):193-8.

Zufall F, Firestein S, Shepherd GM. Cyclic nucleotide-gated ion channels and sensory transduction in olfactory receptor neurons.Annu Rev BiophysBiomolStruct. 1994;23:577-607.

Zufall F, Shepherd GM & Barnstable CJ (1997). Cyclic nucleotide gated channels as regulators of CNS development and plasticity. CurrOpinNeurobiol7, 404–412.

Results

Manuscript 1

Submitted

Conformational rearrangements in the transmembrane domain of CNGA1 channels upon gating revealed by Single Molecule Force Spectroscopy

Sourav Maity, Mazzolini Monica, Manuel Arcangeletti,
Alejandro Valbuena[1], Paolo Fabris, Marco Lazzarino and Vincent Torre[#]

Author Contributions:

S.M performed SMFS experiments and data analysis and prepared the figures; M.M. and M.A. performed electrophysiology and mutagenesis; S.M., P.F and V.T performed bioinformatics; S.M., M.M., A.V. and V.T. designed the experiments; S.M., M.M., M.L. and V.T. participated in interpreting results and contributed to writing the paper; V.T. supervised the project.

We are thankful to Dr. Loredana Casalis for providing us the gold surface and M. Lough for checking the Engligh.

Gating-induced conformational rearrangements in the transmembrane domain of CNGA1 channels revealed by single molecule force spectroscopy

Sourav Maity[1], Mazzolini Monica[1-2], Manuel Arcangeletti[1], Alejandro Valbuena[1], Paolo Fabris[1], Marco Lazzarino[2-3] and Vincent Torre[1*]

1 - International School for Advanced Studies (SISSA) Neuroscience Area via Bonomea 265, 34136 Trieste (Italy)
2 - CBM S.c.r.l., Area Science Park, Basovizza 34149 Trieste, (Italy)
3 - IOM-CNR, Basovizza, Area Science Park, 34149 Trieste (Italy)

Corresponding author: Correspondence and requests for materials should be addressed to Prof. Vincent Torre: torre@sissa.it

We combined single molecule force spectroscopy (SMFS) with mutagenesis, bioinformatics, heterologous expression and electrophysiology to study the conformational changes of cyclic nucleotide-gated (CNG) channels induced by the binding of cyclic nucleotides. We engineered functional channels with appropriate SMFS fingerprints, which were then heterologously overexpressed in *Xenopus laevis* oocytes, and we investigated the gating of CNGA1 channels embedded in a physiological-like membrane. Force spectra in the closed state (without cGMP) show four peaks with forces of approximately 50-60 pN, but in the open state (with cGMP), three additional peaks appear. We demonstrate that during gating: i- CNGA1 channels also undergo conformational changes in the transmembrane domain; ii- the voltage sensor S4 is mechanically coupled to the transmembrane domain S5 in the open state but not in the closed state; and iii- these conformational changes are responsible for the pore widening and voltage gating observed with large permeant cations.

Atomic force microscopy (AFM) is a powerful technique used for surface imaging, measurements of sample mechanics and for the analysis of molecular interactions. Single molecule force spectroscopy (SMFS) using an AFM to apply force to unfold a molecule or a polymer[1-4]. The obtained force-distance (F-D) curves characterize the stretching of the molecule; the resulting sequence of unfolding force peaks and their magnitude allows for the identification of folded and unfolded regions, thus providing insight into the interactions between and within domains of the molecule that stabilize secondary structures[1-5]. SMFS has been used to identify the conformational changes of membrane proteins belonging to the rhodopsin family[6-8] and other proteins, such as the Na^+/H^+ antiporter, the BetP symporter, the KpOmpA transmembrane protein, the β2-adrenergic receptor, T4 lysozyme and the leucine-binding protein[9-14].

Ion channels are membrane proteins that play a major functional role, and they are grouped in superfamilies[15,16]: the superfamily of voltage-gated ion channels comprises Na^+, K^+ and Ca^{2+} channels, whose gating (transitions between the open and closed conformation) depends on the membrane voltage. This superfamily also includes cyclic nucleotide-gated (CNG) channels[17-21] that are voltage dependent[21] but are opened by the binding of cyclic nucleotides to the cyclic nucleotide-binding (CNB) domain[17,19-20].

CNG channels are opened by the binding of cyclic nucleotides. In vertebrates, seven members of the CNG channel gene family have been identified[19,22] and are grouped into two subtypes, CNGA (CNGA1-CNGA5) and CNGB (CNGB1 and CNGB3). CNGA1, CNGA2, CNGA3, and CNGA5 (but not CNGA4) can form cyclic nucleotide-activated homotetrameric channels, while CNGB1 and CNGB3 are modulatory subunits that cannot form functional homomeric channels. Hydropathicity and biochemical analyses of CNGA1 channels[17] - 690 amino acid residues (aas) long - have revealed six transmembrane α-helices (known as S1, S2, S3, S4, S5 and S6) that span the lipid bilayer (see Supplementary Fig.SI1); these helices are linked by non-spanning loops, which are either extracellular or intracellular. Ion permeation occurs through a pore region between S5 and S6, and electrophysiological experiments have identified 20 aas that form the P-helix (V348-L358) and the selectivity filter (T359-P367)[23-28]. The N- and C-terminal ends are both cytoplasmic, and the C-terminal end (N400-D690) is a large domain composed of the C-linker (N400-E482) and the CNB domain (A483-N610)[29,30]. The CNB domain shares 20% sequence identity with other cyclic nucleotide-binding proteins, such as the CNB domain of HCN channels[31] and MlotiK1 channels[32], and it consists of three α-helices and eight stranded anti-parallel β-rolls. The functional properties of

CNG channels have been investigated extensively[19,33,34], and a low-resolution architecture[35], partial crystal structures of the CNB domain[31,36,37], a crystal structure of the isolated C-terminal end from L621 to D690 (Ref.38) and a mimic of the pore[39] are available. However, the full-length channel has never been crystallized, and the conformational changes that are associated with gating are poorly understood.

In this study, we demonstrate how SMFS can be used to examine the gating of CNGA1 channels that are overexpressed in membranes from *Xenopus laevis* oocytes[21,40-41] (i.e., almost *in situ*); the plasma membrane of these oocytes contains few native membrane proteins[41-44]. We identify F-D curves using bioinformatics analysis and by engineering proteins that are composed of CNGA1 channels linked at their C-termini to a SMFS marker, i.e., a protein with a known unfolding pattern that act as a fingerprint. Our results provides three new insights on the structure of the transmembrane domain of CNGA1 channels: i- the S4 domain is mechanically coupled to S3 in the closed state but to S5 in the open state; ii - there are well defined pathways for the unfolding of the transmembrane domain; iii -the degree of folding of α-helices forming the transmembrane domain varies.

RESULTS

Several constructs were designed to identify the F-D curves obtained from the unfolding of CNGA1 channels, to explore different pulling configurations and to validate a hypothesis of the molecular mechanisms. All these constructs had cGMP-activated currents that were measured using electrophysiological experiments (Supplementary Fig.SI1). We performed SMFS experiments in the presence and the absence of cGMP, i.e., in the open and closed states of these channels. We performed SMFS experiments using both uninjected oocytes and oocytes injected with the mRNA coding for CNGA1 channels.

SMFS of the CNGA1 channels

In our SMFS experiments, the obtained F-D curves could not only represent the unfolding of the full CNGA1 channel but also the unfolding of endogenous proteins and/or of the partial unfolding of CNGA1 channels. To identify the F-D curves obtained from the unfolding of the full CNGA1

channels, we have designed a method that analyzes the F-D curves obtained using CNGA1 channels with appropriate mutations and/or channels that bear specific fingerprints.

The N- and C-termini and some loops between the transmembrane helices of the CNGA1 channels are cytoplasmic, and the AFM tip could attach to all aas in these different positions. If the tip starts the unfolding from residue D690 (i.e., from the C-terminus) - assuming that the length of a single residue is 0.4 nm (Ref.1**) -** the complete stretch corresponds to a contour length (Lc) of approximately 240 nm (from the end of the C-terminal to the beginning of S1); if the AFM tip started the unfolding from residue M1 (i.e., from the N-terminus), the complete stretch corresponds to an Lc of 180 nm (from the N-terminus to the end of S6). Therefore, we have restricted our analysis to those F-D curves that had a maximum Lc value larger than 240 nm and only approximately 1% of the F-D curves passed this filtering step. These F-D curves were too diverse to be ascribed to the unfolding of the same protein. To identify the F-D curves obtained from the unfolding of CNGA1 channels, we developed a two-step method based on the following characteristics: first, these F-D curves must be distinguishable and must be found only in SMFS experiments performed using membranes extracted from injected oocytes expressing CNGA1 channels at a high level; these F-D curves must stand out from the other F-D curves and must form a cluster of F-D curves with similar features (*identification*) (Fig.1a-e; Supplementary Fig.SI3). Second, these F-D curves are "good" candidates as F-D curves from the unfolding of CNGA1 channels, but they must be further validated by an appropriate fingerprint that is clearly visible in the F-D curves (*validation*).

The bioinformatics analysis was based on the coding of the F-D curves (Fig.1a). From the analysis of the corresponding (F,Lc) histogram (Fig.1b) (Ref.45), three different coding schemes were obtained (Fig.1c-e). Once the F-D curves were coded in appropriate strings of symbols, clustering methods developed in Computer Science were used (Supplementary 3-5). At the end of the bioinformatics analysis (Supplementary 3-5), we *identified* three major clusters of similar F-D curves that were obtained only from the membranes extracted from injected oocytes (Supplementary Fig.SI3). To *validate* these clusters as clusters obtained from the unfolding of CNGA1 channels in the closed state, we used the construct CNGA1-N2B-HisTag, and we selected those F-D curves that exhibited an N2B fingerprint (black curves in Fig.1f) that showed an initial segment of ~85 nm (in Lc) without unfolding events which are typical characteristics of the N2B

construct[46-48]. This initial segment was followed by an additional segment resembling the peaks obtained for the CNGA1 channel construct (red curves in Fig.1f).

We found the F-D curves that showed these characteristics; furthermore, when these curves were displaced by Lc of 85 nm, the resulting curves superimposed precisely with the curves present in the Cluster 1-CS (Fig.1f; Supplementary Fig.SI3a). This cluster was identified using the bioinformatics analysis and was therefore, *validated* as representing the F-D curves obtained from the C-terminus unfolding of the CNGA1 channels (Supplementary 6). These last F-D curves were used as a template to find additional F-D curves (Fig.1g) that when shifted by less than ± 5 nm aligned properly with those of Fig.1f and include also F-D curves from Cluster 2-CS (Supplementary Fig.SI3b).

Some F-D curves (violet curves in Fig.1g) ended with a force peak that had an Lc of approximately 234 nm, whereas the remaining F-D curves (red curves in Fig.1g) were longer and had an additional force peak with an Lc of approximately 276 nm. This behavior is attributed to the variability of the final detachment. We computed the corresponding value of Lc (see Methods) for each point of the F-D curves to compare the Lc histograms for the selected curves from the CNGA1 channel (Fig.1h in red) and for the curves from the construct CNGA1-N2B-HisTag (Fig.1h in black). When the Lc histogram that was obtained from the construct CNGA1-N2B-HisTag is shifted by 85 nm, the histogram of Lc maxima from the CNGA1 channels has five common peaks (Fig.1i).

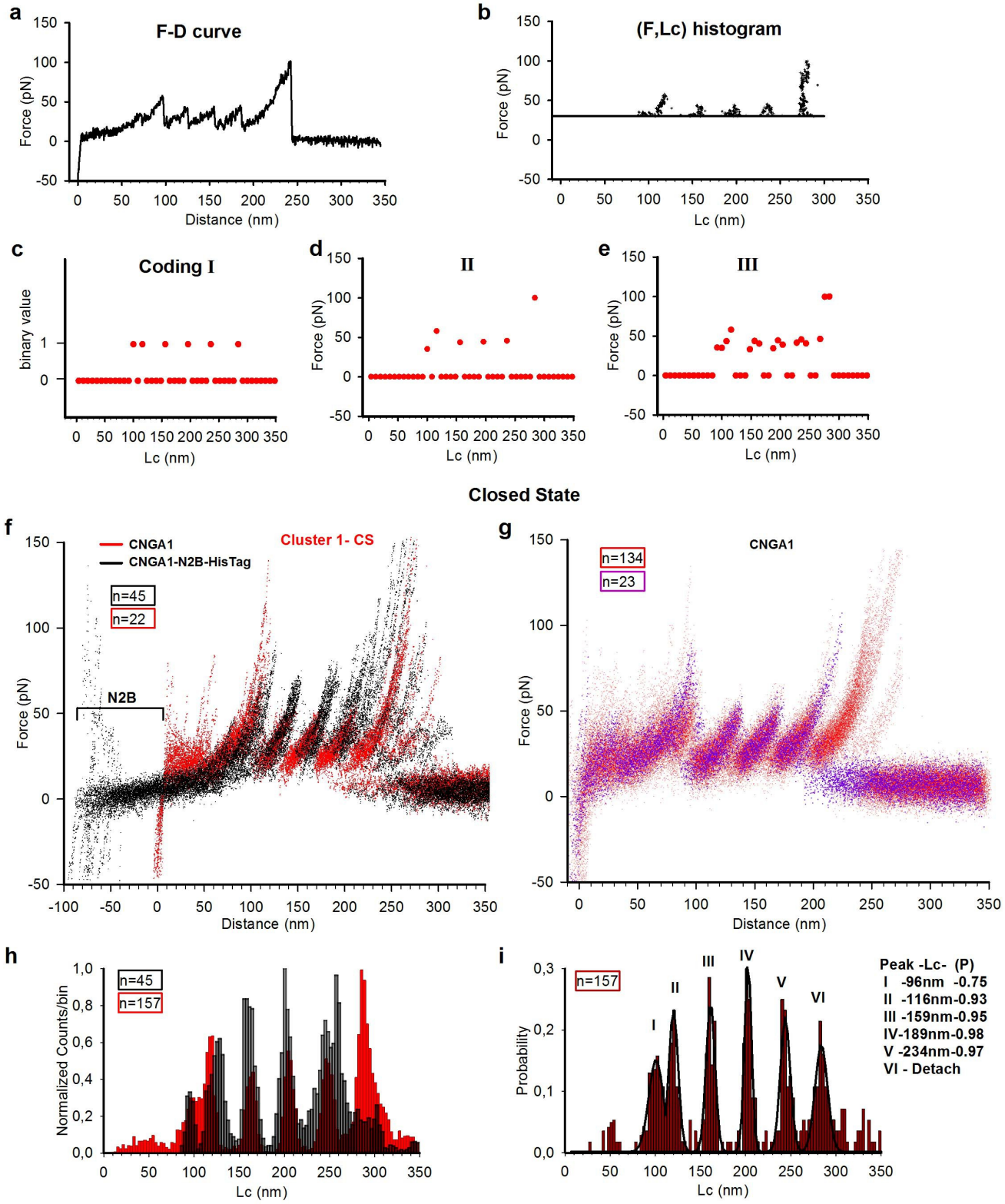


Fig.1: F-D curves from CNGA1 and CNGA1-N2B-HisTag in the closed state. a: *An example of an F-D curve.* ***b:*** *transformation of the F-D curve in (**a**) into an (F,Lc) histogram (Ref.45); if the F-D curve fits well in a piecewise manner using the WLC model, the resulting (F,Lc) histogram is composed of a series of almost vertical segments located at the corresponding value of Lc, and the upper value of each of these segments corresponds to the value of the force peak F.* ***c-e:*** *three different coding schemes of increasing complexity (see Methods) that are all based on the processing of the (F,Lc) histogram in (**b**); coding scheme I (**c**) considers only the location of the force peak (the corresponding value of Lc) and is a binary coding; coding II*

*(**d**) considers the location and the amplitude of the force peak and is not a binary coding; coding III (**e**) fits all sample values of F above 30 pN. Red dots represent the final coding of the F-D curves. **f:** 22 F-D curves from Cluster 1-CS of CNGA1 (red) detected using bioinformatics analysis (Supplementary 4-5) and 45 F-D curves from the construct CNGA1-N2B-HisTag (black) in the closed state. The construct with N2B has an initial flat region of 85 nm, followed by force peaks matching the unfolding events observed with the CNGA1 construct. **g**: superimposition of 157 F-D curves obtained from injected oocytes showing the peak force location of CNGA1. 23 F-D curves (violet) end with a force peak with an Lc of ~234 nm; 134 F-D curves (red) have an additional force peak with an Lc of ~276 nm. **h**: superimposition of histograms of normalized counts/bin against Lc (see Methods) from the 157 F-D curves of panel (**b**) (now all in red) and 45 F-D curves from the CNGA1-N2B-HisTag constructs (black).). **i:** Histogram of Lc (with Gaussian fit for the different peaks) from the F-D curves in (**g**) with 5 force peaks located at 96±3, 116±3, 159±3, 189±5, 234±6 nm and the detachment with the probability (P) of 0.75, 0.93, 0.95, 0.98 and 0.97, respectively.*

We performed SMFS in the open state (Fig.2) to determine the conformational changes that occur upon gating[19,33-34]. Clustering procedures *identified* two major groups of F-D curves that were found only from membranes extracted from injected oocytes. The last force peak of the first cluster (Cluster 1-OS) had an Lc of less than 300 nm (blue curves in Fig.2a), which was similar to that for Cluster 1-CS (Fig.1f). However, we found another cluster (Cluster 2-OS) of F-D curves in which the last force peak had an Lc that was greater than 350 nm (cyan curves in Fig.2a); these curves appeared to be the sequential unfolding of the longer protein. The use of the N2B fingerprint (green curves in Fig.2a) *validated* that Cluster 1-OS resulted from the unfolding of a single CNGA1 channel and suggested that Cluster 2-OS was obtained from the unfolding of two CNGA1 channel subunits that interact via the N-terminus of one subunit and the C-terminus of its neighbor[49,50]. The Lc histogram from the F-D curves of these two clusters superimpose very precisely up to 250 nm (Supplementary Fig.SI7a). We found other F-D curves that could be ascribed to the unfolding of CNGA1 channels in the open state (Fig.2b) when we used Cluster 1-OS as the template. Lc histograms that were obtained in both the closed and open states from the construct CNGA1-N2B-HisTag, and similar histograms that were shifted by 85 nm and obtained from the CNGA1 channels have common peaks (Fig.2c). The histogram of the Lc that was obtained from the CNGA1 F-D curves in the open state has eight peaks (Fig.2d) and was noticeably different from those obtained for the closed state (Fig.1i; Fig.2d; Supplementary Fig.SI7b): first, there is a force peak with an Lc of 84±3 nm and a mean force of 60±15 pN, which was rarely observed in the Lc histogram that was obtained from the closed state channel. Second, the force peak with an Lc of 159± 3 nm (50±12 pN) that was observed from the closed state channel is replaced by two force peaks with Lc values of 144±3 nm (mean force of 65±22 pN) and 171±3 nm (mean force of 80±18 pN). All the other force peaks that are present in the closed state are also observed in the open

state (Supplementary Fig.SI8).

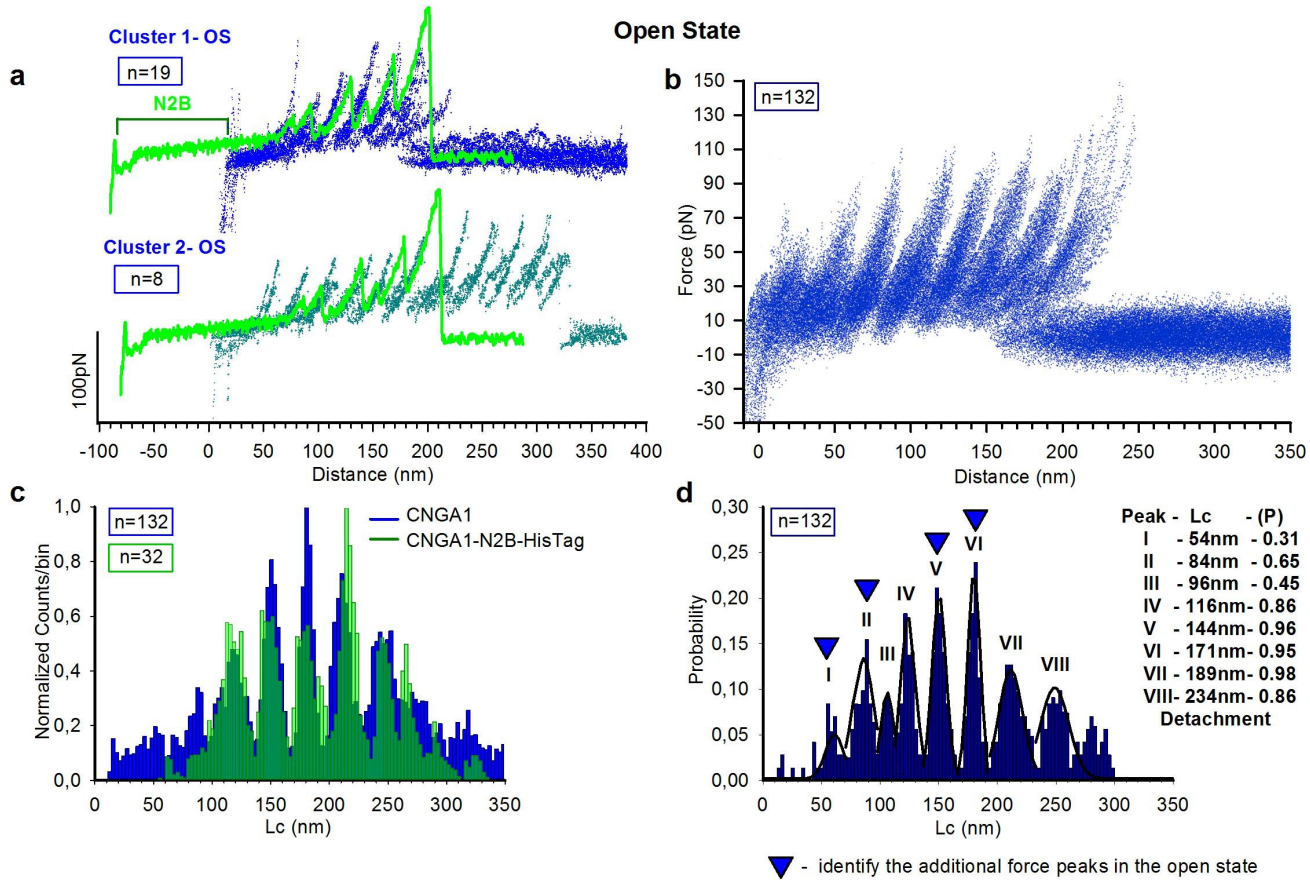


Fig.2: Unfolding of CNGA1 and CNGA1-N2B-HisTag in the open state. a: *example of F-D curves obtained from the unfolding of single CNGA1 (blue and cyan) and CNGA1-N2B-HisTag (green) constructs in the open state (19 curves for cluster 1-OS, 8 curves for cluster 2-OS, 1 curve for CNGA1-N2B-HisTag). The construct with N2B has an initial flat region of 85 nm, followed by peaks that correspond to the unfolding events that were observed in the CNGA1 construct.* ***b****: Superimposition of 132 F-D curves that were obtained from injected oocytes in the open state using the curves of Cluster 1-OS as a template.* ***c:*** *Superimposition of the histograms of normalized counts/bin against Lc from the F-D curves of panel (**b**) (blue) and 32 F-D curves from the CNGA1-N2B-HisTag (green) constructs in the open state.* ***d:*** *Histogram of Lc (with Gaussian fit for the different peaks) from the F-D curves in (**d**) with 8 peaks located at 54±3, 84±3, 99±3, 116±3, 144±3, 171±3, 189±5, 234±6 nm and the detachment; the probabilities of the unfolding are 0.31, 0.65, 0.45, 0.86, 0.96, 0.95, 0.98 and 0.86, respectively.*

For a further validation, we also performed SMFS on the CNGA1-CNGA1 tandem[51] construct (Supplementary 9) and the F380C construct[52] (Supplementary 10). The results of the experiments with the CNGA1-CNGA1 tandems show that in the open state, CNGA1 channels can be unfolded as a sequence of two linked subunits. The unique repetitive unfolding pattern of the F-D curves obtained from identical experiments also confirms that the F-D curves of Figs.1-2 represent the unfolding of the CNGA1 channel from the C-terminal end in the closed and open states, respectively. In the open state, the mutant channel F380C is known to form a disulfide bond

between the exogenous C380 and the endogenous C314 (Ref.52), and the F-D curves obtained from its unfolding are expected to be 25.4 nm shorter than those obtained from the CNGA1 channels, in agreement with the experimentally observed gap of 26±2 nm (Supplementary Fig.SI10).

The comparison of the F-D curves of Figs.1-2 also demonstrates that the unfolding pathway of CNGA1 channels is different in the open and closed states and that the unfolding of CNGA1 channels in the open state is characterized by the presence of three additional peaks.

Identification of the cytoplasmic domain

The Lc histograms in the closed (Fig.1) and open (Fig.2) states show force peaks with values of Lc ranging from 50 to 250 nm. It is important to identify the part of the F-D curves that corresponds to the unfolding of the cytoplasmic and transmembrane domains. This question can be answered by having F-D curves that represent the unfolding of *only* the cytoplasmic domain of CNGA1 channels in which the extracellular loop of S6 is strongly anchored to the substrate and the other portion of the transmembrane domain cannot be unfolded (Fig.3a-b). Therefore, we constructed the mutant channel P366C-HisTag, in which we inserted a cysteine at position P366 on the extracellular loop of S6[28]. If this exogenous cysteine forms a covalent bond with a gold substrate – with a breaking force of 1.4 nN (Ref.53) – the transmembrane domain cannot be unfolded, and only the region from D690 to P366 will be unfolded, corresponding to approximately 130 nm (Fig.3c). To identify these shorter F-D curves, we inserted a fingerprint at the C-terminal end of the mutant channel P366C; this finger was composed of two I27 modules,[1,46,54], and we performed SMFS experiments with the construct P366C-$(I27)_2$-HisTag. If the cantilever tip attaches to the HisTag and because CNGA1 unfolds with forces below 200 pN we expect[4] to unfold initially the cytoplasmic domain and S6 from D690 up to P366 and then the two I27 modules (Fig.3a-b).

We found that many of F-D curves obtained from membranes extracted from oocytes injected with the construct P366C-$(I27)_2$-HisTag exhibited the expected fingerprint (Fig.3b and black curve in Fig.3c). In the closed state, we observed a force peak with an Lc of 96±3 nm (curve I in Fig.3d). In the open state, we observed two additional force peaks with Lc values of 54±3 and 84±3 nm (curves III and IV, respectively, in Fig.3d). All these force peaks were observed in CNGA1 channels

(red and blue curves in Fig.3d) and in the construct P366C-$(I27)_2$-HisTag (black and green curves in Fig.3d). In the closed state, the F-D curves from the construct P366C-HisTag (cyan curve in Fig.3c) had two peaks with Lc values of 96 ± 3 and 116 ± 3 nm followed by a detachment. For both constructs, these features were preceded by a less frequent and lower force peak with an Lc that varied between 80 and 100 nm (Fig.3e). When the same experiments were repeated in the open state, the force peaks had similar values for both the CNGA1 channel and the construct P366C-$(I27)_2$-HisTag (Fig.3f). The F-D curves for the P366C-$(I27)_2$-HisTag and CNGA1 constructs can be closely superimposed for the portions of the curve up to the force peaks with Lc values of 116 ± 3 nm in both the closed and open states, which indicates that the cytoplasmic domain of the CNGA1 channels unfolds before the transmembrane segments S1-S6. These results show that the portion of the F-D curves with tip-sample separation (TSS) values less than 116 nm (corresponding to ~290 aas) represents the unfolding of the cytoplasmic domain comprising both the CNB domain and the C-linker (from D690 to N400). The assignment of force peaks, in this region of the F-D curves, to the unfolding of specific molecular domains is more difficult.

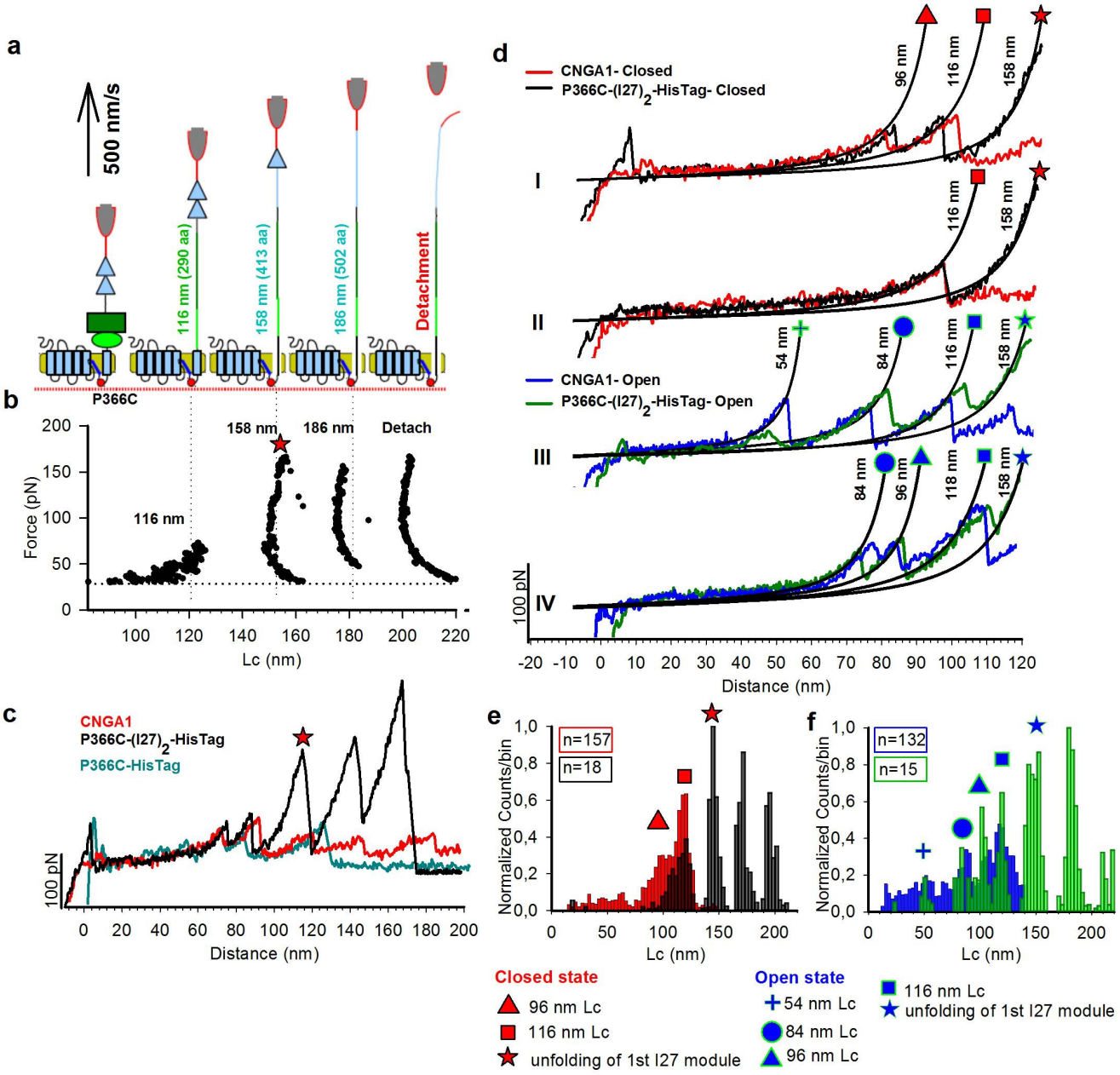


Fig.3: Unfolding of CNB domain. a*: schematic representation of the unfolding sequence, with the expected Lc values and the relative aa length in parenthesis and single module of I27 (cyan triangle), CNB domain (green rectangle), C-linker (green oval), transmembrane domains (cyan rectangles), mutation P366C (red dots).* ***b:*** *(F,Lc) plot from a typical curve of P366C-(I27)$_2$-HisTag. The finger-print is composed by two force peaks of about 200 pN separated by ~28 nm[1,46,54] with the detachment peak with a value of Lc corresponding to unfolding of the cytoplasmic domain from D690 to P366 and the two I27 modules [i.e. (D690-P366)·0.4 nm+2·28nm)] equal to 186 nm.* ***c****: superimposition of the F-D curves of CNGA1 (red), P366C-HisTag (cyan) and P366C-(I27)$_2$-HisTag (black). The F-D curves from the construct P366C-HisTag are shorter than those from CNGA1 and have a detachment event at 120-140 nm. Force peaks with Lc values at 96 and 116 nm are present in the F-D curves from the CNGA1, P366C-HisTag and P366C-(I27)$_2$-HisTag constructs.* ***d****: different unfolding pathways in the open and closed states; I and II: superposition of the F-D curves from CNGA1 channels (red) and from the construct P366C-(I27)$_2$-HisTag (black) in the closed state. In I, there are two force peaks with Lc values of 96 and 116 nm, but there is only one force peak in II; III and IV: superposition of the F-D curves from CNGA1 channels (blue) and from the constructs P366C-(I27)$_2$-HisTag (green) in the open state. In III, there are three force peaks with Lc values of 54, 84 and 116 nm, and in IV, there are*

three force peaks with Lc values of 84, 96 and 118 nm. The force peak with an Lc of 158 is only present in the construct P366C-$(I27)_2$-HisTag. ***e****: superimposition of histograms of normalized counts/bin against Lc from the F-D curves obtained from P366C-$(I27)_2$-HisTag (black) and from CNGA1 channels (red) in the closed state* ***f****: the same as in (**e**) for P366C-$(I27)_2$-HisTag (green) and from CNGA1 channels (blue) in the open state. The data in (**e-f**) are from CNGA1 channels as described in Figs.1-2.*

Conformational changes upon gating in the cytoplasmic domain

Our results indicate three conformational changes that occur in three regions upon gating in CNGA1 channels: the cytoplasmic domain, the transmembrane domain and at the C- and N-termini. The assignment of the secondary structure of CNGA1 channels corresponding to the observed force peaks in the closed and open states (Figs.1-2) is reported in Table 1.

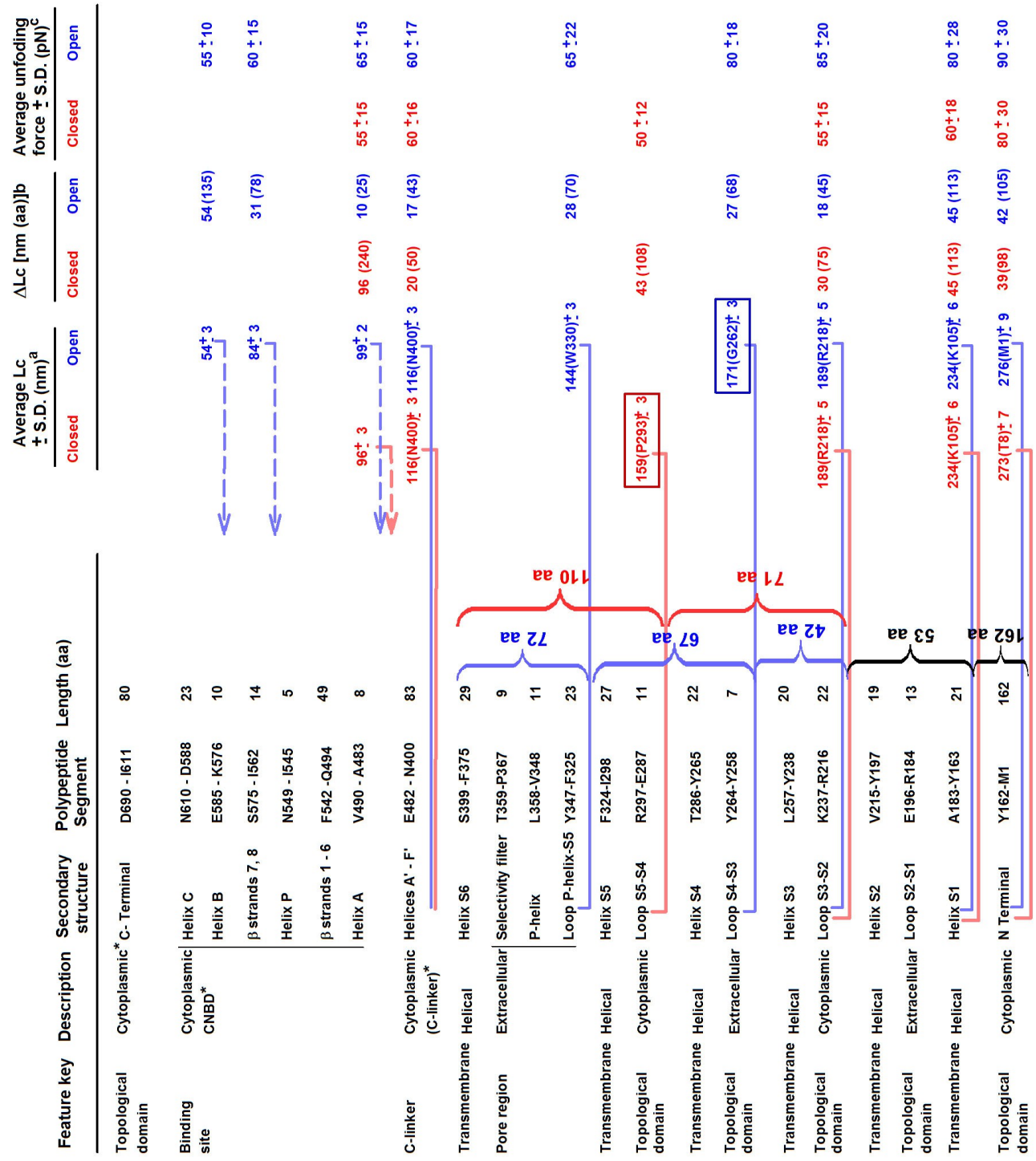


Table 1: Assignment of secondary structure and different domains of CNGA1 channels to the observed force peaks in the closed and open states *The first and second columns show the key features of the functional domains identified in CNGA1 channels according to the available literature[23-30]; the third column reports the presumed secondary structure[17,31]; the fourth and fifth columns contain the corresponding polypeptide segments and the associated aa number; the remaining columns indicate the Lc, ΔLc and force values of the corresponding force peaks in the closed and open states. The length and the polypeptide segments are obtained from the results of Figs.1-2. The number of aas for the transmembrane domains has been calculated by considering the membrane thickness (~5 nm or ~13 aa)[3,44]. The initial and final aa of each polypeptide segment are only indicative a: average Lc values of the force peaks with standard deviation (S.D.) in nm, with the number inside the brackets representing the average position of the force*

*peaks in the aa sequence. b: average ΔLc in nm and the corresponding number of aas (in brackets). c: average value of the unfolding force with the corresponding S.D. in pN. *The unfolding of the cytoplasmic segments in the C-terminus cannot be exactly determined.*

The unfolding of the cytoplasmic segment up to an Lc of 80 nm in the closed state usually requires forces below 25-35 pN, consistent with the notion that these domains do not have a well-defined conformation[30] and with their problematic crystallization[31,35-37]. In the open state, there is a peak, which appears with a probability of approximately 0.31, that has an Lc of 54±3 nm with a force of 55±10 pN; this peak is either absent or present with a very low force in the closed state. The 3D structure of the CNB domain of HCN channels has been determined in the presence of cAMP using X-ray crystallography[31,37] and in the absence using NMR spectroscopy[36]. The comparison of these structures shows that in the absence of cAMP, some α-helices are either not folded or only partially folded. Therefore, the force peak with an Lc of 54±3 nm (~135 aa) present in the open state (Fig.4) could correspond to the unfolding of α-helices in the CNB domain of CNGA1 channels, possibly the C (23 aa long) and B (10 aa long) helices, which are not folded in the closed state[30]. The force peaks with an Lc of 84±3 nm in the open state and 96±3 nm in the closed state could be the unfolding of the C-linker or of part of it; these peaks occur with a probability between 0.45 and 0.65.

Conformational changes upon gating in the transmembrane domain

The force peaks with an Lc between 120 to 250 nm appear with a probability close to 1 and correspond to the unfolding of the transmembrane domains that unfold sequentially according to the aa sequence of the protein[3,55] with force peaks correspond to unstructured regions, such as the connecting loops[11]; these force peaks can be reliably assigned to specific domains of CNGA1 channels (Fig.4a-b).

In the closed state (Fig.4a-b), three peaks with Lc equal to 159±3, 189±3 and 234±3 nm are observed requiring forces varying between 50 and 60 pN (Fig.1; Supplementary 8). The peak at 159±3 nm corresponds to the unfolding of S6-P-helix-S5 segments (S399-L301) followed by an unstructured loop (N300-N291) that causes a drop in the pulling force; the second peak at 189±3 nm corresponds to the unfolding of S4 and S3 (N291-V215), and the last peak at 234±3 nm -

before detachment - corresponds to the unfolding of S2 and S1 (R216-E100). The force peak that has an Lc of 159±3 nm for the closed state (corresponding to a residue near P293) splits into two force peaks for the open state with Lc of 144±3 (~360 aa) and 171±3 (~428 aa) nm; these peaks have higher forces of 65±22 and 80±18 pN, respectively (Fig.2; Fig.4a; Supplementary 8). This splitting reveals an important conformational change in the transmembrane domain that could be controlled by the S4-S5 linker via the helix-breaker proline P293: in the open state, the unfolding of S6-P-helix-S5 occurs in two steps. The first step consists of the unfolding of S6 and the P-helix (V348-S399); this unfolding is followed by a drop in the force at an unstructured loop (F325-Y347). In the second step, S5 is unfolded together with S4 (Y265-F324). Therefore, in the open state, S5 is mechanically coupled to S4, whereas in the closed state, S5 is more strongly connected to the P-helix and S6, as suggested by the obtained F-D curves (Fig.4b).

Homology modeling of the structure of CNGA1 channels based on the Kv1.2 channel[56,57] and the use of an improved algorithm for predicting α-helix folding suggest that the stretch of aas from Y265 to F324, corresponding to the region comprising the S4 and S5 transmembrane domains, has a good propensity to have an α-helix fold and the 3D structure is shown in Fig.4c. To validate this interpretation of the SMFS data (Fig.4a-c; Table 1), we performed electrophysiological experiments using CNGA1 mutant channels with the following rationale: if S4 in the open state is mechanically coupled by the S4-S5 linker to S5 (Fig.4c), point mutations in the S4-S5 linker are expected to propagate to the pore region and affect ionic permeation in the open state. We have identified a specific residue in the S4-S5 linker that corresponds to a conserved proline (P293 in the CNGA1 channel), which - when inserted into an α-helix - is expected to modify the ideal helical structure; therefore, we constructed the mutant channel P293A, in which P293 is substituted by the α-helix-forming alanine. It is well-known that CNGA1 channels do not inactivate[21], i.e., no time-dependent decline in the cGMP-activated current is observed, neither in the presence of small permeant cations such as Na^+ nor in the presence of larger organic cations such as ethylammonium. In the mutant channel P293A, however, the cGMP-activated current declined in the presence of ethylammonium but not in the presence of Na^+ (Fig.4d), indicating that mutations in the in the S4-S5 linker affect ionic permeation and therefore validating the notion that in the open state, the S4-S5 linker is mechanically coupled to the pore region.

The determination of the exact value of Lc for a given force peak is limited by the shift (less than

±5 nm) used to align different F-D curves (Figs.1-2). This limitation is circumvented when the increase in contour length (ΔLc) is measured. ΔLc is a structural parameter determined by the number of aa residues involved in the folded structure[58] providing also information on the kinetic barriers for the unfolding[10]. We have analysed the variability of the ΔLc between the force peaks with Lc values around 159 to 189 nm, corresponding to the unfolding of S4 and S3 in the closed state ($\Delta Lc_{closed(3/4)}$), and the variability of the ΔLc between the force peaks with Lc values around 171 to 189 nm, corresponding to the unfolding of S3 in the open state ($\Delta Lc_{open(3)}$). We have also computed the sum of ΔLc corresponding to the unfolding of all the transmembrane domains from S6 to S1 in the closed ($\Sigma\Delta Lc_{closed(TM)}$) and open ($\Sigma\Delta Lc_{open(TM)}$) state.

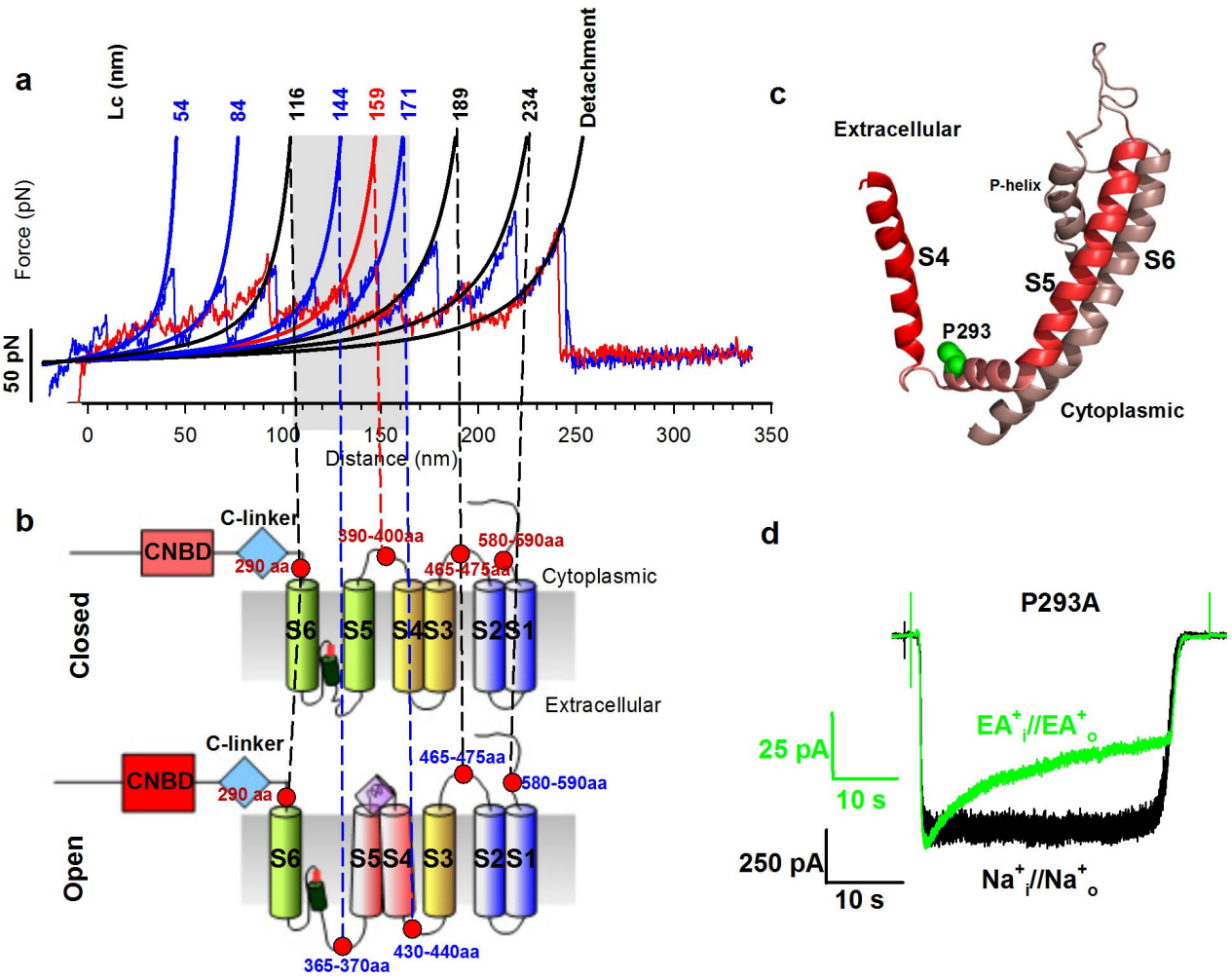


Fig.4: Conformational changes in the transmembrane domain during the gating of CNGA1 channels. a: *Superposition of two representative F-D curves in the closed (red) and open (blue) states. Continuous black lines obtained from the fitting with WLC model. Numbers indicate the corresponding values of Lc.* ***b:*** *Schematic representation of hypothesized interactions between the transmembrane helices in the closed and open states. Red dots indicate the approximate location of the force peaks and the numbers of the corresponding aas.* ***c:*** *Homology model of the S4 and S6 transmembrane domains of the CNGA1 channel based on the molecular structure of the Kv1.2 channel; the conserved P293 is indicated in green (modified*

from Ref.56). ***d:*** *Electrophysiological recordings from the P293A mutant channel in the presence of* Na^+ *(black) and ethylammonium (green) at –60 mV.*

The distribution of $\Delta Lc_{closed(3/4)}$ has three distinct peaks, corresponding to the stretching of a different number of aas (Fig.5a-c) which have a similar unfolding force (i.e. same mechanical stability). The unfolding with a ΔLc of 26 nm occurs with a low probability (14%) and corresponds to the unfolding of S3 and S4 from approximately P293 up to E230 (Fig.5d). The most probable (77%) unfolding of S3 and S4 occurs with a longer ΔLc of 34 nm from approximately P293 up to R218. In this case, residues from E230 to R218 are mechanically coupled to S3, possibly being part of the α-helix forming S3 (Fig.5e). The unfolding with a ΔLc of 50 nm occurs with a lowest probability (9%) and corresponds presumably to the unfolding of S2 and S4 from approximately P293 up to T170 (Fig.5f). In this case, S2 appears to be mechanically coupled to S4 and S3, suggesting that the connecting loop between S2 and S3 acts as a rigid handle and possibly forms a short α-helix. The distribution of $\Sigma\Delta Lc_{closed(TM)}$ has several peaks at about 113, 123 and 131 nm (Fig.5g-h) indicating a significant variability of the unfolding of the entire transmembrane domain. This variability is likely to originate from the unfolding of folded and partially folded α-helices forming the transmembrane domain[58]. Similar results were obtained when we analysed F-D curves obtained in the open state of CNGA1 channels.

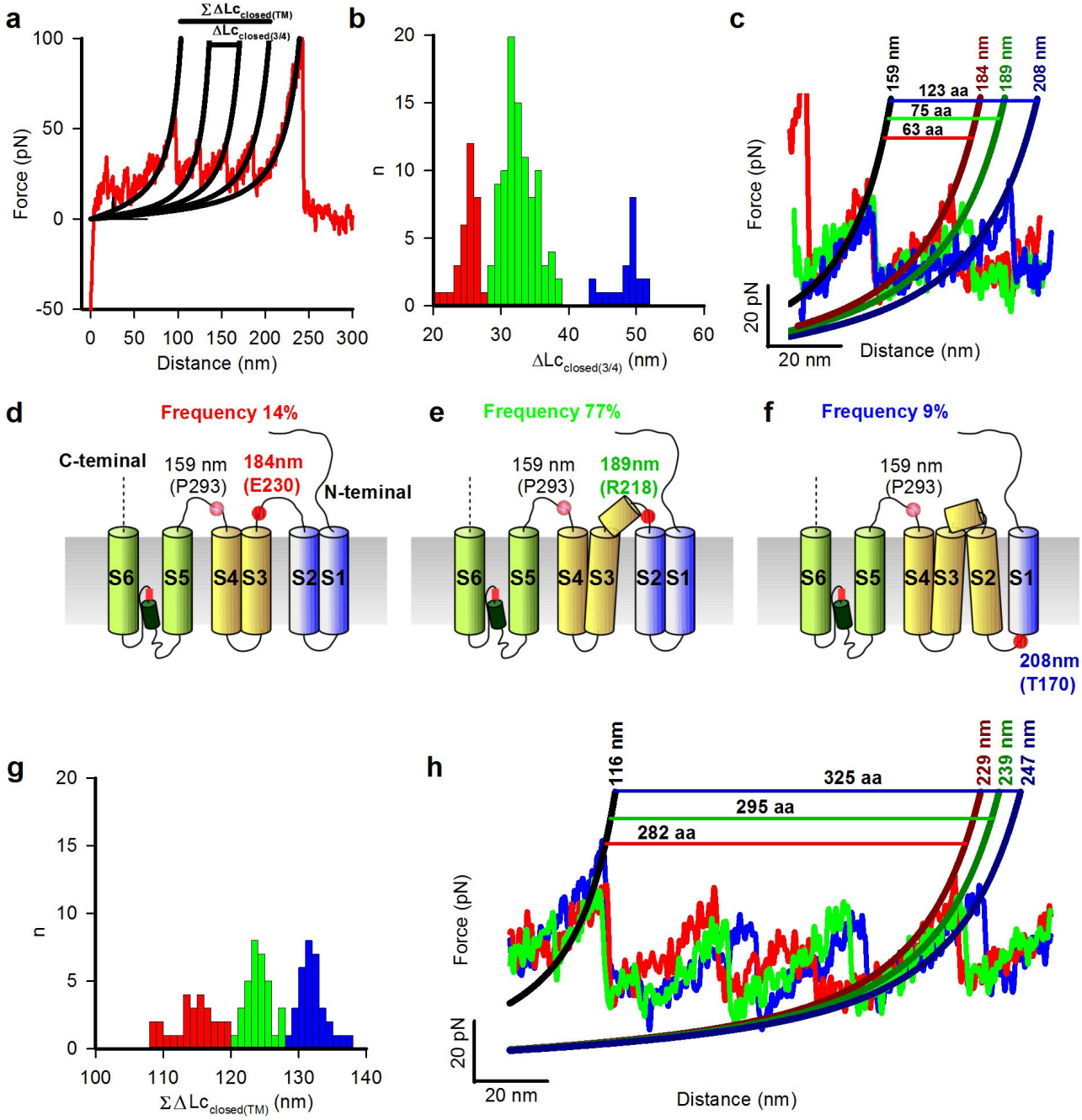


Fig.5: Variability of the unfolding of the transmembrane segments. a: *A representative F-D curve in the closed state (the same as in Fig.4a). Continuous black lines obtained from the fitting with WLC model. Horizontal lines indicate the corresponding values of ΔLc and ΣΔLc.* ***b:*** *Histogram of $\Delta Lc_{closed(3/4)}$ (bin 2 nm) in the closed state between the force peaks with Lc values around 159 to 189 nm. Colors correspond to different clusters of F-D curves with values of $\Delta Lc_{closed(3/4)}$ around 26, 34 and 50 nm.* ***c:*** *Superposition of representative F-D curves of the three clusters of panel (**b**) and the relative number of stretched aas.* ***d-f:*** *Schematic representation of the corresponding unfolding pathways. Small red dots or circles indicate the position of the force peaks, and the number is the corresponding value of Lc. The aa that presumably corresponds to the force peak is also indicated.* ***g:*** *$\Sigma\Delta Lc_{closed(TM)}$ histogram (bin 1 nm) in the closed state for the force peaks as in panel (**a**) with values around 113, 123 and 131 nm.* ***h:*** *Superposition of representative F-D curves of the three clusters of panel (**g**) and the relative number of stretched aas.*

These results (Figs.4-5) show that the transmembrane α-helices S2, S3, S4 and S5 have a variable mechanical coupling and that this coupling differs in the closed and open states.

In the open state, the force necessary to unfold the transmembrane domain varies between 65 and 85 pN, while in the close state, this force varies between 50 and 60 pN (Figs.1-2; Supplementary 8). This difference in the unfolding force was observed when CNGA1 channels were fused with the N2B fingerprint and in the CNGA1-CNGA1 tandem construct. The force necessary to unfold the transmembrane segments of other membrane proteins increases upon the binding of the appropriate ligands[13], and the observed increase in the unfolding force is comparable to the force observed in CNGA1 channels. In β2-adrenergic receptors, ligand binding is thought to change the conformational and mechanical properties of the transmembrane segments of the receptor itself[13]. The breaking of a single H-bond requires a force of approximately 4 pN (Refs. 8,59), whereas the breaking of a single hydrophobic bond requires approximately 30 pN (Refs. 8,53) and breaking a single noncovalent bond requires a larger force of approximately 160 pN (Refs. 8,53). Therefore, the observed increase in the force necessary to unfold the transmembrane domain in the open state can be accounted for by the establishment of some H-bonds and/or of one or two additional hydrophobic interactions. Another possibility is that in the open state, more residues are folded as α-helices, such as the residues in the loop connecting S4 and S5.

In the closed state, some F-D curves terminate with a force peak with an Lc of 234 nm, but other F-D curves have an additional force peak with an Lc of approximately 276 nm (Fig.1f). Nonetheless, all these F-D curves are very similar prior to the force peak with an Lc of approximately 234 nm. If the N-terminus does not interact with the C-terminus of a neighboring subunit or with other domains of the same subunit, detachment of the cantilever tip can occur simultaneously with the unfolding of S1; however, without N-terminal interactions, an additional pull is required for the final detachment. Some of the F-D curves obtained in the open state were longer (Fig.2a-b), and the force peaks appeared to be replicated, suggesting that these curves do not correspond to the unfolding of a single CNGA1 subunit but instead represent the unfolding of two interacting subunits. Indeed, in the open state, the N-terminus of one subunit strongly interacts with the C-terminus of a neighboring subunit[49,50]; thus, two neighboring subunits are almost linked together. Under these conditions, when the cantilever tip unfolds one CNGA1 subunit, an additional subunit can subsequently be unfolded. Therefore, our results indicate a possible interaction between the N- and C-termini in the closed state, and this interaction is clearly potentiated in the open state. In the open

state, we never observed “duplicated” F-D curves when pulling with the constructs CNGA1-N2B-HisTag because in these constructs, the added domains do not allow the N- and C-termini to interact as they do in the CNGA1 channels. The interaction between the C- and N-termini in the open state can explain also the lower success of SMFS at obtaining good and complete F-D curves (Fig.4) because the cantilever tip has limited access to the C-terminus.

DISCUSSION

We have performed SMFS experiments to recover structural information about CNGA1 channels in a physiological-like environment, thus avoiding purification and reconstitution into lipid bilayers. F-D curves obtained from the unfolding of CNGA1 channels expressed in *Xenopus laevis* oocytes were identified using bioinformatics analysis validated by the N2B and I27 fingerprints and were corroborated by experiments with CNGA1-CNGA1 tandem constructs and with the mutant channel F380C. These F-D curves representing the unfolding of CNGA1 can be used as a template to identify the F-D curves obtained from native CNG channels from rod outer segments (Maity et al., in preparation). SMFS experiments can detect changes in the interactions of proteins that are not necessarily associated with changes in the secondary structure[60,61], and for proper interpretation of the changes in the portion of the F-D curves (Fig.3) corresponding to the cytoplasmic domain, we have used additional experimental insight from NMR and X-ray approaches (Table 1).

Our SMFS experiments provide new insights on the function of CNGA1 channels and show that S4 in the closed state is mechanically coupled to S3, but in the open state is mechanically coupled to S5; this conclusion is confirmed by electrophysiological experiments (Fig.4d) and leads to a better understanding of the gating in CNG channels. P293 in the S4-S5 linker and G262 in the loop connecting S3 and S4 could be the hinges regulating the mechanical coupling between these α-helices.

Our SMFS experiments also reveal a significant variability in the unfolding of the transmembrane domains (Fig.5). This variability can originate from the existence of multiple unfolding pathways[10,14,58] and/or from the existence of different initial conformations[10,14,58]. The variability of $\Delta Lc_{closed(3/4)}$ (Fig.5) indicates the existence of multiple unfolding pathways associated to a ragged energy landscape[8,14,67]. These observations are in agreement with the notion that the mechanical

coupling between transmembrane α-helices varies in different CNGA1 channels and that the connecting loops can have a different degree of mechanical rigidity, presumably originating from their highly dynamic conformation.

The variability of $\Sigma\Delta Lc_{(TM)}$ (see Fig.5) can be explained if these α-helices are folded but also partially folded so that the unfolding of part of these α-helices requires a very small force. These observations suggest that the transmembrane α-helices of CNGA1 channels have a variable folding and do not have a fixed mechanical stability[10,14,58].

Which is the possible molecular mechanism underlying the gating of CNGA1 channels? The results presented here and previous experimental observations suggest a more accurate view of the molecular mechanism underlying gating. Upon the binding of cGMP to the CNB domain that is almost located at the C-terminal end of CNGA1 channels, several conformational changes occur in the cytoplasmic domain, the transmembrane domain and the C- and N-termini. The structural information obtained in the present investigation and in similar ion channels, such as the K, HCN and MlotK channels[31,32], provides a better picture of the gating in CNG channels. Upon the binding of cGMP, the entire cytoplasmic domain acquires a more structured conformation[30], either as a dimer of dimers[30,35,62] or as a tetramer[63]. Similarly to the observations in MlotK channels[64], the cytoplasmic domain of CNGA1 could move vertically towards the membrane and could induce rotations, vertical shifts and tilts of the S5 and S6 transmembrane domains[24,27], leading to widening of the filter region, where the gate is located[26,28].

The S4-S5 linker of CNGA1 channels is composed of 11 aas, in which the α-helix breaker P293 is flanked on the right by predominantly hydrophobic residues and flanked on the left by predominantly hydrophilic residues[57], and the segment TNYP has a low propensity to be in an α-helical conformation[56]. Electrophysiological experiments with chimeric channels show that the C-linker interacts with the S4-S5 linker[57]. Therefore, we propose that in the closed state, the cytoplasmic end of S4 is unfolded, but in the open state it becomes an α-helix due to its interaction with the C-linker; this change increases the mechanical coupling between S5 and the voltage sensor in S4. As a consequence, CNGA1 channels acquire voltage gating in the open state, as recently shown when large cations such as Cs^+, methylammonium and dimethylammonium are the permeant ions[21]. A similar interaction between the cytoplasmic domain and the S4-S5 linker has been observed in Kv1.2 and hERG channels[65].

In summary, three new insights on CNGA1 channels emerge from this analysis: i- the S4 domain is mechanically coupled to S3 in the closed state but to S5 in the open state (Fig.4); ii - there are well defined pathways for the unfolding of the transmembrane domain (Fig.5); iii -the degree of folding of α-helices forming the transmembrane domain varies (Fig.5) possibly also assuming a 3_{10} helix conformation[66,67]. These experimental observations, obtained from SMFS, show that the transmembrane domain has a dynamical structure. Moreover, our results show that SMFS is a powerful tool for analyzing ion channels, suggesting that the same approach combining SMFS with informatics, mutagenesis and fusion with known fingerprints can be used to study other ion channels and membrane proteins and to detect their conformational changes at the single molecule level and in a physiological-like environment.

Acknowledgments: We thank Dr. Loredana Casalis for providing us the gold surface and Dr. Michael Pusch for reading the manuscript. We thank Dr. Mariano Carrión-Vázquez for providing proteins for controls and sequences for designing protein construct. We thank M. Lough for checking the English. This work was supported by FOCUS Contract no. 270483 (FP7-ICT-2009-6) from the EU.

Author Contributions: S.M performed SMFS experiments, data analysis and prepared the figures ; M.M. and M.A. performed electrophysiology and mutagenesis; P.F., S.M. and V.T performed bioinformatics; S.M., M.M., A.V. and V.T. designed the experiments; S.M., M.M. and V.T. participated in interpreting results; S.M., M.M., M.L. and V.T. contributed to writing the paper; V.T. supervised the project.

Competing financial interests: The authors declare no competing financial interests.

METHODS

Molecular biology and CNGA1 constructs with different fingerprints

Eight different channel constructs (see Fig.SI1) were cloned into the expression vector known as pGEM-HE. The constructs are the CNGA1 channel cloned from bovine retinal rod photoreceptors, as previously described[17], that consists of 690 aa (here named the CNGA1 channel), the CNGA1 channel with a tag that is composed of six histidines (HisTag) at the C-terminus (CNGA1-HisTag), the CNGA1 with an N2B module (CNGA1-N2B-HisTag)

and the CNGA1 with two I27 modules followed by the same tag [CNGA1-$(I27)_2$-HisTag]. The HisTag was used to increase the probability of a specific attachment between the protein and the functionalized cantilever tip. We used two fingerprints, the N2B module (the shortest titin isoform)[46-48] and two I27 modules (Ig module 27 of the I band of titin)[46,54], to differentiate the F-D curves that were obtained from the unfolding of the CNGA1 channels from those of native membrane proteins that were present in the oocyte plasma membrane. The N2B module is composed of 210 aas with the mechanical properties of a random coil unfolding that does not show any unfolding events[47,48]. The unfolding of the I27 modules (each 89 aa long) shows the characteristic sawtooth pattern, in which successive force peaks have an increase in contour length (ΔLc) of ~28 nm and amplitudes of approximately 200 pN. The CNGA1-CNGA1 tandem[51] construct was obtained using two identical CNGA1 subunits linked by a short 10 aa linker (GSGGTELGST) between the C-terminal end of the first subunit and the N-terminal end of the second subunit. We performed a point mutation in an exposed loop of the channel (P366C) [two constructs: P366C-HisTag and P366C-$(I27)_2$-HisTag] that is able to bind the gold surfaces used in our experiments and two mutant channels F380C and P293A. Single-residue mutagenesis was performed as described using the Quick Change Site-Directed Mutagenesis kit (Stratagene). Point mutations and cloning were confirmed by sequencing, using the sequencer LI-COR (4000 L). The constructs CNGA1-N2B-HisTag and CNGA1-$(I27)_2$-HisTag were subcloned by DNA2.0 (Menlo Park, CA). cDNAs were linearized and were transcribed to RNA *in vitro* using the mMessage mMachine kit (Ambion, Austin, TX).

Heterologous expression system and sample preparation

Purified RNA of the different constructs was injected into *Xenopus laevis* oocytes ("Xenopus express" Ancienne Ecole de Vernassal, Le Bourg 43270, Vernassal, Haute Loire, France). Oocytes were prepared as described[30]. Injected eggs were maintained at 18°C in a Barth solution that was supplemented with 50 mg/ml of gentamycin sulfate and contained the following (in mM): 88 NaCl, 1 KCl, 0.82 $MgSO_4$, 0.33 $Ca(NO_3)_2$, 0.41 $CaCl_2$, 2.4 $NaHCO_3$, and 5 TRIS-HCl (pH 7.4 buffered with NaOH). The usual salts and reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA).

The vitelline membrane of oocytes was removed mechanically. These oocytes were incubated on a freshly cleaved mica sheet (or on a gold surface) for 1-30 minutes in Standard solution (110 NaCl, 10 HEPES and 0.2 EDTA; pH 7.4 buffered with NaOH) with or without 2 mM cGMP. We removed

the cytoplasmic content (yolk and granules) using 5-10 washings with the Standard solution after the membrane had been attached to the surface.

Electrophysiological recordings

The functionality of all constructs was verified by exposure to 2 mM cGMP (saturating concentration) and then recording the channel opening using electrophysiological measurement in the excised patch configuration (see Fig.SI1a-g). cGMP-gated currents in a voltage-clamp condition were recorded using a patch-clamp amplifier (Axopatch 200, Axon Instruments Inc., Foster City, CA, USA) 2-6 days after RNA injection at room temperature (20-24°C) using borosilicate glass pipettes with resistances of 2-5 MΩ. The perfusion system allowed a complete solution change in less than 1 s. During the experiments, oocytes were kept in Ringer's solution containing the following (in mM): 110 NaCl, 2.5 KCl, 1 $CaCl_2$, 1.6 $MgCl_2$, and 10 HEPES-NaOH (pH 7.4 buffered with NaOH). The Standard solution on both sides of the membrane consisted of (in mM) 110 NaCl, 10 HEPES and 0.2 EDTA (pH 7.4 buffered with NaOH). We used Clampex 10.0, Clampfit 10.1, and SigmaPlot 9.0 for data acquisition and analysis. Data are usually given as the mean ± s.e.m. We attempted SMFS only in oocytes in which the measured cGMP-activated current was larger than 1 nA at ±100 mV.

Atomic force microscopy and cantilever functionalization

The NanoWizard 3 AFM system (JPK) and an inverted optical microscope (Olympus IX71) were used under liquid conditions in Standard solution (110 mM NaCl, 10 mM HEPES and 0.2 mM EDTA pH 7.4 buffered with NaOH) with or without 2 mM cGMP. Rectangular silicon nitrite gold-coated cantilevers (HYDRA2R-50NGG from APPNANO) were functionalized and were used to localize plasma membrane patches and to perform SMFS experiments. For imaging of the membrane patch, the AFM system was operated in liquid using the tapping mode[44] with ~14 kHz as the operating frequency. The cantilever spring constant was ~ 0.08 N/m and was calculated before the start of each experiment by using the equipartition theorem. A 0.4 NA/10X objective was used to localize the area of oocyte incubation. The scan rate for AFM imaging was kept between 1 and 0.5 Hz, depending on the image size (from 20 to 0.5 μm) and on the condition of each sample. AFM images were acquired with a resolution of 512 pixels. Cantilever tip functionalization was based on

the thiol and nitrilotriacetic acid-Ni^{2+} specificity for gold and the HisTag, respectively[68,69]. Tips were first cleaned in ethanol for 15 minutes, dried under an N_2 flow, and exposed to UV light for 15 minutes. The tips were further incubated for 5 minutes in chloroform and dried again under an N_2 flow. These three steps were repeated one more time to obtain cleaner tips. Cantilevers were then incubated for 30 minutes in 10 µM dithiobis-C_2-NTA (Dojingo Technologies, Japan), washed with ethanol and dried in an N_2 flow. The tips were then incubated for 20 minutes in 100 µM $NiSO_4$ and rinsed with MilliQ water before being dried in a N_2 flow. Finally, the functionalized tips were incubated for 20 minutes in 10 mM 6-mercapto-1-hexanol to avoid nonspecific adsorption and were dried under an N_2 flow. Salts and reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA). To check the cantilever functionalization and activity of NTA, a control experiment was performed for each set of functionalized cantilevers. First, we performed an SMFS experiment on a polypeptide chain composed of 8 modules of I27. The pulling efficiency was almost 10%, but in presence of 50, 100 and 200 mM imidazole, the efficiency decreased to 1%, 0.5% and 0%, respectively (result not shown).

SMFS experiments and data processing

SMFS experiments were performed using membrane extracted from injected oocytes expressing constructs at a high level (about 1000-5000 channels/µm^2) and in membrane extracted from uninjected oocytes as control. Oocytes were attached to a mica substrate or a gold surface (in case of cysteine mutant), and clean fragments of membrane remained anchored to the substrate with the intracellular side exposed to the bathing medium and to the cantilever tip of the AFM. We used the AFM in liquid and in tapping mode[44] to image the membrane patches. We observed membrane patches that had a high or a low protrusion density in their surface patterns (Supplementary Fig.SI1h-i). In both cases, as would be expected from a membrane patch[44], structures emerging from the mica had a height of 4-6 nm. Membrane patches with a high protrusion level had additional peaks that were 2-3 nm in height (Supplementary Fig.SI1i) and were taken as suitable samples for SMFS experiments.

After the localization of the membrane patch we moved the AFM tip over the imaged area (usually 1-2 µm²). Using the matrix scanning mode, the AFM tip was pushed into the surface with a contact force of 1 nN for 0.5 s, to give the protein a chance to adsorb on the stylus, and then retracted

with a constant speed of 500 nm/s while the force exerted between the tip and surface was recorded. In approximately 20% of the cases, the tip was able to absorb a molecule, providing a sawtooth-like F-D curve, and if the magnitude of the force of these F-D curves was larger than 45 pN, the curve was saved. In this manner, during an experimental session lasting up to 8-10 hours, we collected a total of approximately 10.000 F-D curves and approximately 200.000 and 300.000 F-D curves from membranes extracted from uninjected (control) oocytes in the presence and absence of cGMP, respectively; we also collected more than 500.000 F-D curves from membranes extracted from oocytes injected with mRNA for the CNGA1 channels. Approximately 30% of these F-D curves had only nonspecific adhesion events, and the remaining 70% showed very diverse unfolding pathways, which could originate from the unfolding of endogenous proteins or the partial unfolding of CNGA1 channels. Therefore, we had to identify the F-D curves from the unfolding of the full CNGA1 channel, and we had to distinguish these curves from all other unfolding events. We set up an initial filtering to remove the F-D curves showing only nonspecific adhesions and all other invalid F-D curves (Supplementary 2). Then, the remaining F-D curves were fitted to the worm-like chain (WLC) model[1] with a persistence length (Lp) of 0.4 nm, and the corresponding contour length (Lc) was calculated using 0.4 nm as the length of a single amino acid (aa). All data points for Lc were summarized in histograms and fitted using Gaussian model. In text and in the figures, maxima of the Gaussian fittings are expressed as the mean ± standard deviation (SD).

Bioinformatics analysis

We performed bioinformatics to identify F-D curves that were present in only the SMFS experiments that used membranes extracted from injected oocytes. The method had two steps. First, each F-D curve was mapped to a sequence of symbols that represented the location and amplitude of the force peaks (coding), and then these sequences were assembled in groups with similar properties (clustering). The coding step was based on the transformation of the F-D curves into a histogram of force and contour lengths (F,Lc) (Ref.45) (Supplementary 4). We used the worm-like chain (WLC) model where Lp is the persistence length (equal to 0.4 nm), a parameter that represents the stiffness of the molecule.

For each tip-sample separation (D or TSS) value, the WLC model is used to compute the

corresponding value of Lc that is obtained by solving the third order polynomial: $4\lambda^3 + \omega\,\lambda^2 - 1 = 0$, where $\lambda = 1 - D/Lc$ and $\omega = 4F(x,Lc)/\alpha - 3$ $\alpha = k_bT/Lp$. This equation has three roots, and the root of interest is the real root $\lambda*$ such that $0 < \lambda* < 1$. In this manner, each point of the F-D curve (F,D) (e.g., the curve in Fig.1a) is transformed into a corresponding point (F,Lc), and each F-D curve is transformed into an (F,Lc) histogram (see Fig.1b). The three roots of the equation were obtained using a MATLAB routine. Due to this transformation, each portion of the F-D curve that is fitted perfectly by the WLC model is mapped to a perfect vertical segment. The transformation of an F-D curve is therefore a histogram and not a function (or curve) in the (F,Lc) plane.

Histogram of normalized counts of Lc/bin: given a set of F-D curves, we computed the histogram of the normalized counts/bin of Lc values in the following manner:

- the Lc axis of the (F,Lc) profile is divided into bins (from 1 to 10 nm);
- all points with a value of F larger than 30 pN are counted in the corresponding bin and summed over all sets of the F-D curves; and
-the final histogram of counts/bin is normalized so that its maximal value is equal to 1.

This histogram is used to quantify the occurrence of points in the F-D curves that correspond to a given value of Lc.

Histogram of maxima of Lc: given an (F,Lc) plot, it is possible to extract an (F,Lc) profile and to compute the local maxima in the following manner:

- the Lc axis is divided into fixed intervals of 1-10 nm (typically 5 nm);
- we extract the maximum value of F in each interval to obtain the corresponding (F,Lc) profile;
- if all values of F in an interval are below 30 pN, the value of the force in that interval is set equal to 0 pN; and
- the local maxima of the (F,Lc) profile are computed.

From a set of F-D curves, we compute the histogram of the local maxima of the corresponding (F,Lc) profile, and in each bin (the bin size for this computation can be different from the bin size that was used in the (F,Lc) profile to detect local maxima), we sum the number of local maxima. The final histogram is then divided by the total number of F-D curves that were considered. The histogram of the local maxima shows the probability of obtaining F-D curves that have a force peak with a given value of Lc. Three different, increasingly complex coding schemes were considered. The simplest coding scheme (Coding I) only codes the location of force peaks and only considers the value of Lc neglecting the corresponding value of F. If the F-D curve has n force peaks with values of Lc and F equal to Lc_i and F_i for i=1,..n and having selected a bin width ΔLc

(varying between 1 and 10 nm) each F-D curve is converted into a sequence of symbols of the type (0,0,1,0,0,0,1,...). All symbols are set to 0 with the exception of some values of 1 located at position k corresponding to the integer value of $Lc_i/\ \Delta Lc$ (see red circles in Fig.1c). This coding scheme is a binary code and was considered for its simplicity. For the next scheme, given a bin width ΔLc, such as 5 or 10 nm, each F-D curve is converted into a sequence of symbols (0,0,F1,0,0,0,F2,...), where Fi is the value of the k_{th} force peak, and if Lck is the corresponding value of Lc, the symbol Fk is located at the position that corresponds to the integer value of $Lck/\Delta Lc$ (Coding II and red circles in Supplementary Fig.1d). A more complex coding scheme (Coding III) considers all points in the F,Lc histograms with values of F larger than 30 pN; for each interval between $k\Delta Lc$ and $(k+1)\ \Delta Lc$, this scheme selects the maximum value of the histogram F,Lc in that interval (see red circles in Fig.1e). In this case, the F-D curve is coded in the sequence of symbols $s_1,s_,{}_2,s_,{}_3,,..s_n$, where s_k is the maximal value of F in the interval between $k\Delta Lc$ and $(k+1)\Delta Lc$. Once the F-D curves are transformed into symbol sequences, a distance $D(x_i,x_j)$ between the two sequences x_i and x_j must be defined.

We then applied different clustering procedures with the rationale that clusters containing F-D curves that only originate from membranes extracted from injected oocytes are “good” candidates to represent the unfolding of CNGA1 channels. For the clustering step (Supplementary 5), we consider all F-D curves with a maximal value of Lc larger than 220 nm that are obtained from membranes extracted from injected and uninjected oocytes, and we look for a cluster of similar F-D curves that are *only* obtained from injected oocytes. We used clustering methods developed in Computer Science to identify objects or patterns with similar features. At the end of the informatics analysis, we *identified* three major clusters (Supplementary Fig.SI3f-h and Supplementary 5) of similar F-D curves that were obtained only from membranes extracted from injected oocytes. Other clusters were nondiscriminatory and were composed of F-D curves obtained from membranes extracted from injected and uninjected oocytes (Supplementary Fig.SI3i-k and Supplementary 5).

REFERENCES

1- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-1112 (1997).

2- Engel, A. & Gaub, H.E. Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* **77,** 127-148 (2008).

3- Müller, D.J., Wu, N. & Palczewski, K. Vertebrate membrane proteins: structure, function, and insights from biophysical approaches. *Pharmacol. Rev.* **60,** 43-78 (2008).

4- Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.* **41,** 4781-4796 (2012).

5- Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl. Acad. Sci. USA* **106,** 13791-13796 (2009).

6- Oesterhelt, F. *et al.* Unfolding pathways of individual bacteriorhodopsins. *Science* **288,** 143-146 (2000).

7- Sapra, T.K. et al. Detecting molecular interactions that stabilize native bovine rhodopsin. *J. Mol. Biol.* **358,** 255-69 (2006).

8- Kawamura, S. *et al.* Kinetic, energetic, and mechanical differences between dark-state rhodopsin and opsin. *Structure* **21,** 426-37 (2013).

9- Kedrov, A., Ziegler, C. & Muller, D.J. Differentiating ligand and inhibitor interactions of a single antiporter. *J. Mol Biol*. **362,** 925-932 (2006).

10- Peng, Q. & Li, H. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. USA*. **105,** 1885-1890 (2008).

11- Ge, L., Perez, C., Waclawska, I., Ziegler, C. & Muller, D.J. Locating an extracellular K+-dependent interaction site that modulates betaine-binding of the Na+-coupled betaine symporter BetP. *Proc. Natl. Acad. Sci. USA* **108,** E890-898 (2011).

12- Bosshart, P.D. *et al.* The transmembrane protein KpOmpA anchoring the outer membrane of Klebsiella pneumoniae unfolds and refolds in response to tensile load. *Structure* **20,** 121-127 (2012).

13- Zocher, M., Zhang, C., Rasmussen, S.G., Kobilka, B.K. & Müller, D,J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor. *Proc. Natl. Acad. Sci. USA*. **109,** E3463-472 (2012).

14- Kotamarthi, H.C., Sharma, R., Narayan, S., Ray, S. & Ainavarapu, S.R. Multiple unfolding pathways of leucine binding protein (LBP) probed by single-molecule force spectroscopy (SMFS). *J. Am. Chem. Soc.* **135,** 14768-14774 (2013).

15- Jan, L.Y. & Jan, Y.N. A superfamily of ion channels. *Nature* **345,** 672 (1990).

16- Yu, F.H., Yarov-Yarovoy, V., Gutman, G.A. & Catterall, W.A. Overview of molecular relationships in the voltage-gated ion channel superfamily. *Pharmacol. Rev.* **57,** 387–395 (2005).

17- Kaupp, U.B. *et al.* Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel. *Nature* **342,** 762-766 (1989).

18- Anderson, P.A. & Greenberg, R.M. Phylogeny of ion channels: clues to structure and function. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* **129,** 17-28 (2001).

19- Kaupp, U.B. & Seifert, R. Cyclic nucleotide-gated ion channels. *Physiol. Rev.* **82,** 769-824 (2002).

20- Matulef, K. & Zagotta, W.N. Multimerization of the ligand binding domains of cyclic nucleotide-gated channels. *Neuron* 2002 **36,** 93-103 (2002).

21- Marchesi, A., Mazzolini, M. & Torre, V. Gating of cyclic nucleotide-gated channels is voltage dependent. *Nat. Commun.* **3,** 973 (2012). doi: 10.1038/ncomms1972.

22- Tetreault, M.L., Henry, D., Horrigan, D.M., Matthews, G. & Zimmerman, A.L. Characterization of a novel cyclic nucleotide-gated channel from zebrafish brain. *Biochem Biophys Res Commun.* **348,** 441-449 (2006).

23- Becchetti, A., Gamel, K. & Torre, V. Cyclic nucleotide-gated channels. Pore topology studied through the accessibility of reporter cysteines. *J. Gen. Physiol.* **114,** 377-392 (1999).

24- Flynn, G.E. & Zagotta, W,N. Conformational changes in S6 coupled to the opening of cyclic nucleotide-gated channels. *Neuron* **30,** 689-698 (2001).

25- Flynn, G.E. & Zagotta, W.N. A cysteine scan of the inner vestibule of cyclic nucleotide-gated channels reveals architecture and rearrangement of the pore. *J. Gen. Physiol.* **121,** 563-582 (2003).

26- Contreras, J.E., Srikumar, D. & Holmgren, M. Gating at the selectivity filter in cyclic nucleotide-gated channels. *Proc. Natl. Acad. Sci. USA* **105,** 3310-3314 (2008).

27- Nair, A.V., Nguyen, C.H. & Mazzolini, M. Conformational rearrangements in the S6 domain and C-linker during gating in CNGA1 channels. *Eur. Biophys. J.* **38,** 993-1002 (2009).

28-Mazzolini, M., Anselmi, C. & Torre, V. The analysis of desensitizing CNGA1 channels reveals molecular interactions essential for normal gating. *J. Gen. Physiol.* **133,** 375-86 (2009).

29- Nair, A.V., Anselmi, C. & Mazzolini, M. Movements of native C505 during channel gating in CNGA1 channels. *Eur. Biophys. J.* **38,** 465-478 (2009).

30- Mazzolini, M., Punta, M. & Torre, V. Movement of the C-helix during the gating of cyclic nucleotide-gated channels. *Biophys. J.* **83,** 3283-3295 (2002).

31- Zagotta, W.N. *et al.* Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425,** 200-205 (2003).

32- Mari, S.A. *et al.* Gating of the MlotiK1 potassium channel involves large rearrangements of the cyclic nucleotide-binding domains. *Proc. Natl. Acad. Sci. USA* **108,** 20802-20807 (2011).

33- Craven, K.B. & Zagotta, W.N. CNG and HCN channels: two peas, one pod. *Annu. Rev. Physiol.* **68,** 375-401 (2006).

34- Mazzolini, M., Marchesi, A., Giorgetti, A. & Torre, V. Gating in CNGA1 channels. *Pflugers. Arch.* **459,** 547-555 (2010).

35- Higgins, M.K., Weitz, D., Warne, T., Schertler, G.F. & Kaupp, U.B. Molecular architecture of a retinal cGMP-gated channel: the arrangement of the cytoplasmic domains. *EMBO J.* **21,** 2087-2094 (2002).

36- Schünke, S., Stoldt, M., Lecher, J., Kaupp, U.B. & Willbold, D. Structural insights into conformational changes of a cyclic nucleotide-binding domain in solution from Mesorhizobium loti K1 channel. *Proc. Natl. Acad. Sci. USA* **108,** 6121-6126 (2011).

37- Lolicato, M. *et al.* Tetramerization dynamics of C-terminal domain underlies isoform-specific cAMP gating in hyperpolarization-activated cyclic nucleotide-gated channels. *J. Biol. Chem.* **286,** 44811-44820 (2011).

38- Shuart, N.G., Haitin, Y., Camp, S.S., Black, K.D. & Zagotta, W.N. Molecular mechanism for 3:1 subunit stoichiometry of rod cyclic nucleotide-gated ion channels. *Nat. Commun.* **2,** 457 (2011). doi: 10.1038/ncomms1466.

39- Derebe, M.G., Zeng, W., Li, Y., Alam, A. & Jiang, Y. Structural studies of ion permeation and Ca2+ blockage of a bacterial channel mimicking the cyclic nucleotide-gated channel pore. *Proc. Natl. Acad. Sci. USA* **108,** 592-597 (2011).

40- Gorostiza, P. *et al.* Molecular handles for the mechanical manipulation of single-membrane proteins in living cells. *IEEE Trans Nanobioscience* **4,** 269-276 (2005).

41- Santacroce, M. *et al.* Imaging of Xenopus laevis oocyte plasma membrane in physiological-like conditions by atomic force microscopy. *Microsc Microanal* **19,** 1358-1363 (2013).

42- Orsini, F. et al. Atomic force microscopy characterization of Xenopus laevis oocyte plasma membrane. *Microsc. Res. Tech.* **69,** 826-834 (2006).

43- Santacroce, M. *et al.* Atomic force microscopy imaging of Xenopus laevis oocyte plasma membrane purified by ultracentrifugation. *Microsc Res Tech.* **71,** 397-402 (2008).

44- Orsini, F. *et al.* Intermittent contact mode AFM investigation of native plasma membrane of Xenopus laevis oocyte. *Eur. Biophys. J.* **38,** 903-910 (2009).

45- Bosshart, P.D., Frederix, P.L. & Engel, A. Reference-free alignment and sorting of single-molecule force spectroscopy data. *Biophys. J.* **102,** 2202-2211 (2012).

46- Li, H., Oberhauser, A.F., Fowler, S.B., Clarke, J. & Fernandez, J.M. Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. USA* **97,** 6527-31

(2000).

47- Li, H. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998-1002 (2002).

48- Watanabe, K. et al. Molecular mechanics of cardiac titin's PEVK and N2B spring elements. J. Biol. Chem. 277, 11549-11558 (2002).

49- Gordon, S.E., Varnum, M.D. & Zagotta, W.N. Direct interaction between amino- and carboxyl-terminal domains of cyclic nucleotide-gated channels. *Neuron* **19,** 431-441 (1997).

50- Rosenbaum, T. & Gordon, S.E. Dissecting intersubunit contacts in cyclic nucleotide-gated ion channels. *Neuron* **33,** 703-713 (2002).

51- Mazzolini, M., Nair, A, & Torre, V. A comparison of electrophysiological properties of the CNGA1, CNGA1tandem and CNGA1cys-free Channels. *Eur. Biophys. J.* **37,** 947-959 (2008).

52- Nair, A.V., Mazzolini, M., Codega, P., Giorgetti, A. & Torre, V. Locking CNGA1 channels in the open and closed state. *Biophys. J.* **90,** 3599-3607 (2006).

53- Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H.E. How strong is a covalent bond? *Science* **283,** 1727-1730 (1999).

54- Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: a comparison. Proc. Natl. Acad. Sci. USA **96,** 3694-3699 (1999).

55- Andreopoulos, B. & Labudde, D. Efficient unfolding pattern recognition in single molecule force spectroscopy data. *Algorithms. Mol. Biol.* **6,** 16 (2011).

56- Anselmi, C., Carloni, P. & Torre, V. Origin of functional diversity among tetrameric voltage-gated channels. *Proteins* **66,** 136-146 (2007).

57- Kusch, J. *et al.* Role of the S4-S5 linker in CNG channel activation. *Biophys. J.* **99,** 2488-2496 (2010).

58- Peng, Q., Fang, J., Wang, M. & Li, H. Kinetic partitioning mechanism governs the folding of the third FnIII domain of tenascin-C: evidence at the single-molecule level. *J. Mol. Biol.* **412,** 698-709 (2011).

59- Kedrov, A., Janovjak, H., Sapra, K.T. & Müller, D.J. Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. Annu. Rev. Biophys. Biomol. Struct. **36,** 233-260 (2007).

60- Dietz, H. & Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci USA.* **101,** 16192-16197 (2004).

61- Dietz, H. & Rief, M. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci USA.* **103,** 1244-1247 (2006).

62- Liu, D.T., Tibbs, G,R., Paoletti, P. & Siegelbaum, S.A. Constraining ligand-binding site stoichiometry suggests that a cyclic nucleotide-gated channel is composed of two functional

dimers. Neuron **21,** 235-248 (1998).

63- Matulef, K. & Zagotta, W,N. Multimerization of the ligand binding domains of cyclic nucleotide-gated channels. *Neuron* **36,** 93-103 (2002).

64- Kowal, J. et al. Ligand-induced structural changes in the cyclic nucleotide-modulated potassium channel MloK1. *Nat. Commun.* **5,** 3106 (2014). doi: 10.1038/ncomms4106.

65- Ng, C.A. *et al.* The S4-S5 linker acts as a signal integrator for HERG K+ channel activation and deactivation gating. *PLoS One* **7,** e31640 (2012). doi: 10.1371/journal.pone.0031640.

66- Clayton, G.M., Altieri, S., Heginbotham, L., Unger, V.M. & Morais-Cabral, J.H. Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. *Proc. Natl. Acad. Sci. USA.* **105,** 1511-1505 (2008).

67- Yarov-Yarovoy, V. et al. Structural basis for gating charge movement in the voltage sensor of a sodium channel. *Proc. Natl. Acad. Sci. USA* **109,** E93-102 (2012).

68- Murata, M. Novel biosensor for the rapid measurement of estrogen based on a ligand-receptor interaction. *Anal. Sci.* **17,** 387-390 (2001).

69- Verbelen, C., Gruber, H.J. & Dufrêne, Y,F. The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. *J. Mol. Recognit.* **20,** 490-494 (2007).

SUPPLEMENTARY INFORMATION

Conformational rearrangements in the transmembrane domain of CNGA1 channels upon gating revealed by Single Molecule Force Spectroscopy

Sourav Maity[1], Mazzolini Monica[1-2], Manuel Arcangeletti[1], Alejandro Valbuena[1], Paolo Fabris[1], Marco Lazzarino[2-3] and Vincent Torre[1*]

1 - International School for Advanced Studies (SISSA) Neuroscience Area via Bonomea 265, 34136 Trieste (Italy)

2 - CBM S.c.r.l., Area Science Park, Basovizza 34149 Trieste, (Italy)

3 - IOM-CNR, Basovizza, Area Science Park, 34149 Trieste (Italy)

Supplementary Information 1: Electrophysiological recordings.

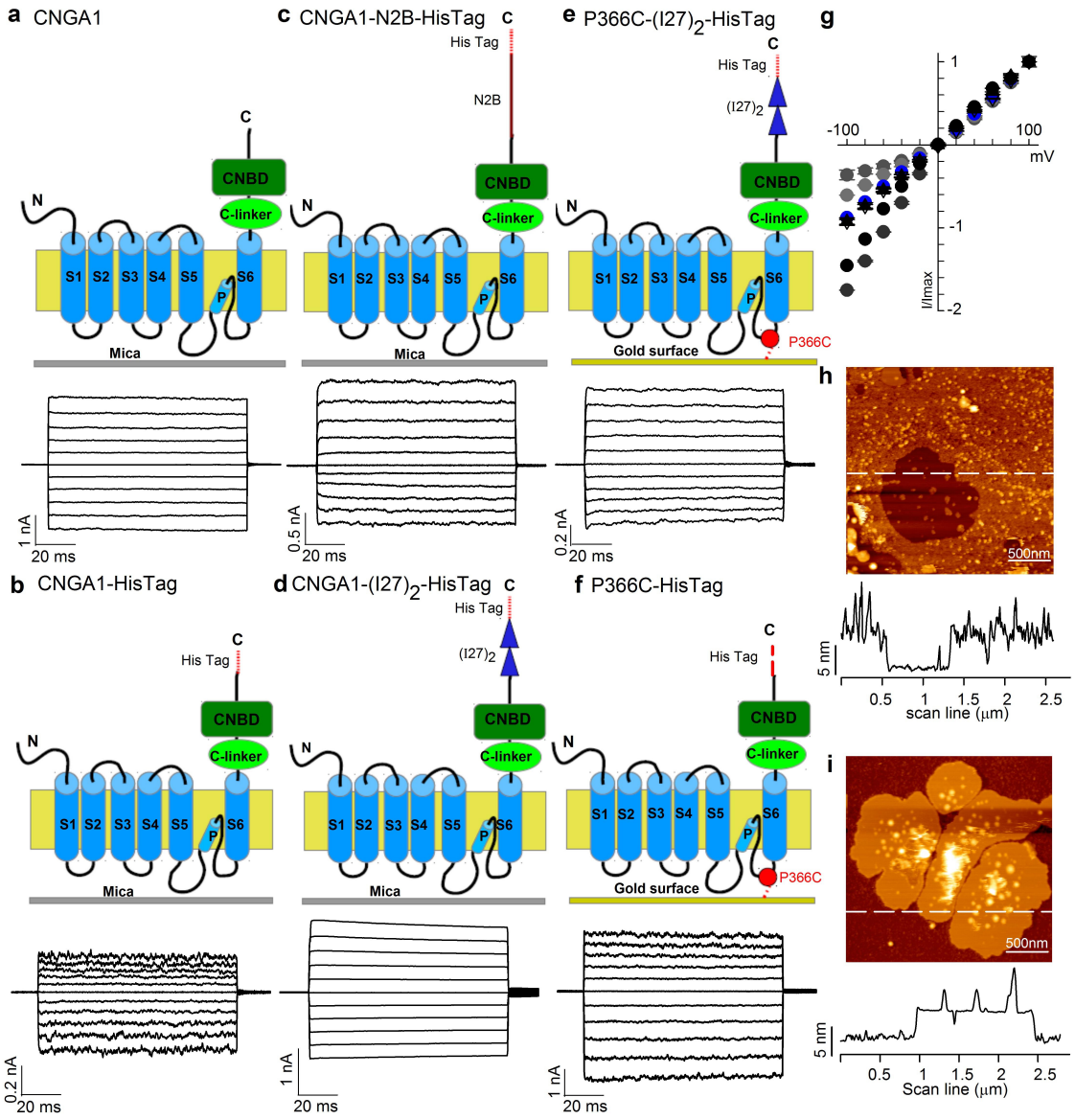


Fig. SI1: The constructs and finger-prints that were utilized. ***a-f:*** *schematic representation (upper panel) and corresponding current recordings (lower panel) of the CNGA1 channel (**a**), CNGA1-HisTag (**b**), CNGA1-N2B-HisTag (**c**), CNGA1-(I27)$_2$-HisTag (**d**), P366C-(I27)$_2$-HisTag (**e**) and P366C-HisTag (**f**). For current recordings of the CNGA1-CNGA1 tandem see Ref.3. The current recordings are measured in an inside-out excised patch under voltage-clamp conditions; voltage steps are from -100 to 100 mV. Curves obtained from the average of at least 5 recordings in the presence of 2 mM cGMP, minus the average of at least 2 recordings in the closed state.* ***g:*** *I/V relationships of the cGMP-activated current from the CNGA1 channel (blue circles) and of the construct CNGA1-HisTag (dark gray circles), CNGA1-N2B-HisTag (gray circles), CNGA1-(I27)$_2$-HisTag (white triangles), P366C-(I27)$_2$-HisTag (black triangles) and mutant channel P366C-HisTag (middle gray circles).* ***h-i:*** *Representative AFM images of membrane patches with a high or a low protrusion density with the corresponding height profile taken along the dashed lines.*

Only oocytes that had been injected with the construct CNGA1-HisTag (Fig. SI1b) had a lower level of expression. Currents activated with cGMP were slightly inward-rectifying, and the mutant

channels F380C showed significant rectification as already described[6].

Pulling experiments with the construct CNGA1-$(I27)_2$-HisTag did not produce F-D curves that contained the unfolding of the CNGA1 channels or the sawtooth pattern that is attributable to the unfolding of the I27 modules. The reason for this failure is that the force necessary to unfold the I27 modules is approximately 200 pN, which is significantly larger than that required to unfold CNGA1 channels (see Fig. 4 and 5). The construct CNGA1-$(I27)_2$-HisTag could therefore be pulled away from the membrane without unfolding the I27 modules. Only the experiments obtained with the construct P366C-$(I27)_2$-HisTag are reported in the main text and are illustrated in Fig.3.

Supplementary Information 2: Initial filtering of F-D curves.

The F-D curves that were collected by the AFM JPK software were filtered to remove unsuitable cases. Filtering was based on the analysis of the pushing (red curves) and pulling signal (black curves) in the following way:

- Treatment of the retraction signal by median and variance filtering. If the number of peaks of the filtered pulling signal was < 2, the curve was discarded (Fig.SI2a);

- Comparison of the distance between the retraction and extension signal. If the distance was > 30 pN, the curve was discarded (Fig.SI2b);

- Comparison of crossings between the extension and retraction signal. If the distance between two crossings was > 20 pN, the curve was discarded (Fig.SI2c);

- Computation of the variance σ^2 of the extension signal to estimate the motion of baseline. If σ^2 was > 10 pN^2, the curve was discarded (Fig.SI2d);

- Comparison between the initial slope of the extension Sp and retraction Sr signal. If (1-Sp/Sr) was > 1, the curve was discarded (Fig.SI2d,e);

- Calculation of the maximum pulling force (F). If F was < 30 pN, the curve was discarded (Fig.SI2f);

- Calculation of the extension signal drift. If the maximum amplitude was > 35 pN in the "flat" region, the curve was discarded;

- Comparison of the distance between the retraction and extension signal. If the distance was > 30 pN the curve was discarded. In such cases, the retraction signal was not complete;

- Comparison of the length of tip-sample distance (x) of the extension and retraction signal. If the length difference was > 10%, the curve was discarded;

- Computation of the (F,Lc) plot of the pulling signal. The algorithm determines if there is at least one peak (extracted as described in Supplementary 3) in the range between 80 and 300 nm and that the last peak is in a range between 250 and 350 nm (Fig.SI2a-f).

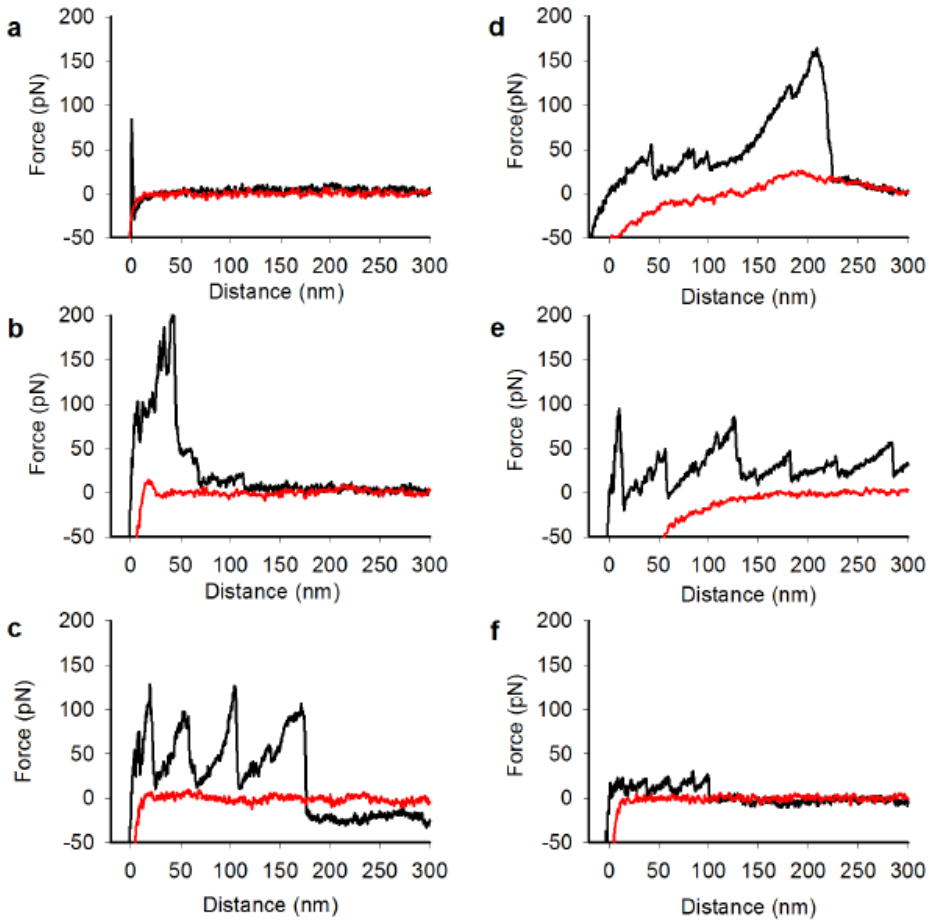


Fig. SI2: Examples of F-D curves that were discarded by our filtering steps. ***a:*** *absence of force and tip-sample distance that demonstrates the absence of an interaction between the cantilever and the protein.* ***b:*** *high initial nonspecific force that is most likely due to interactions with proteins on the oocyte membrane surface;* ***c:*** *short tip-sample distance showing an Lc value that does not correspond to our constructs;* ***d-e:*** *protein is unfolded before the pulling cycle started;* ***f:*** *low force and short tip-sample distance that are not characteristic of our constructs. The force-extension data for the approach of the cantilever to the surface are shown in red, and data for the retraction are shown in black.*

Supplementary Information 3: Computation of Lc, Lc histograms and of histogram of Max Lc

F-D curves composed of a sequence of points (F,D) were transformed into a histogram of points (F,Lc), using Bosshart procedure[7].

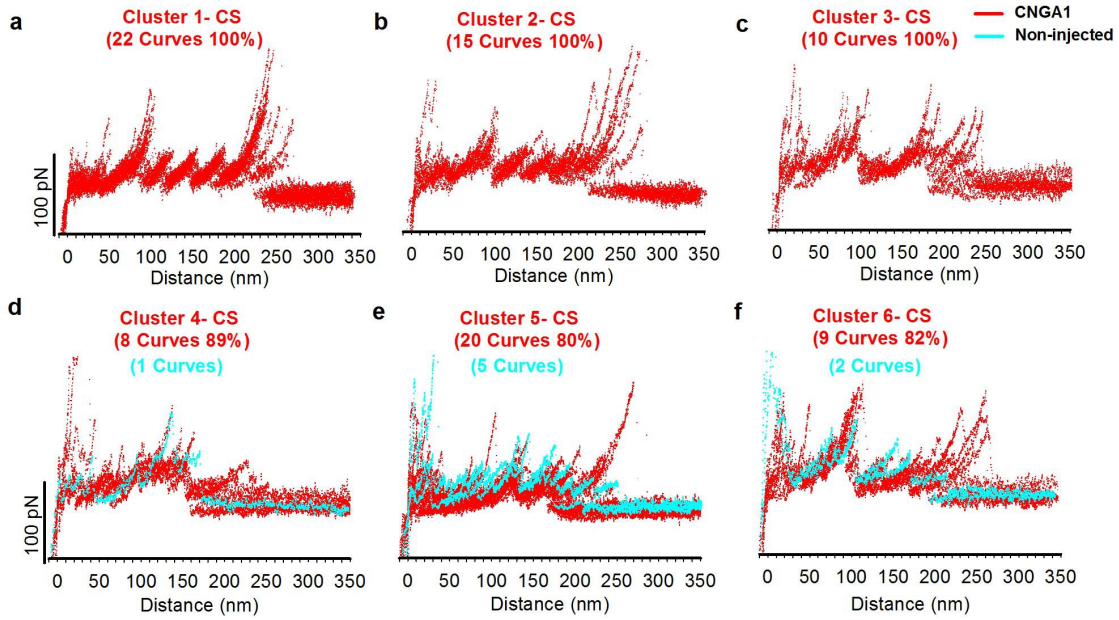


Fig. SI3: Clustering of F-D curves: a-c: *three different clusters containing curves from membrane extracted from injected oocytes for CNGA1 channels in the closed state.* ***d-f:*** *promiscuous clusters containing curves from membrane extracted from injected (in red) and non injected (in cyan) oocytes.*

Supplementary Information 4: Coding of F-D curves and distances between them.

Methods of Computer Science and Pattern Recognition for clustering objects, patterns or instances with similar features require a measure of their similarity or more precisely a distance between them. In order to have a distance between different F-D curves we coded F- D curves in an appropriate sequence of symbols and we constructed similarity measurements – or distances – between coded F-D curves. The coding was based on the observation that F-D curves obtained in SMFS experiments are well fitted by the superposition of curves obtained from the WLC model (see Methods). We used three different distances: classic MAE (Mean Absolute Error), Hamming and Binary-Force. These distances are used with the different coding schemes. The MAE is used with Coding III (F,Lc), and the Hamming-Binary is used with Coding I and II.

The MAE distance, which is expressed in force (pN), is computed as:

$$MAE = 1/n \sum_{k=1}^{n} | x_i(k) - x_j(k) |$$

where n is the total number of points and x_i and x_j are the coded sequence of symbols.

The Hamming-force distance is computed as:

$$Hamming - force = \sum_{k=1}^{n} | x_i(k) - x_j(k) |$$

where n is the total number of points and x_i and x_j are the coded sequence of symbols.

For the Binary force each peak of the F-D curves is individually compared with every other peak with a selected curve used as model. If the distance (in bin size) is in the range [-1, 0, 1], the error is computed as in Hamming-Force. Otherwise, the error is set to zero. The distance matrices that were used are shown in Fig.SI4a-c. After the filtering step that uses all of the curves from from membrane extracted from injected and non injected F-D curves, it is possible to distinguish the injected F-D curves for each distance matrix.

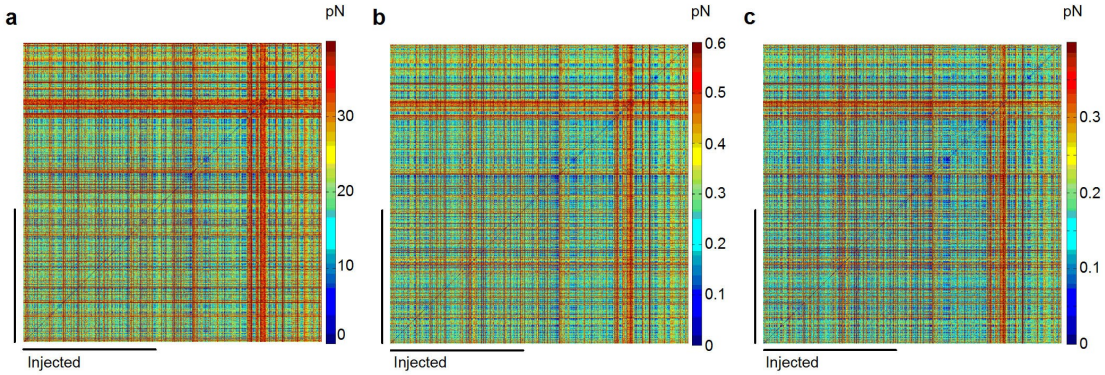


Fig. SI4: Distance matrix with 1585x1585 entries*. **a:** MAE distance, **b:** Hamming distance, **c:** Binary distance. Blue represents the minimum value (i.e., the distance between curves is very low, indicating that the curves are similar). Red represents the maximum values (i.e., the distance between curves is high, indicating that the curves are different). The number of F-D curves is the same for both axes. The black line represents the F-D curves obtained from membrane extracted from injected oocytes.*

Supplementary Information 5: Clustering of F-D curves.

Standard clustering procedures can be used after the F-D curves have been converted into symbol sequences $x_1,x_2,.....x_n$ and a $D(x_i,x_j)$ between symbol sequences is available (see Supplementary 4). We have used and compared average, centroid, complete, median, single, ward and weighted clustering algorithms (all these algorithms are agglomerative clustering algorithms from the MATLAB statistics toolbox http://www.mathworks.it/it/help/stats/clusterdata.html, see also Ref. 8). All these algorithms need to have the total number of clusters $N_{cluster}$ assigned, and the choice of $N_{cluster}$ is crucial. We have circumvented this problem by mixing the F-D curves from membrane extracted from injected and noninjected oocytes, and we have varied the value of $N_{cluster}$. For values of $N_{cluster}$ that are less than 10, none of the tested clustering algorithms are able to distinguish between F-D curves that were obtained from injected and from noninjected oocytes. When $N_{cluster}$ is between 50 and 100, clusters of F-D curves can be observed from injected oocytes with probabilities of 0.8, 0.9 and 1, i.e., in which 80, 90 and 100% of curves are from injected oocytes. We found that when we used the clustering algorithm "complete" and when $N_{cluster}$ was larger than 100, several clusters from injected oocytes could be detected with a probability that was larger than 0.8. We selected as the value of $N_{cluster}$ the value for which we had the largest clusters of F-D curves from injected oocytes with a probability of 1. Examples of obtained clusters are shown in Fig.SI3 of the main text: there are several clusters with promiscuous F-D curves (Fig.SI3d-f), i.e. obtained from both injected (red curves) and not injected oocytes (cyan curves). These F-D curves presumably represent the unfolding of proteins or molecules forming the plasma membrane of *Xenopus laevis* oocytes and therefore can be obtained from either injected or not injected oocytes. These F-D

curves have force peaks of about 60 pN. As shown in Fig.SI5a we found one cluster (Cluster1-CS) with 22 F-D curves obtained only from injected oocytes. These F-D curves have force peaks with values of Lc of 5 force peaks that are located at 96±3, 116±3, 159±3, 189±5 and at 234±6 nm and forces of 60 pN. We did not find any cluster with more than 5 F-D curves obtained only from not injected oocytes.

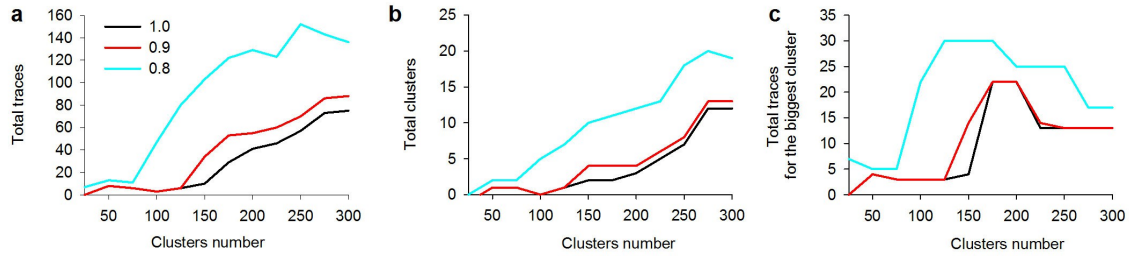


Fig. SI5: Clustering of F-D curves as a function of $N_{cluster}$ a: *relation between total number of F-D curves from membrane extracted from injected oocytes with probabilities of 0.8, 0.9 and 1 as a function of $N_{cluster}$.* ***b:*** *relationship between total number of clusters from membrane extracted from injected oocytes with probabilities of 0.8, 0.9 and 1 as a function of $N_{cluster}$.* ***c:*** *relationship between total number of F-D curves of the largest cluster from membrane extracted from injected oocytes with probabilities of 0.8, 0.9 and 1 as a function of $N_{cluster}$. Clustering was obtained by the algorithm "complete" using the MAE distance. As shown in c, when $N_{cluster}$ is about 200, clusters with more than 20 F-D curves, only curves from membrane extracted from injected oocytes can be observed. Clustering was therefore performed with a value of $N_{cluster}$ equal to 200.*

Supplementary Information 6: Intracluster and Intercluster distances.

Given the distance matrix *Dist(x,y)* and a cluster C1 the intracluster distance IntraD is defined[9] as:

$$IntraC_1 = (1/(n^2-1)) \sum_{i,j} Dist(x_i,x_j)$$

where n is the total number of sequences in C1 and x_i i=1,...n are the sequences in C1. Given two clusters C1 and C2 the intercluster distance $InterC_1C_2$ is defined[9] as:

$$InterC_1C_2 = (1/n\,m) \sum_j \sum_i Dist(x_i,y_j)$$

Where n and m are the total number of sequences in cluster C_1 and C_2 respectively and x_i i=1..n and y_j j=1,..m are the sequences present in C_1 and C_2 respectively. A cluster of finger-prints f_n FP identifies cluster C_i among a set of other clusters C_k k=,…,N if

$InterFPC_i$ < InterFP C_k for all k different from i

A cluster of finger-prints f_n FP identifies very well cluster C_i among a set of clusters C_k k=,…,N if

$InterFPC_i$ < InterFP C_k for all k different from I and

$InterFPC_i$ is very similar (within 10%) to $IntraC_i$ and IntraFP.

In Fig.1a the set of fingerprints with the N2B construct identifies cluster 1 among all clusters with obtained as described below according to all considered coding and distances and identifies very well cluster 1 according coding I. As the F-D curves from the CNGA1-N2B-Histag construct have slightly higher forces; according to coding II and III InterFP-clusters using MAE distance is higher for IntraFP-clusters and about twice to Intra-clusters, while using Hamming-force distance InterFP-clusters is higher for IntraFP-clusters and for Intra-clusters (see table SI6).

In Fig.1, clusters were obtained with $N_{cluster}$ = 200 and by selecting those clusters in which at least with 80% of curves were obtained from membrane extracted from oocytes. With this procedure we identified 15 clusters and 3 of them were obtained only from membrane extracted from injected oocytes. F-D curves with the N2B fingerprint identified cluster I as obtained from the unfolding of CNGA1 channels. Table SI6 reports the Intracluster distance for the 3 clusters and for the set of F-D curves (FP) with the N2B fingerprint and the Intercluster distance between each cluster and FP:

Coding	Distance	Cluster	Intracluster	IntraFP	Intercluster (with FP)
I	Hamming	1	0.34	0.32	0.33
I	Hamming	2	0.48	0.32	0.57
I	Hamming	3	0.42	0.32	1.09
II	Hamming-force	1	2.35e-10	3.16e-10	3.34e-10
II	Hamming-force	2	2.22e-10	3.16e-10	3.53e-10
II	Hamming-force	3	2.92e-10	3.16e-10	3.89e-10
III	MAE	1	1.24e-11	2.12e-11	2.27e-11
III	MAE	2	9.67e-12	2.12e-11	2.35e-11
III	MAE	3	1.38e-11	2.12e-11	2.68e-11

Table SI6: Intercluster and intracluster distance

Supplementary Information 7: Comparison between closed and open state.

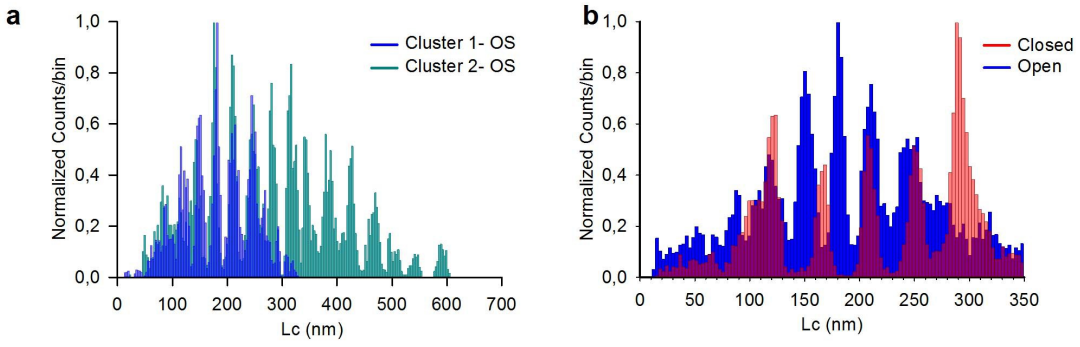


Fig. SI7: Two clusters of CNGA1 channels in the open state and comparison between open and closed state. a: *superimposition of histograms of normalized counts/bin against Lc (see Supplementary 3) from the F-D curves from the unfolding of a single (in blue) and two concatenated (in cyan) CNGA1 channels in the open state.* ***b:*** *superposition of histograms of normalized counts/bin compared with Lc for CNGA1 in the open state (in blue) and the closed state (in red).*

Supplementary Information 8: Distribution of force amplitudes in the closed and open state.

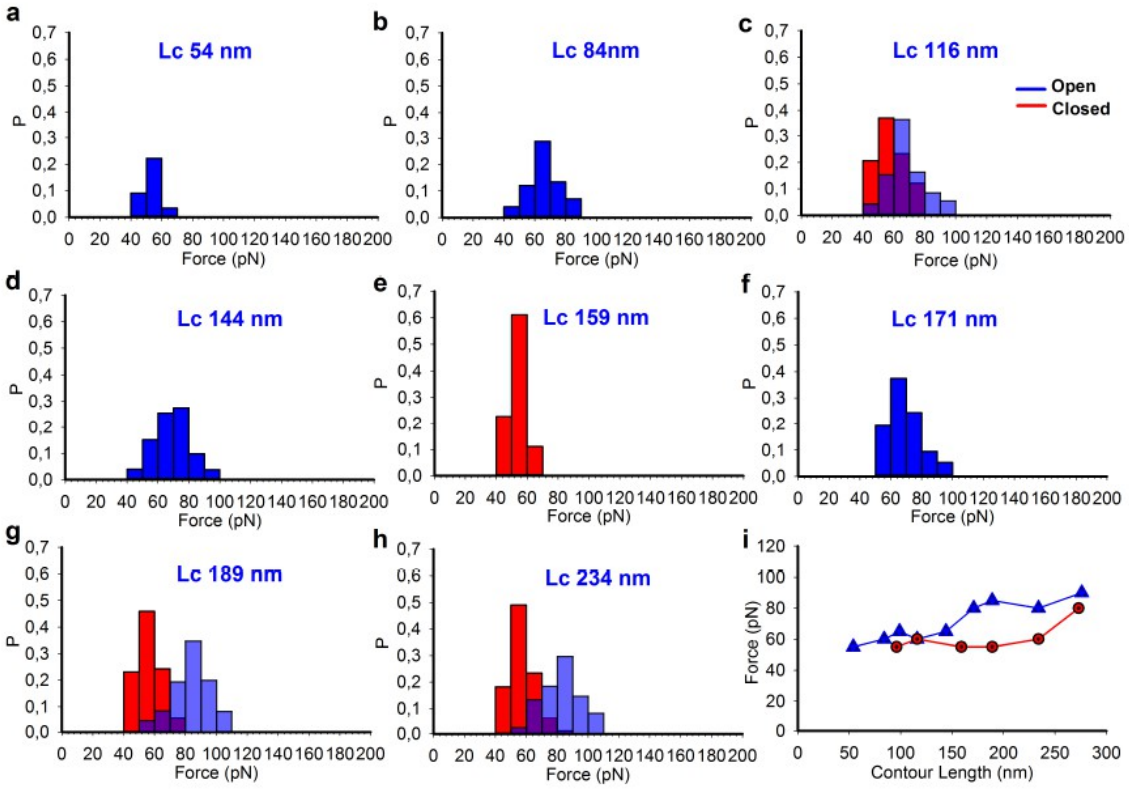


Fig. SI8: Force distribution histograms. a-h: *histograms corresponding to the different Lc for 157 F-D curves of CNGA1 construct in the closed (red) and 132 F-D curves in the open (blue) state.* ***i:*** *plot comparison between the force in the close (red) and open (blue) state.*

Supplementary Information 9: SMFS of the CNGA1-CNGA1 tandem.

In order to obtain a further validation that F-D curves of Fig.1b and 2b are obtained from the unfolding of CNGA1 channels in the closed and open state respectively we performed SMFS on the CNGA1-CNGA1 tandem construct obtained from the concatenation of two identical CNGA1 subunits linked by a short linker (10 aa between the C-terminal end of the first subunit and the N-terminal end of the second subunit). When the mRNA of the CNGA1-CNGA1 tandem construct is injected into oocytes there is a good expression of functional channels with properties very similar to those of CNGA1 channels[10]. F-D curves obtained from the unfolding of CNGA1-CNGA1 tandem are expected to have the largest value of Lc above 400 nm and to have well defined repeats of force peaks separated by an appropriate value of ΔLc.

Each CNGA1 subunit contains seven endogenous cysteines: C35, C169, C186, C314, C481, C505 and C573. C35 near the N-terminal is thought to interact with C481 in the open state, but not in the closed state[11,12]. C169 and C186 are in S1 and C314 is located in S5 and is known to interact with residues at position 380 in the S6 domain[13]. C481 and C505 are located in the C-linker and in the CNB domain respectively[14]. C573 is located near the C-terminus of the channel[10] and does not seem to have any structural and functional role. In the absence of the formation of C-C bonds, unfolding of the second subunit will be repeated at a distance corresponding to 690+10 aa, i.e. at about 280 nm, but in the presence of a C-C bond between C35 and C481 – promoted by the concatenation of the two subunits - the distance between the two repeats will be shorter by about 102 nm, i.e. at 178 nm.

We performed SMFS experiments with the CNGA1-CNGA1 tandem construct and in the closed state, we obtained 22 F-D curves with clear force repeats (Fig. SI9a) and the analysis of (Fc,Lc) histograms indicated that these repeats were separated by about 180 ± 6 nm (Fig. SI9b). In the open state, we obtained 34 F-D curves with a repeated pattern (Fig. SI9c,d) very similar to what observed for the unfolding of CNGA1 channels in the open state (Fig.2). Force peaks from the tandem constructs in the open state were larger, varying from 75 to 120 pN, than those obtained in its absence, in agreement with what seen for CNGA1 channels (compare Fig.1 and Fig.2).

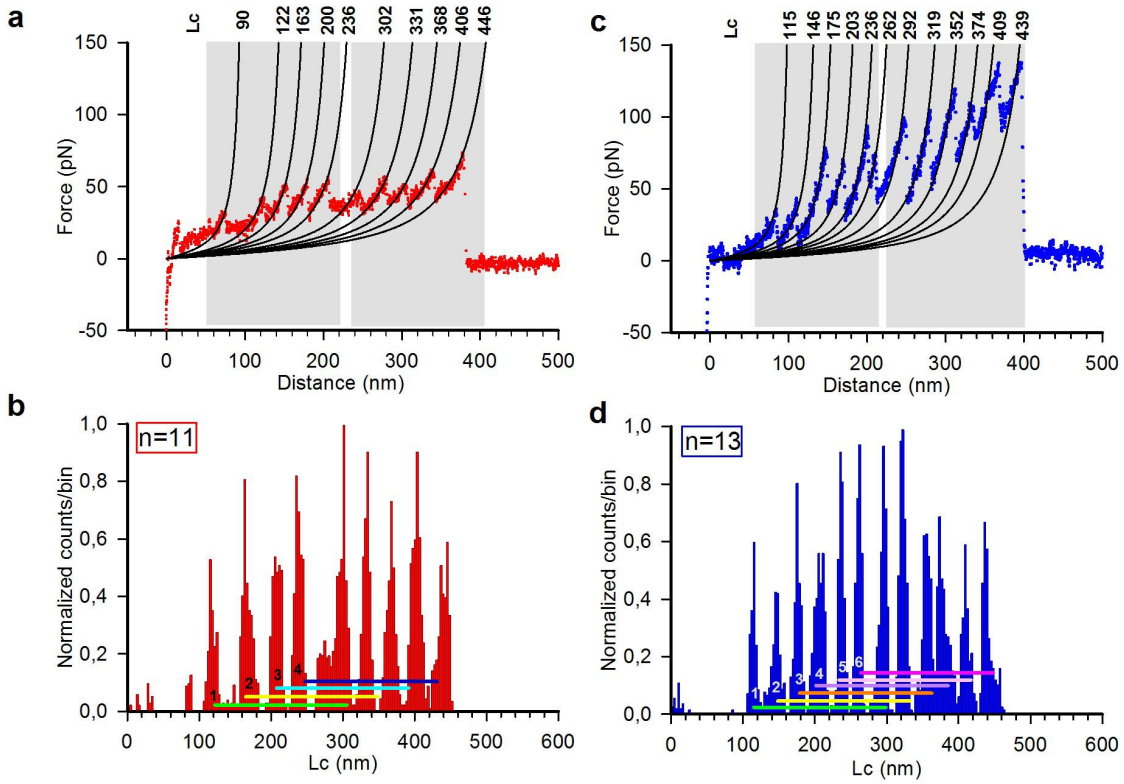


Fig. SI9: F-D curves from CNGA1-CNGA1 tandem in the closed state open and closed state. a,c: *F-D curves recorded during the unfolding of single construct in the closed state (red) and in the open state (blue).* ***b,d:*** *histograms of normalized counts/bin against Lc (see Supplementary 3) from the F-D curves of panel a (in red) and F-D curves of panel c (in blue) construct in the closed and in the open state, respectively. The colored lines (180 nm) indicate that these repeats are separated by about 180 ± 6 nm.*

Supplementary Information 10: SMFS of the mutant channels F380C.

An additional test for the validation of F-D curves ascribed to the unfolding of CNGA1 channels is represented by the unfolding of mutant channels F380C (Fig.SI10a). This channel in the presence of cGMP is locked in the open configuration by the formation of S-S bond between the native C314 and the exogenous cysteine in position 380 (Ref. 13). Since S-S bond has a breaking force of about 1.4 nN (Ref. 15) and CNGA1 channels have a mechanical stability below 150 pN and it is obvious that we unfold the CNGA1 channel without breaking the S-S bond. In the presence of a S-S bond, the total stretched length decreases by $N_{aa} \cdot l_{aa}$, where N_{aa} is the total number of aa in the S-S loop and l_{aa} the mean length of a single aminoacid assumed to be 0.4 nm (Rief al., 1999); as a consequence F-D curves from the mutant channel F380C are expected to be shorter by 66 aa (F380-C314) ~ 26 nm of Lc (Fig.SI10a). We performed SMFS on mutant channels F380C in the presence of 2 mM cGMP with 10 µm CuP, agent promoting the formation of S-S bonds[13]. We used F-D curves obtained from the CNGA1 channels in the open state (Fig.2) as a template to identify F-D curves obtained from the unfolding of the mutant channel F380C.

We identified 28 curves which when superimposed and aligned had the initial force peaks corresponding to those of the CNGA1 channels (Fig. SI10b): F-D curves from the mutant channel F380C (green curves) superimposed to those from the CNGA1 channels (blue curves) and had common force peaks with Lc around 54, 86 and 116 nm (black WLC curves in Fig. SI10b). The force peaks seen in CNGA1 channels with a value of 189 ± 5 and 234 ± 6 nm (red WLC curves in Fig.SI10b) in the mutant channel F380C appear shifted by about 26-28 nm to the left (green WLC curves in Fig.SI10b). These results indicate F-D curves from the mutant channel F380C have a gap of 26 $\pm$ 2 nm compared to those from the CNGA1 channels. Inspection of Lc histograms show a similar result: the initial and final portion of F-D curves from mutant channel F380C and CNGA1 channels have the force peaks with the same value of Lc, provided that approximately 66 amino acids are deleted (Fig.SI10c,d). If mutant channel F380C in the open state has a disulfide bond between the endogenous C314 and the exogenous cysteine in position 380, F-D curves obtained from its unfolding are expected to be 25 ± 2 nm shorter than those obtained from the CNGA1 channels, in agreement with the experimentally observed gap of 26 $\pm$ 2 nm (Fig.SI10c,d). This result provides a further experimental validation of the identification of F-D curves of Fig.1-2 as obtained from the unfolding of CNGA1 channels and of the existence of a S-S bond in mutant

channels F380C in the open state.

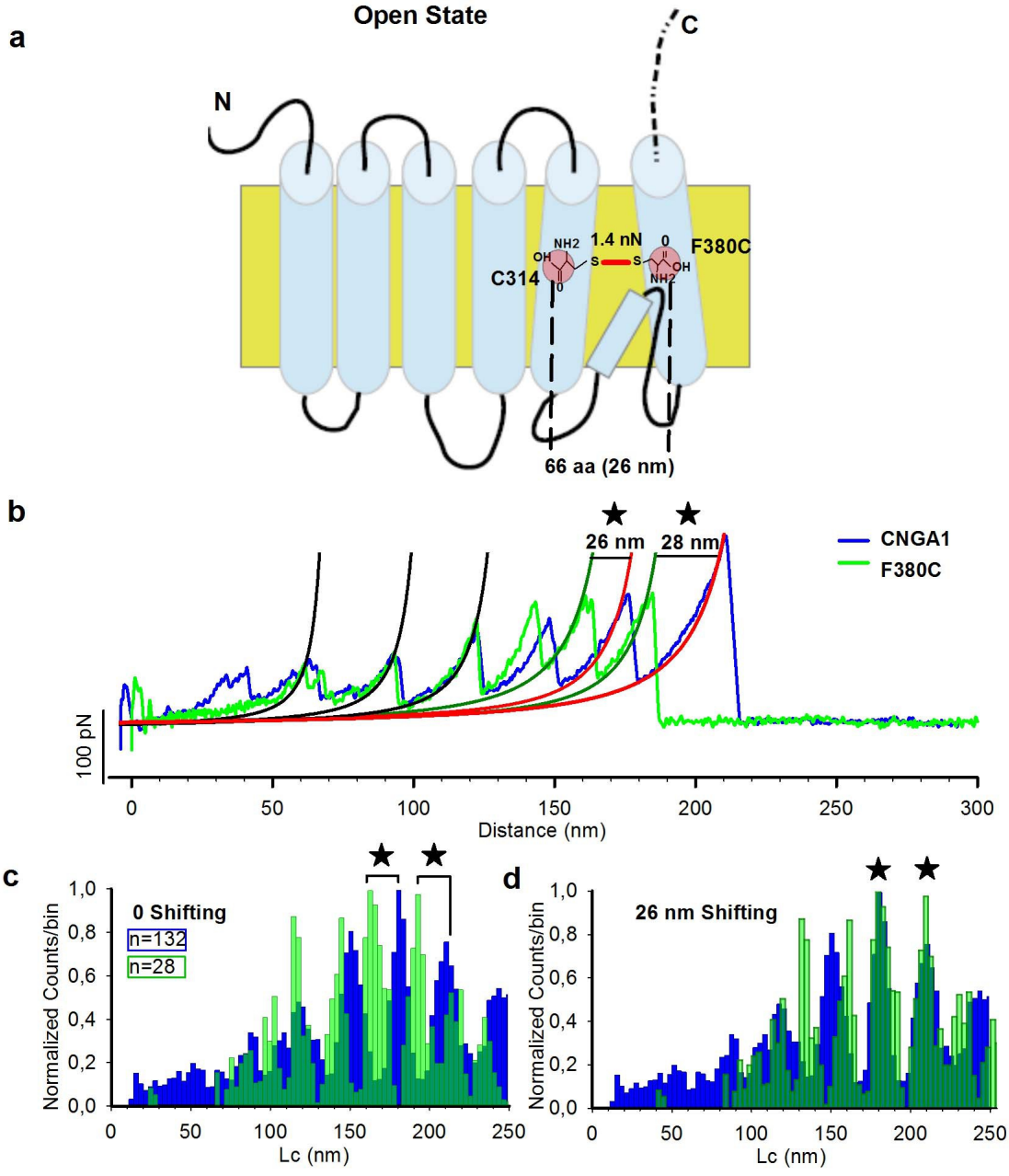


Fig. SI10: Unfolding of F380C mutant channel in the open state. *a: schematic representation of interaction between the endogenous C314 and the exogenous cysteine in position 380 in mutant channel F380C. b: superposition of a representative F-D curve from CNGA1 (blue) and from mutant channel F380C (green) in the presence of 2 mM cGMP; the F-D curve from mutant channel F380C was obtained in the presence of 10 µm CuP to promote the formation of S-S bonds. F-D curves were fitted using the WLC model. The first three force peaks have very similar values of Lc, while the last two force peaks (indicated by the black star) of mutant channel F380C are shifted to the left by 26 nm, corresponding to the full extension of the stretch of aa from C314 to C380. c: superimposition histogram representing the normalized frequency of events versus the Lc for CNGA1 (blue) and mutant F380C (green) channels. d: the same as in d but with a shift of 26 nm.*

References

1-Kaupp, U.B. *et al.* Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel. *Nature* **342,** 762-766 (1989).

2-Gordon, S.E. & Zagotta, W.N. Subunit interactions in coordination of Ni2+ in cyclic nucleotide-gated channels. *Proc. Natl. Acad. Sci. USA* **92,** 10222-10226 (1995).

3-Mazzolini, M., Nair, a. & Torre, v. A comparison of electrophysiological properties of the CNGA1, CNGA1tandem and CNGA1cys-free channels. *European biophysics journal* 947-59 (2008).

4-Murata, M. Novel biosensor for the rapid measurement of estrogen based on a ligand-receptor interaction. *Anal. Sci.* **17,** 387-90 (2001).

5-Verbelen, C., Gruber, H.J. & Dufrêne, Y,F. The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. *J. Mol. Recognit.* **20,** 490-494 (2007).

6-Nair, A.V., Mazzolini, M., Codega, P., Giorgetti, A. & Torre, V. Locking CNGA1 channels in the open and closed state. *Biophys J.* **90,** 3599-607 (2006).

7-Bosshart, P.D., Frederix, P.L. & Engel, A. Reference-free alignment and sorting of single-molecule force spectroscopy data. *Biophys. J.* **102,** 2202-2211 (2012).

8-Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. "14.3.12 Hierarchical clustering". The Elements of Statistical Learning (2nd ed.). New York: Springer. pp. 520–528 (2009).

9-Daliri, M.R. & V. Torre, V. Shape and Texture Clustering: The Best Estimate for the Cluster Numbers. *Image and Vision Computing* **27,** 1603-1614 (2009).

10-Mazzolini, M., Nair, A, & Torre, V. A comparison of electrophysiological properties of the CNGA1, CNGA1tandem and CNGA1cys-free Channels. *Eur. Biophys. J.* **37,** 947-959 (2008).

11- Rosenbaum, T. & Gordon, S.E. Dissecting intersubunit contacts in cyclic nucleotide-gated ion channels. *Neuron* **33,** 703-713 (2002).

12- Gordon, S.E., Varnum, M.D. & Zagotta, W.N. Direct interaction between amino- and carboxyl-terminal domains of cyclic nucleotide-gated channels. *Neuron* **19,** 431-441 (1997).

13- Nair, A.V., Mazzolini, M., Codega, P., Giorgetti, A. & Torre, V. Locking CNGA1 channels in the open and closed state. *Biophys. J.* **90,** 3599-3607 (2006).

14- Nair, A.V., Anselmi, C. & Mazzolini, M. Movements of native C505 during channel gating in CNGA1 channels. *Eur. Biophys. J.* **38,** 465-478 (2009).

15- Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H.E. How strong is a co valent bond? *Science* **283,** 1727-1730 (1999).

Manuscript 2

Ready to submit

Unfolding variability of CNGA1 channels in the open and closed states using single molecule force spectroscopy

Sourav Maity, Mazzolini Monica and Vincent Torre[1*]

Author contributions:

S.M. Performed the experiments, data analysis and prepared the figures.

S.M., M.M. And V.T. contributed to writing the paper; V.T. supervised the project.

Unfolding variability of CNGA1 channels in the open and closed states using single molecule force spectroscopy

Sourav Maity[1], Monica Mazzolini[1] and Vincent Torre[1*]

1 - International School for Advanced Studies (SISSA) Neuroscience Area via Bonomea 265, 34136 Trieste (Italy)

[*]Corresponding author: torre@sissa.it

Running title: Unfolding variability of CNGA1 channels

Keywords: CNG channels, Ion channels, Gating, Single molecule force spectroscopy, SMFS

We have analyzed with single molecule force spectroscopy (SMFS) the unfolding pathways of a single subunit of CNGA1 channels expressed in *Xenopus laevis* oocytes and we investigated the variability of the unfolding both in the presence of 2 mM cGMP (open state) and in its absence (close state). There was a significant variability during the unfolding of the cytoplasmic domain both in the closed and open state. The unfolding of the transmembrane domains was much less variable and force peaks with a given value of contour length (Lc) appeared with a probability close to 1, but the increase of contour length (ΔLc) between two successive force peaks could vary from 32 to 52 nm.

These observations suggest that the different molecular conformations of CNGA1 channels and a ragged free-energy landscape, could explain their problematic crystallization.

INTRODUCTION

Voltage-gated ion channels form a large super-family comprising Na^+, K^+ and Ca^{2+} channels, whose transitions between the open and closed conformation highly depend on the voltage across the membrane. This super-family of ionic channels also includes cyclic nucleotide-gated (CNG) channels(Kaupp et al.,1989; Anderson & Greenberg, 2001; Kaupp & Seifert, 2002; Matulef & Zagotta, 2002) that are also voltage-dependent (Marchesi et al., 2012) but are opened by the binding of cyclic nucleotides of the cyclic nucleotide-binding (CNB) domain (Kaupp et al., 1989; Kaupp & Seifert, 2002; Matulef & Zagotta, 2002). CNG channels have been extensively analyzed and their biochemical and electrophysiological properties have been extensively clarified (Kaupp et al., 1989; Craven & Zagotta, 2006; Mazzolini et al., 2010).

Seven members of the CNG channel gene family have been identified (Kaupp & Seifert, 2002; Becchetti et al., 1999) which are grouped into two subtypes, CNGA (CNGA1-CNGA5) and CNGB (CNGB1 and CNGB3). CNGA1, CNGA2, CNGA3, and CNGA5 (but not CNGA4) when heterologously expressed form functional cyclic nucleotide-activated channels, but CNGB1 and CNGB3 cannot form functional homomeric channels and are modulatory proteins of CNG channels. CNGA1 channels from bovine rods are composed by 690 aminoacids (aas) and hydropathicity and biochemical analyses (Kaupp et al., 1989) have identified six transmembrane α-helices referred as S1, S2, S3, S4, S5 and S6 that are separated by unfolded loops. The N- and C- terminal ends of CNGA1 channels are both cytoplasmic and therefore are not accessible from the extracellular side of the membrane. There is a pore region between S5 and S6 where ion permeation occurs, and electrophysiological experiments have identified 20 amino acids that form the P-helix (V348-P358) and the selectivity filter (T359-P367) (Flynn & Zagotta, 2001, 2003; Contreras et al., 2008; Nair et al., 2009; Mazzolini et al., 2009). The C-terminal (N400-D690) is a large cytoplasmic domain composed of the C-linker (N400-E482) and the CNB domain (A483-N610) (Nair et al., 2009; Mazzolini et al., 2002). The full-length channel has never been crystallized, but a low-resolution architecture (Higgins et al., 2002) obtained from single particle imaging is available and partial crystal structures of the CNB domain (Zagotta et al., 2003; Schunke et al., 2011; Lolicato et al., 2011) have been obtained. In addition, the 3D molecular structure of a channel mimics of the pore (Derebe et al., 2011) is also available.

Single molecule force spectroscopy (SMFS) consists in the application of a force (F) obtained by the tip of a cantilever of an Atomic Force Microscope (AFM) required to unfold a polymer while

the distance (D) between the cantilever tip and the sample is measured with a nanometer resolution. In a force-distance (F-D) curve there are several peaks, each of them representing the unfolding of a segment of the molecule under investigation. The sequence of the unfolding force peaks allows the identification of folded and unfolded regions so to obtain direct information on the secondary structure of the protein (Reif et al., 1997; Engel & Gaub, 2008; Muller et al., 2008). SMFS has been used to study proteins belonging to the rhodopsin family (Oesterhelt et al,. 2000; Sapra et al., 2006; Kawamura et al., 2013) and of other proteins like the Na^+/H^+ antiporter, BetP symporter, KpOmpA transmembrane protein, β2-adrenergic receptor, lysozyme T4L* and leucine binding protein (Kedrov et al., 2006; Bosshart et al., 2012; Zocher et al., 2012; Peng & Li, 2008; Kotamarthi et al., 2013).

In a recent investigation (MS1) we have used SMFS to examine gating of CNGA1 channels that are expressed in *Xenopus laevis* oocytes (MS1), i.e. almost *in situ*. We identified F-D curves by an informatics analysis and by expressing fusion proteins that are composed of CNGA1 channels that are linked at their C-terminal to marker proteins whose unfolding is known. We have identified those domains that change their conformation upon gating.

Variability of the unfolding of molecules obtained with SMFS has been investigated in the lysozyme T4L* (Peng & Li, 2008), in the leucine binding protein (Kotamarthi et al., 2013) where the unfolding occurred in a single or in two steps. These experiments were interpreted as an example of a kinetic partitioning mechanism, originating from the stochastic nature of the unfolding process and/or from the existence of different initial or native conformations. In the present manuscript, we examine in details the variability of the unfolding of CNGA1 channels both in the open (in presence of 2 mM cGMP) and closed state (in the absence of cGMP). Unfolding of CNGA1 channels occurs in several steps, ranging from 5 to 9 steps. We found a high variability during the unfolding of the cytoplasmic domain and a significantly reduced variability when is unfolded the transmembrane domain. These observations provide a new structural information of CNGA1 channels at a single molecule level and hint to the existence of multiple native conformations.

METHODS

Heterologous expression system and sample preparation

Purified RNA coding for the CNGA1 subunit was injected into *Xenopus laevis* oocytes ("Xenopus express" Ancienne Ecole de Vernassal, Le Bourg 43270, Vernassal, Haute Loire, France). Oocytes were prepared as previously described (Sesti et al., 1995). Injected oocytes were maintained at 18°C in a Barth solution supplemented with 50 mg/ml of gentamycin sulfate and contained (in mM): 88 NaCl, 1 KCl, 0.82 $MgSO_4$, 0.33 $Ca(NO_3)_2$, 0.41 $CaCl_2$, 2.4 $NaHCO_3$, and 5 TRIS-HCl (pH 7.4 buffered with NaOH). All the used salts and reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA). The vitelline membrane of oocytes was removed mechanically. We performed SMFS in oocytes in which the cGMP-activated current measured with conventional patch pipettes was larger than 1 nA at ±100 mV (add Ref. NCB). These oocytes were incubated and deposited on a freshly cleaved mica substrate for some tens of minutes in the control solution (which is this solution? Please add the salt composition). The yolk and granules of the cytoplasm were removed by repetitive washings with the control solution. To image the membrane patches, AFM in liquid and in tapping mode[41] was used.

Atomic force microscopy and cantilever functionalization

The AFM NanoWizard 3 (JPK) and an inverted optical microscope (Olympus IX71) were as described in (REF). After localizing plasma membrane patches, we performed SMFS experiments. The cantilever spring constant was ~ 0.08 N/m and was calculated before the start of each experiment by using the equipartition theorem. A 0.4 NA/10X objective was used to localize the area of oocyte incubation. AFM images were acquired with a resolution of 512 pixels. For SMFS experiments, the tip was in contact with the selected membrane patch for approximately 0.5-1 s while we applied a force of 500 pN. The surface was separated at a velocity of 500 nm/s while the force exerted between the tip and surface was recorded. During an experimental session lasting up to 8-10 hours, approximately 10,000 F-D curves were collected in three or four different regions of 1-2 μm size.

We have performed identical SMFS also when the pulling speed was varied from 200 nm/s up to 1500 nm/s.

SMFS experiments and data processing

Clean fragments of oocyte membrane remained anchored to the mica substrate with the intracellular side exposed to the bathing medium (which is this solution? It is Seal Na? Please add the salt composition) and to the cantilever tip of the AFM, was used. After the localization of the membrane patch we moved the AFM tip over the imaged area (usually 1 μm²). Using the matrix scanning mode, the AFM tip was pushed into the surface with a contact force of 1 nN for 0.5 s and then retracted with a constant speed of 500 nm/s. In approximately 20% of cases the tip was able to absorb a molecule providing a sawtooth like F-D curve and if the magnitude of force of these F-D curves was larger than 45 pN, the curve was saved. We performed SMFS and collected approximately 300.000 F-D curves in the absence of 2 mM cGMP (closed state) and 200.000 F-D curves in the presence of 2 mM cGMP (open state) from noninjected (control) oocytes We also collected more than 500.000 F-D curves from oocytes that had been injected with the mRNA of CNGA1 channels in the open and closed state. We applied informatics to identify F-D curves that were present in only the SMFS experiments that used injected oocytes. The method had two steps. First, each F-D curve was mapped in a sequence of symbols that represented the location and amplitude of the force peaks (coding), and then these sequences were assembled in groups with similar properties (clustering). This method is described in details in our previous work (Maity et al., 2014).

Clustering of F-D curves

In order to obtain and characterize different unfolding pathways, we developed a procedure for clustering of F-D curves obtained from the unfolding of CNGA1 channels from the C-terminal. This procedure is composed by the following steps:

1 – mappping of F-D curves in the (Lc,ΔLc) plane, where Lc is the contour length and ΔLc is the increase of contour length. Each force peak in a F-D curve is fitted with the Worm Like Chain (WLC) model (Bustamante et al., Science 1994; Rief et al., Science 1997) so to obtain the corresponding value of Lc of all force peaks present in the F-D curve. Only force peaks with an amplitude larger than 20 pN were considered. In this way given the sequence of force peaks $F_1,F_2\ldots F_{i,\ldots}.F_n$ we obtained the corresponding sequence of values of Lc i.e. Lc_1, Lc_2,… Lc_i …Lc_n. and each value of Lc_i is mapped in the (Lc,ΔLc) plane into the point (Lc_i,,$\Delta Lc_i = Lc_i$-Lc_{i-1}). As shown in Fig.1A, points in the (Lc,ΔLc) plane form some visible and well distinguishable clusters, indicating a high reproducibility of F-D curves for specific values of Lc, such as those around 234

nm.

2 – in order to identify regions of low and high reproducibility we computed the variance (Var) of ΔLc in binwidth of 2 nm, as shown in Fig.1B. We decided to start clustering points in the (Lc,ΔLc) plane where the variance of ΔLc was low, as for values of Lc around 229 to 245 nm, such as those within the black square in Fig.2C.

4 – computation of the histogram of the values of ΔLc within this region. As shown in Fig.1D the histogram of the values of ΔLc show three peaks, as shown by the three colors. The F-D curves corresponding to the colored points of Fig.1D have force peaks between 229 and 250 nm separated by 35, 42 and 52 nm, as shown in Fig.1E.

5 – from this clustering, three sets of F-D curves are obtained and are shown in Fig.1F-H. Each of these sets of F-D curves can be Characterized separately.

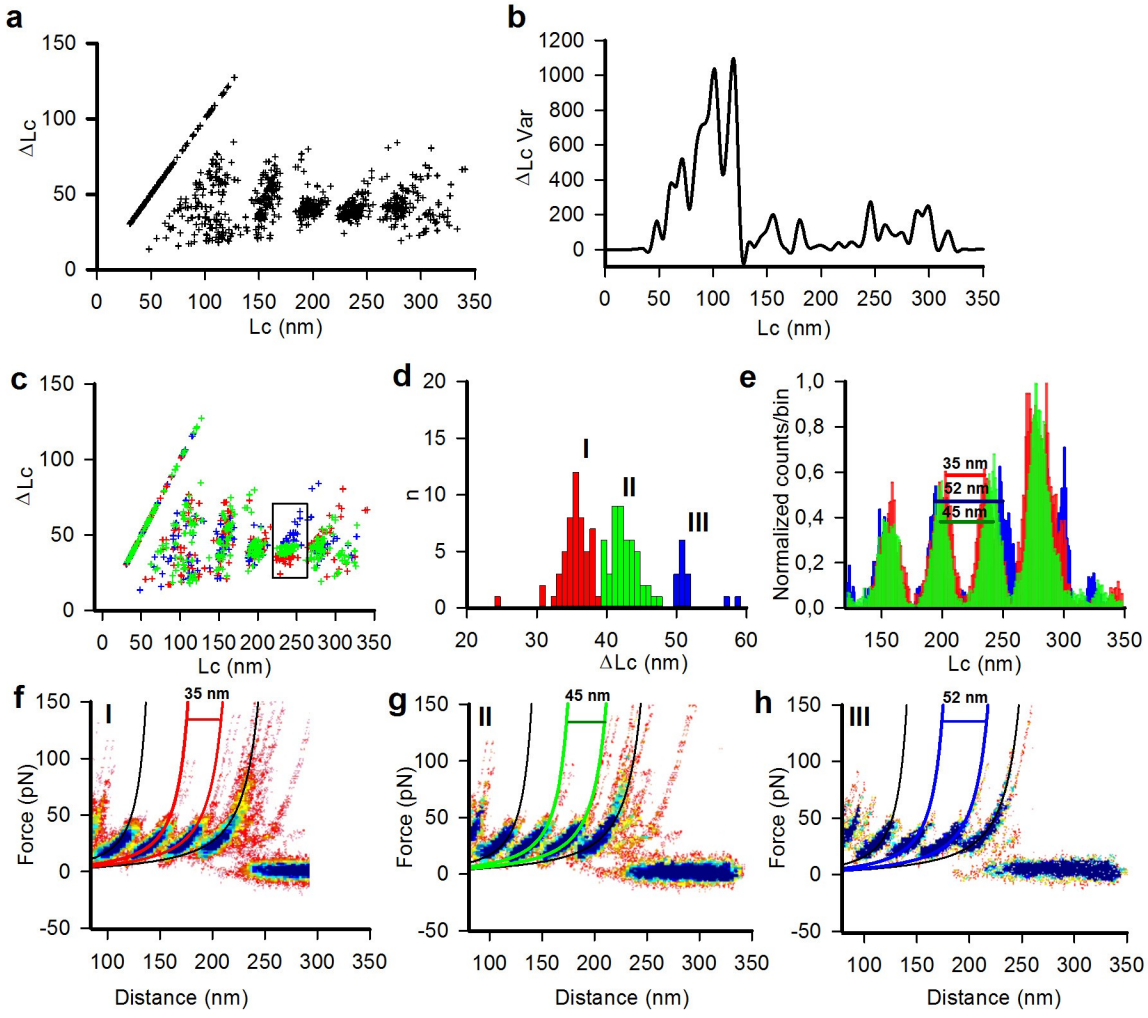


Fig.1. The clusterization in (Lc,ΔLc) plane *a: mapping of F-D curves in the closed state shown in fig.2a in the (Lc,ΔLc) plane as described in the Methods. b: the corresponding plot of the variance of ΔLc in a sliding window of 5 nm. c: clustering of points around 234 nm, in the (Lc,ΔLc) plane as described in the*

Methods (red, green and blue represent different clusters). d: ΔLc histogram of the points (n) within the black box of panel c, three different color define three different clusters (I,II,III in different color code) similar to panel c. e: normalized histogram of the Lc values of F-D curves of the three obtained clusters. f-h:density plot of F-D curves of the three clusters corresponding to three values of ΔLc of 35, 42 and 52 nm between the force peaks occurring with a value of Lc around 220 and 250 nm (point size 1 unit).

RESULTS

We deposited the oocytes on a freshly cleaved mica substrate and we used the AFM in liquid and in tapping mode[41] to image the membrane patches. After the localization of the membrane patch we moved the AFM tip over the imaged area. We performed SMFS and collected approximately 300.000 F-D curves in the absence of 2 mM cGMP and 200.000 F-D curves in the presence of 2 mM cGMP from noninjected (control) oocytes. We also collected more than 500.000 F-D curves from oocytes injected with the mRNA of CNGA1 channels in the absence and presence of 2 mM cGMP.

SMFS was performed in the open (in presence of 2 mM cGMP) and closed state (in the absence of cGMP) from oocytes injected with the mRNA of CNGA1 channels. As used in our previous method (MS1), we applied informatics analysis to identify F-D curves that correspond to the curves obtained only by the unfolding of CNGA1 channels.

Unfolding variability of CNGA1 channels in the open and closed state

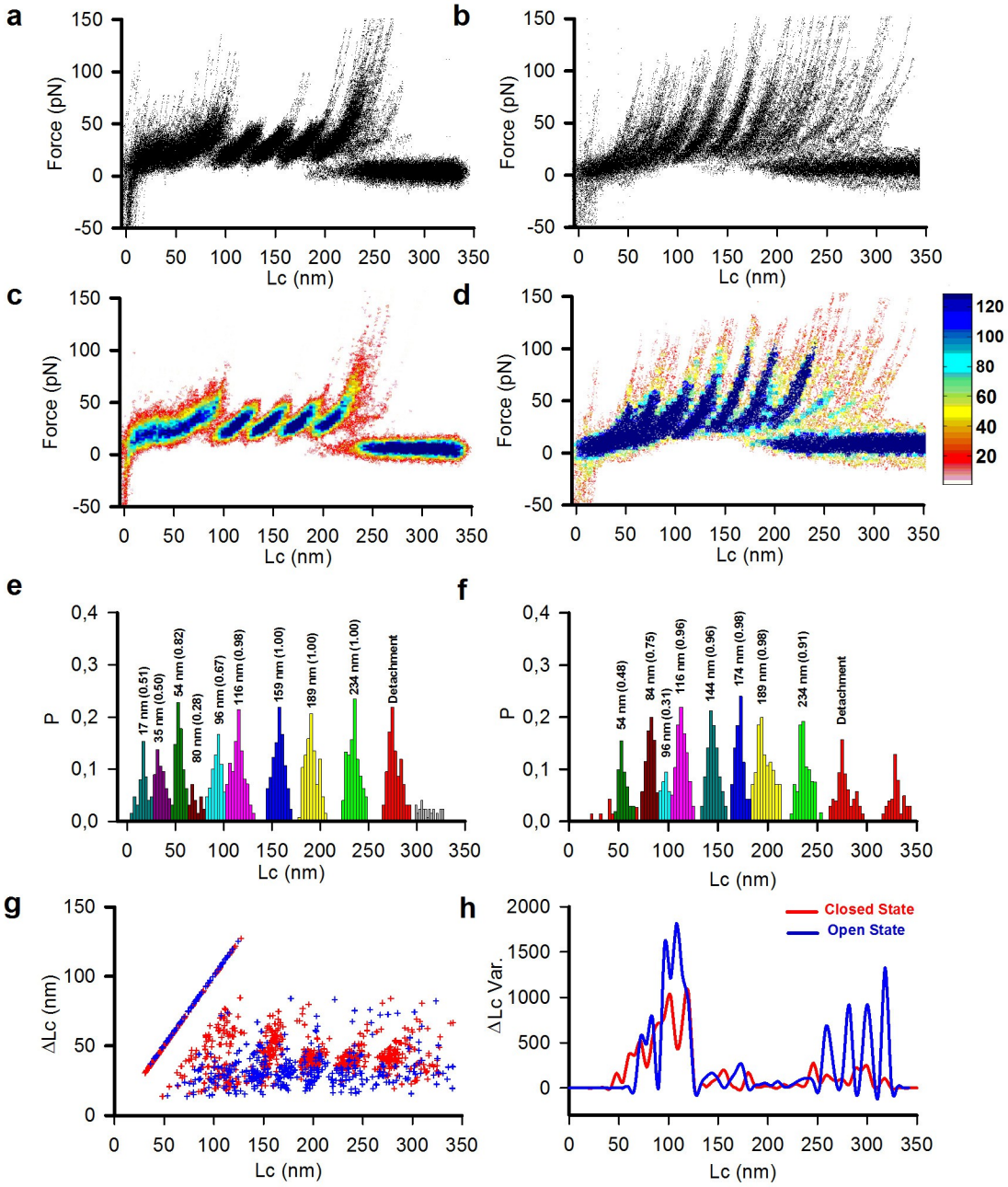


Fig.2. CNGA1 channel in open and closed state: *a and b: Superposition of 126 and 91 F-D curves obtained from the unfolding of CNGA1 channels in the closed and open state, respectively. c and d: density plots of the collected F-D curves in the closed and open state respectively. e: histogram of occurrence of the values of Lc obtained by fitting F-D curves with the WLC model in the closed state. f: as in e but in the open state. In e and f: number above each peak indicate the value of Lc and the probability (P) occurrence of that force peak. g mapping of F-D curves in the closed (red points) and in the open state (blue points) in the (Lc,ΔLc) plane as described in the Methods. h: the comparison of the variance of ΔLc for closed (in red) and open (in blue) states in a sliding window of 5 nm.*

Our SMFS experiments of CNGA1 channels identified 126 F-D curves obtained from the unfolding of CNGA1 from its C-terminal end in the closed state (Fig.2A) and 91 F-D curves in the open state (Fig.2B). The density plot of these F-D curves in the closed and open state are shown in Fig.2C and Fig.2D respectively. Fitting of these force peaks with the Worm Like Chain model (Bustamante et al., Science 1994; Rief et al., 1997) allows the estimation of the contour length (Lc) of these peaks and a better quantification of the unfolding of CNGA1 channels from the C-terminal end (Fig.2E and F). In the closed state, several force peaks with a value of Lc less than 100 nm could be seen and indeed peaks with a force between 20 and 50 pN were detected with a value of Lc of 17± 2.29, 35± 3.82, 54± 2.87, 80± 4.81 and 96± 3.87 nm with an occurrence probability of 0.51, 0.50, 0.82, 0.28, 0.67 respectively. All these force peaks did not appear in all F-D curves and they have an occurrence probability significantly lower than 1. In the closed state we identified four force peaks with a value of Lc of 116± 5.53, 159± 5.77, 189± 6.34 and 234± 6.91 nm which occurred with a probability very close to 1, i.e. these force peaks appeared in almost all F-D curves. These force peaks were followed by the detachment occurring usually with a value of Lc around 274± 7.49 nm.

In the open state, force peaks had larger forces between 50 and 80 pN and three main force peaks with a value of Lc less than 100 nm occurred with a value of Lc of 54± 3.10, 84± 4.18 and 96± 2.54 nm with an occurrence probability of 0.48, 0.75 and 0.31 respectively. Similarly to what observed in the closed state, these force peaks did not appear in all F-D curves. Moreover, in the open state we identified five force peaks with a value of Lc of 116± 4.46, 144± 5.10, 174± 4.74, 189± 7.22 and 234± 5.77 nm which occurred with a probability very close to 1, i.e. these force peaks appeared in almost all F-D curves. These force peaks were followed by the detachment occurring with a larger variability than that observed in the closed state usually, which will be analyzed in more details in Figure 4.

Using the WLC model for each F-D curve we computed the values of all Lc_i corresponding to the force peaks Fi. As a consequence each force peak F_1, F_2… $F_{i,...}F_n$ observed in a F-D curve has a specific value of Lc_i which differs by $\Delta Lc_i = Lc_i - Lc_{i-1}$ from the value of Lc_{i-1} corresponding to the previous force peak, so that each F-D trace can be mapped in a sequence of points in the plane (Lc,ΔLc) and if the F-D trace has n force peaks there will be n associated points in the (Lc,ΔLc) plane . F-D curves will be highly reproducible if when their force peaks are mapped in the (Lc,ΔLc) plane, the corresponding points accumulate in distinct clusters with a limited variability.. When this mapping is applied to F-D curves obtained from the unfolding of CNGA1 channels we

obtain the points in red and blue shown in Fig.2G for the closed and open state respectively. In the closed state, the points (Lc,ΔLc) group together in well visible clusters for values of Lc larger than 150 nm, but points with a value of Lc lower than 150 nm are rather dispersed and do not show any visible pattern. Visual inspection of the original F-D curves shows a high degree of superposition for values of tip-sample separation (distance) between 100 and 250 nm, but a much larger degree of variability for values of distance below 100 nm. We have quantified this variability by computing the variance of the value of ΔLc in binwidth of 5 nm. As shown in Fig.2H (red trace), the variance of the values of ΔLc increases for values of tss from 50 to 120 nm and then drops for values of tss above 130 nm. Force peaks with values of Lc below 120 nm correspond to the unfolding of the cytoplasmic domain of the CNGA1 channels (MS 1) and the observed high value of the variance of ΔLc suggests a significant structural variability of this domain.

In the open state, as shown in Fig.2B and Fig.2D, two additional force peaks are observed with large forces comprised between 50 and 100 pN. The corresponding points of these curves in the (Lc,ΔLc) plane show a larger variability (see blue points in Fig.2G). The variance of ΔLc (see blue traces in Fig.2H) from 130 to 250 nm which is expected to be the unfolding of transmembrane domain is comparable to what we have observed for closed state, but it shows significantly higher value for peaks beyond 250 nm which suggest higher order of variability in the detachment with respect to what we have seen for closed state.

Unfolding variability of the cytoplasmic domain of CNGA1 channels

Some force peaks do not appear in all F-D curves. This behavior is a strong indication for the presence of multiple unfolding pathways (Muller et al., 2002) and/or of the existence of distinct conformation of the cytoplasmic domain (Peng and Li 2008, Kotamarthi et al., 2013). We computed the distribution of the values of ΔLc in the interval of the values of Lc up to 116 nm (Fig. 3), corresponding to the cytoplasmic domain (MS1). These histograms have clear peaks in the closed state at about 17, 35, 54, 96 and 116 nm (Fig.3A,C). The sum of the values of ΔLc in each F-D curve has a single peak, as expected, at about 116 nm. In the open state, there are well resolved peaks at 54, 84 and 116 nm (Fig.3K) and similarly to what observed in the closed state. The sum of the values of ΔLc in each F-D curve has a single peak, as expected, at about 116 nm. These peaks appear with different frequencies indicating the existence of different unfolding pathways. Using the clustering procedure described in the Methods section, we found several clusters of F-D curves with a similar unfolding.

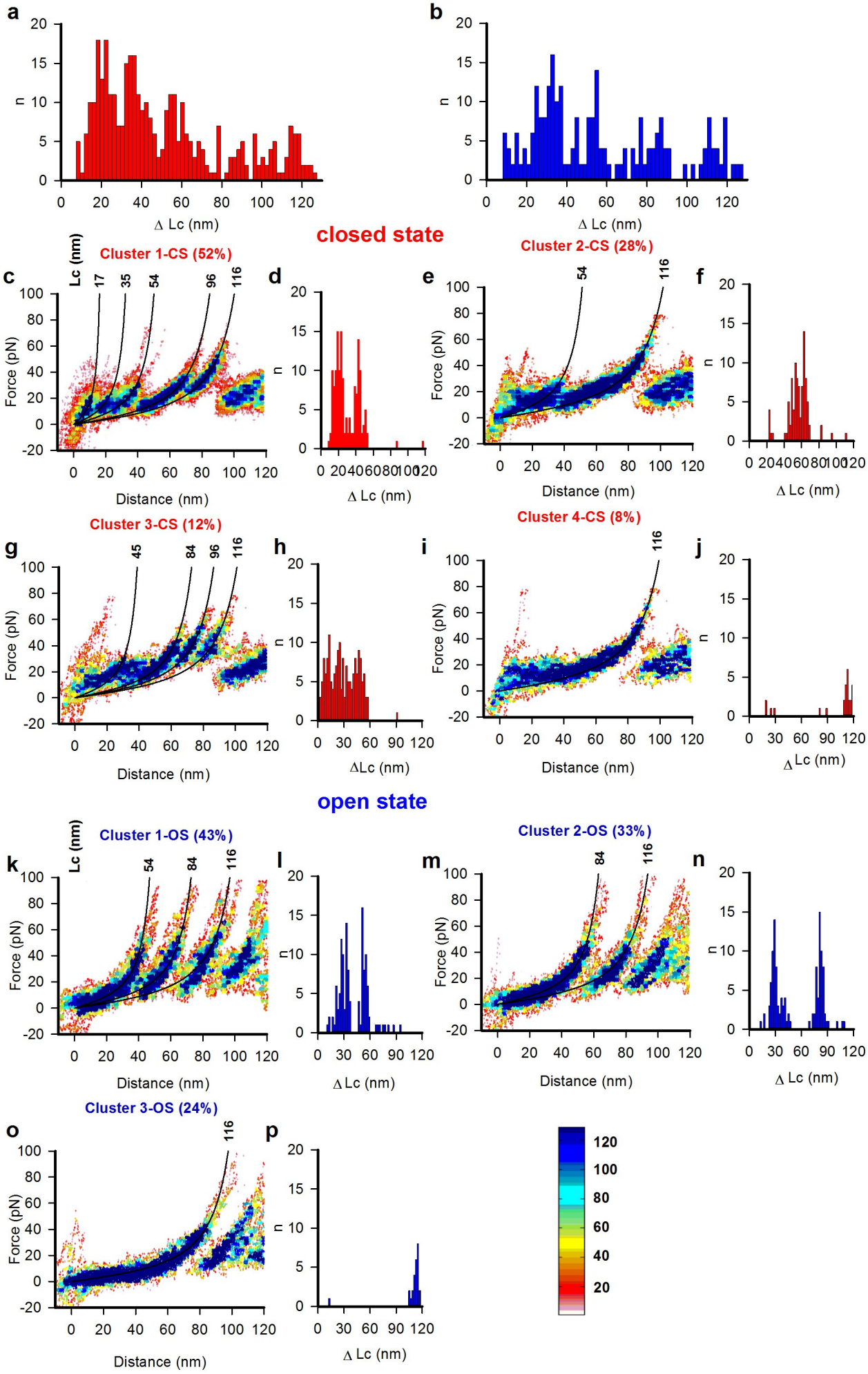

a
b
n
Δ Lc (nm)
closed state
Cluster 1-CS (52%)
Cluster 2-CS (28%)
Cluster 3-CS (12%)
Cluster 4-CS (8%)
Lc (nm)
17
35
54
96
116
45
84
Force (pN)
Distance (nm)
ΔLc (nm)
open state
Cluster 1-OS (43%)
Cluster 2-OS (33%)
Cluster 3-OS (24%)
c
d
e
f
g
h
i
j
k
l
m
n
o
p

Fig.3: Different unfolding pathways of the cytoplasmic domain composed by the CNB and the C-linker. *a-b: the ΔLc histogram (closed state in red and open sate in blue) in the entire C-terminal unfolding, i.e. from 0 to ~120 nm of Lc. A distribution of ΔLc value observed with certain conserved values.. c: Density plot of F-D curves from CNGA1 channels in the closed state for the cluster 1-CS (closed state here named CS) with frequency of occurrence of 52% and force peaks with an Lc of 17± 2.59, 35± 2.46, 54± 4.18, 96± 4.04 and 116± 3.32 nm; d: ΔLc histogram relative to cluster 1-CS. e: Density plot for cluster 2-CS with frequency of occurrence of 28% has force peaks with values Lc of 54± 1.94 and 116± 3.46 nm; f: ΔLc histogram relative to cluster 2-CS. g: Density plot for cluster 3-CS has a frequency of 12% with force peaks with values of Lc of 45± 3.45, 84± 2.80, 96± 3.69 and 116± 3.99 nm; h: ΔLc histogram relative to cluster 3-CS. i: Density plot for cluster 4-CS with only a force peak with a value of Lc 116± 3.40 nm, has a frequency of 8%. j: ΔLc histogram relative to cluster 4-CS. k: Density plot of F-D curves from CNGA1 channels in the open state for the cluster 1-OS (point 1 unit; open state here named OS) cluster 1-OS with force peaks with an Lc of 54± 3.10, 84± 3.60 and 116± 4.80 nm and a frequency of occurrence of 43%; i: ΔLc histogram relative to cluster 1-OS. m: Density plot for cluster 2-OS with force peaks with values Lc of 84± 3.51 and 116± 3.81 nm has a frequency of 33%; n: ΔLc histogram relative to cluster 2-OS.o: Density plots for cluster 3-OS with only a force peak with a value of Lc 116± 3.93 nm has a frequency of 24%; p: ΔLc histogram relative to cluster 3-OS. x: sum of the values of ΔLc in each F-D curve in the open and closed state. In all panels solid black curves represent the fitting with the WLC model and the number indicates the corresponding value of Lc. In all density plots points=1 unit and in all ΔLc histogram bin= 2 nm.*

In the closed state we found four major clusters: cluster 1-CS has a frequency of appearance of 52% and it has peaks at 17± 2.59, 35± 2.46, 54± 4.18, 96± 4.04 and 116± 3.32 nm (Fig-3A); the cluster 2-CS appear with frequency of 28% has peaks at 54± 1.94 and 116± 3.46 nm (Fig-3B); cluster 3-CS possessed 12% of the total events, has peaks at 45± 3.45, 84± 2.80, 96± 3.69 and 116± 3.99 nm (Fig-3C). And then the less frequent cluster 4-CS appears in 8% of all curves and it has a peak at 116± 3.40 nm (Fig-3D). In the open state we found three major clusters: cluster 1-OS with highest frequency of 43% and has force peaks at 54± 3.10, 84± 3.60 and 116± 4.80 nm (Fig.3E); cluster 2-OS with a force peaks at 84± 3.51 and 116± 3.81 nm and frequency of 33% (Fig.3F) and cluster 3-OS with only the force peak at 116± 3.93 nm and frequency of 24%. First observation shows that in the closed state the peaks at 17± 2.59 and 35± 2.46 nm are not observed in the open state, and have a magnitude of low force of 20-35 pN expected from the unfolding of very short α-helices (Reif 1999). The force peak with an Lc of 54± 2.87 nm, in the closed state was either absent or had a force lower than 35 pN, while in the open state it had larger forces often above 45 pN. The force peak with an Lc equal to 84 ± 3.60 nm was seen primarily in the open state with a frequency of 76% and was rarely observed in the closed state.

Variability of detachment

Both the C- and N- terminal of CNGA1 channels are intracellular, so theoretically when we pull from C-terminal the unfolding should end until it reaches the beginning of S1 (residue near K105; ~234 nm), since there will be no anchoring force in the N-terminal. 12% of the curves show exactly to the length that we have expected, i.e. 247± 5.63 nm (Fig.4A); but when we look to the all F-D curves we observed that about 88% of them show an additional force peak at around 274± 7.49 nm (Fig.4B) which is the full length of the channel (690 aa). This behavior is attributed to the variability of the final detachment, which occurs concomitantly with the force peak that has an Lc of approximately 234± 5.77 nm (Fig.4B) or which requires additional pulling (Fig.4B). If the N-terminal is not interacting with the C-terminal of a neighboring subunit or with other domains of the same subunit, detachment of the cantilever tip can occur simultaneously with the unfolding of S1 and in this case the final force peak is expected to have a value of Lc around 234 nm, corresponding to 585 aas (690-585=aa105) as expected from the unfolding of aa1 from the C-terminal (D690) to the end of the S1 α-helix, with the force peak with an Lc of 247± 5.63 nm corresponding to the unfolding of S2 and S1 (from R216 to E100). If the N-terminal interacts with the C-terminal of a neighboring subunit or with other domains of CNGA1 channels an additional pull is required for the final detachment, as observed for the F-D curves of Fig.4B. In this case, detachment occurs with a force peak with a value of Lc around 274± 7.49 nm. Short and longer F-D curves are very similar, but not entirely: indeed short traces have well distinguishable force peaks with a value of Lc about 54± 1.95 and 84± 2.58 nm with probability equal to 0.93 and 0.79 respectively and longer F-D curves do no shows well distinguishable force peaks around 54 and 84 nm and the force peak at 35± 6.89 nm has a probability of 0.41. These observations suggest that if the unfolding of the CNGA1 channel requires force peaks with an Lc at 54± 1.95 and 84± 2.58 nm, its C-terminal is less likely to interact strongly neither with the N-terminal or other domains of neighboring subunits.

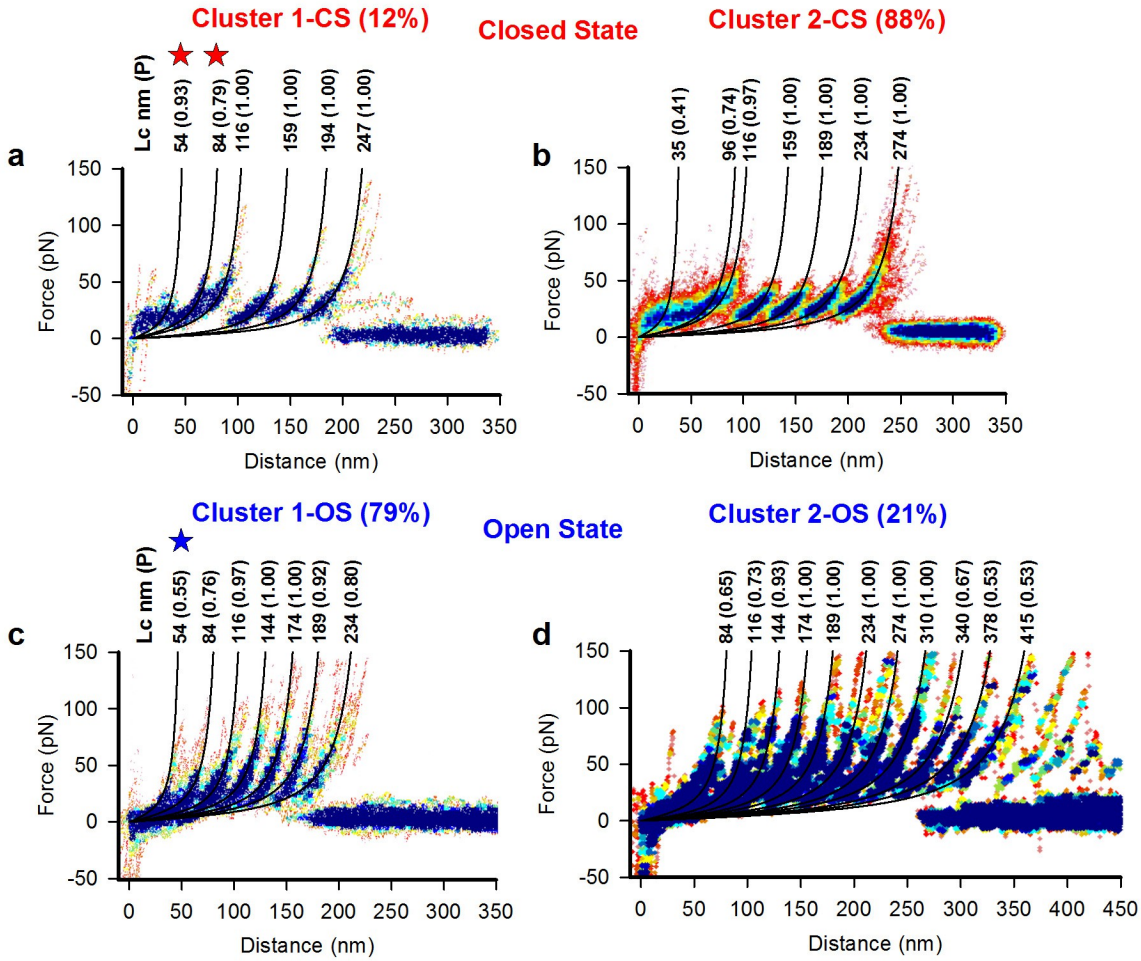


Fig.4: Variability of detachment. a-b*: Density plots of F-D curves from the unfolding CNGA1 channels in the closed state ending with a force peak with a value of Lc around 247± 5.63 and 274± 7.49 nm respectively (point 1 unit). F-D curves in (a) have the detachment coinciding with the unfolding of the last transmembrane segment S1, while for those in (b) detachment occur with an extra peak. F-D curves in (a) have initial force peaks with a value of Lc of 54± 1.95 and 84± 2.58 nm (see red stars), not present in F-D curves in (b). Clusters in (a) and (b) occurred with a frequency approximately of 12 and 88%.* ***c-d****: Density plots of F-D curves from the unfolding CNGA1 channels in the open state ending with a force peak with a value of Lc of 234± 5.77 and significantly longer respectively (point size 1 unit in c and 1.5 unit in d). F-D curves in (c) have the detachment coinciding with the unfolding of the last transmembrane segment S1, while for those in (d) detachment occur with a large value of Lc. Almost 50 % of F-D curves in (c) have a force peak with a value of Lc of 54± 3.10 nm (see blue star), not present in F-D curves in (d). Clusters in (c) and (d) occurred with a frequency approximately of 79 and 21%. In all panels solid black curves represent the fitting with the WLC model and the number indicates the corresponding value of Lc.*

In the open state, some of F-D curves had the final detachment with a force peak with a value of Lc about 234± 2.85 nm (Fig.4C) with a frequency of 79%, but other F-D curves were significantly longer (Fig.4D). In these F-D curves, force peaks appeared replicated, as obtained from the unfolding of not a single CNGA1 subunit, but from the concatenation of two subunits. This possibility is

not surprising, as it is known that in the open state the N-terminal of one subunit strongly interacts with the C-terminal of a neighboring subunit (Gordon et al., 1997, Rosenbaum & Gordon, 2002) so that two neighboring subunits are almost linked together: under these conditions when the cantilever tip unfolds one CNGA1 subunit an additional subunit can subsequently unfolded. Therefore our results indicate a possible moderate interaction between the N- and C- terminals in the closed state, which is clearly potentiated in the open state. In the open state we never observed "duplicated" F-D curves when pulling with the constructs CNGA1-N2B-HisTag, because in these constructs the added domains do not allow the N- and C-terminals to interact as in the CNGA1 channels (result not shown, see also MS1). The interaction between the C- and N- terminals in the open state can explain also the lower success of SMFS to obtain good and complete F-D curves (Fig.4), because the cantilever tip has a limited access to the C-terminal.

In the open state, shorter F-D curves have a force peak with an Lc around 54± 3.10 nm with a value of probability equal to 0.55, this force is not observed in longer F-D curves, suggesting that – similarly to what observed in the closed state - when the unfolding of the CNGA1 channel requires a force peak with an Lc around 54± 3.10 nm, its C-terminal is less prone to interact with the N-terminal of a neighboring subunit. This possibility is also supported by the observation that F-D curves obtained from the tandem construct CNGA1-CNGA1 do not show force peaks with values of Lc around 54 nm (MS1).

Variability of the unfolding of the transmembrane domain

As shown in Fig.2H the variance of ΔLc is relatively small between approximately 120 and 245 nm, corresponding to the unfolding of the transmembrane domain. In Fig.3 and 4 we have seen that the significant variability of the unfolding of the cytoplasmic domain and of the detachment could be associated to structural properties and concomitant different unfolding pathways. It is possible, therefore, that structural properties of both the C- and N-terminal domains of CNGA1 channels are influenced at some extent by the specific unfolding of the transmembrane domains S1-S6. Therefore we analyzed in details the variability of ΔLc during the unfolding of the transmembrane domains and see whether different values of ΔLc between force peaks associated to the unfolding of the transmembrane segments could influence the appearance of force peaks in the cytoplasmic domain. The clustering described in the methods has identified clusters of F-D curves (six for the closed state and five for the open state) with significant differences of the value of ΔLc in the unfolding of the transmembrane domain.

As shown in Fig.5, in the closed state, the distribution of the values of ΔLc associated to the force peaks with an Lc around 189 and 234 nm has three main peaks (Fig.5A and B) with values of 35, 45 and 52 nm. Assuming that the length of a stretched aminoacid is 0.4 nm (MS1), the unfolding of the S2 and S1 could involve about 88 or 113 or 130 aas. F-D curves in which the unfolding of S2 and S1 requires stretching approximately 88 aas do not have clear and large force peaks in the cytoplasmic domain and show only a weak force peak with a value of Lc equal to 35± 5.87 nm, clearly visible in the density plot (Fig.5B,C). F-D curves in which the unfolding of S2 and S1 is achieved by stretching a larger number of aas, i.e. of approximately 113 and 130 aas, have all a force peak in the cytoplasmic domain with an Lc equal 96± 3.87 nm (Fig.5D) and F-D curves with a value of ΔLc about 52 nm have several force peaks in the cytoplasmic domain with values of Lc equal to 35± 1.95, 84± 2.10 and 96± 2.58 nm (Fig.5E). Similarly, clusterization of the values of ΔLc associated to the values of Lc around 189 nm provides three main clusters (Fig.5F and G): F-D curves with a value of ΔLc around 25 nm (63aa) in the cytoplasmic domain have a force peak at 96± 2.18 nm with a probability of 0.82 and occasional force peaks at 35± 1.90 and 54± 3.41 nm but those with a value of ΔLc around 30 nm have force peak at 96± 2.81 with a probability of 0.91. F-D curves with a value of ΔLc around 49 nm have force peaks at 35± 2.35, 84± 2.10 and 96± 1.85 nm.

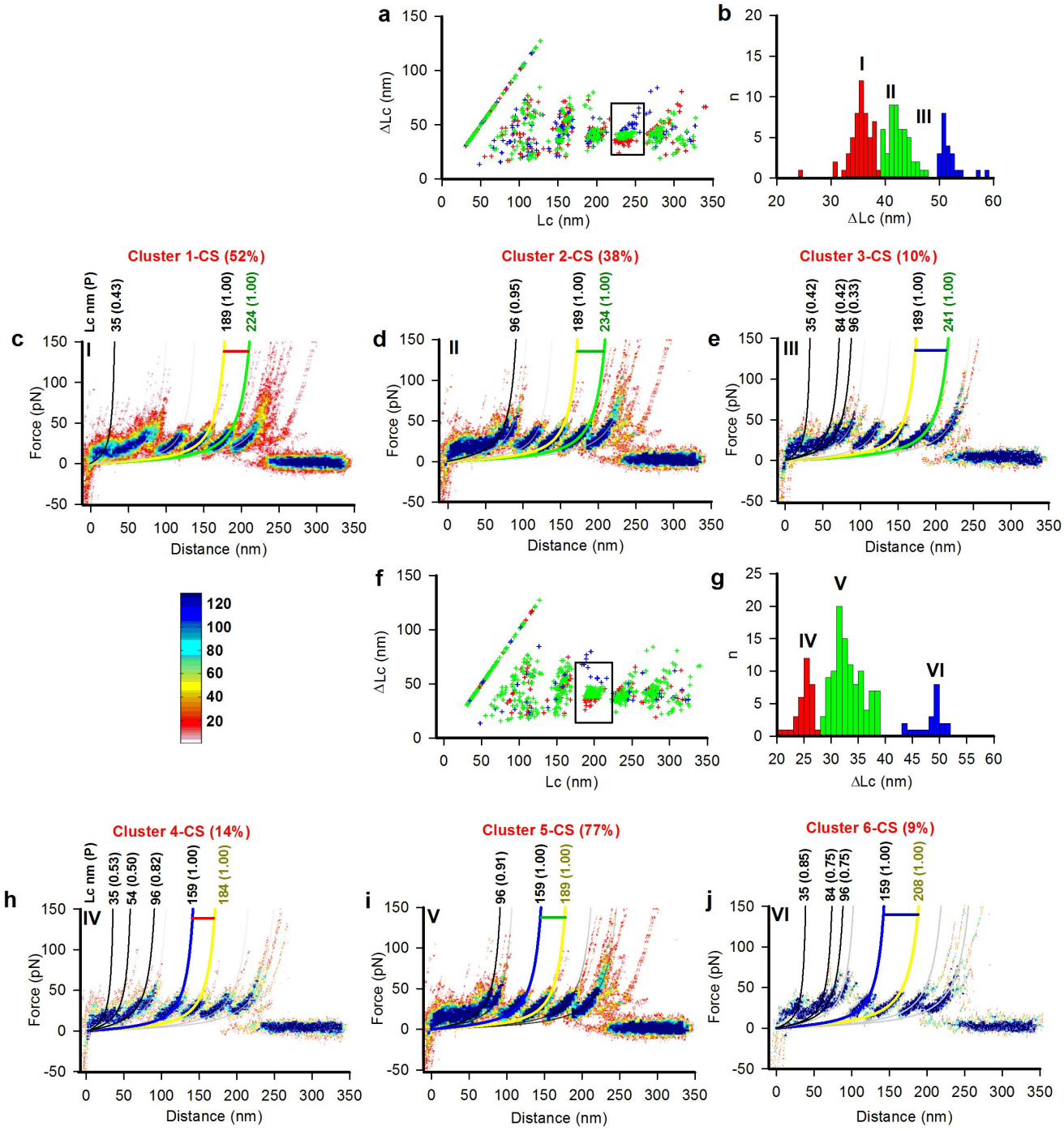


Fig.5: Variability of the unfolding of the transmembrane domain in the closed state. a: *the clusterization around 234 nm (black box) in (Lc-ΔLc) plane;* ***b:*** *the ΔLc histogram (bin 2 nm) from all the points inside the box from (a), different colours are for corresponding clusters from (a);* ***c-e****: the density plot of F-D curves of those three cluster from (a), showing three different unfolding pathways of the transmembrane domain.* ***c:*** *cluster 1-CS, has a peak at Lc of 224± 6.77 nm with ΔLc of 35 nm has a force peak with an Lc of 35± 5.87 nm and a frequency of 52%; d: cluster 2-CS with force peak at 234± 2.18 nm of Lc with ΔLc of 45 nm, appearing with a force peak with values of Lc of 96± 3.87 nm and frequency of 38%; e: cluster 3-CS has the peak at Lc value of 241±1.90 nm with a ΔLc value of 52 nm has force peaks with values of Lc of 35± 1.95, 54± 2.10 and 96± 2.58 nm; all these clusters have common force peaks at a Lc values around 116, 159, 189 and the*

detachment around 274 nm. f: the clusterization around 189 nm (black box) in (Lc-ΔLc) plane; g: the ΔLc histogram (bin 2nm) from all the points inside the box from (f), different colours are for corresponding clusters from (f); h-j: the density plot of F-D curves of those three cluster from (f), showing three different unfolding pathways of the transmembrane domain. h: cluster 4-CS, has a peak at Lc of 184± 2.54 nm with ΔLc of 25 nm has a force peak with an Lc of 35±1.90, 54± 3.41 and 96± 1.85 nm and a frequency of 14%; i: cluster 5-CS with force peak at 189±2.94 nm of Lc with ΔLc of 30 nm, appearing with a force peak with values of Lc of 96± 2.81 nm and frequency of 77%; j: cluster 6-CS has the peak at Lc value of 208± 2.19 nm with a ΔLc value of 49 nm has force peaks with values of Lc of 35± 2.35, 84± 2.10 and 96± 1.85 nm; all these clusters have common force peaks at a Lc values around 116, 159, 234 and the detachment around 274 nm. d All the density plot of each cluster have fitted with WLC where colour chains represent the peak under clusterization; while the black chain represents the peak appearing in the cytoplasmic domains (point size 1unit). The numbers corresponding to the WLC chain are the Lc value with probability.

The results of Fig.5 shows that in the closed state when the value of ΔLc corresponding to the unfolding of the transmembrane segments S3-S4 or S2-S1 is large there are more force peaks during the unfolding the cytoplasmic domain.

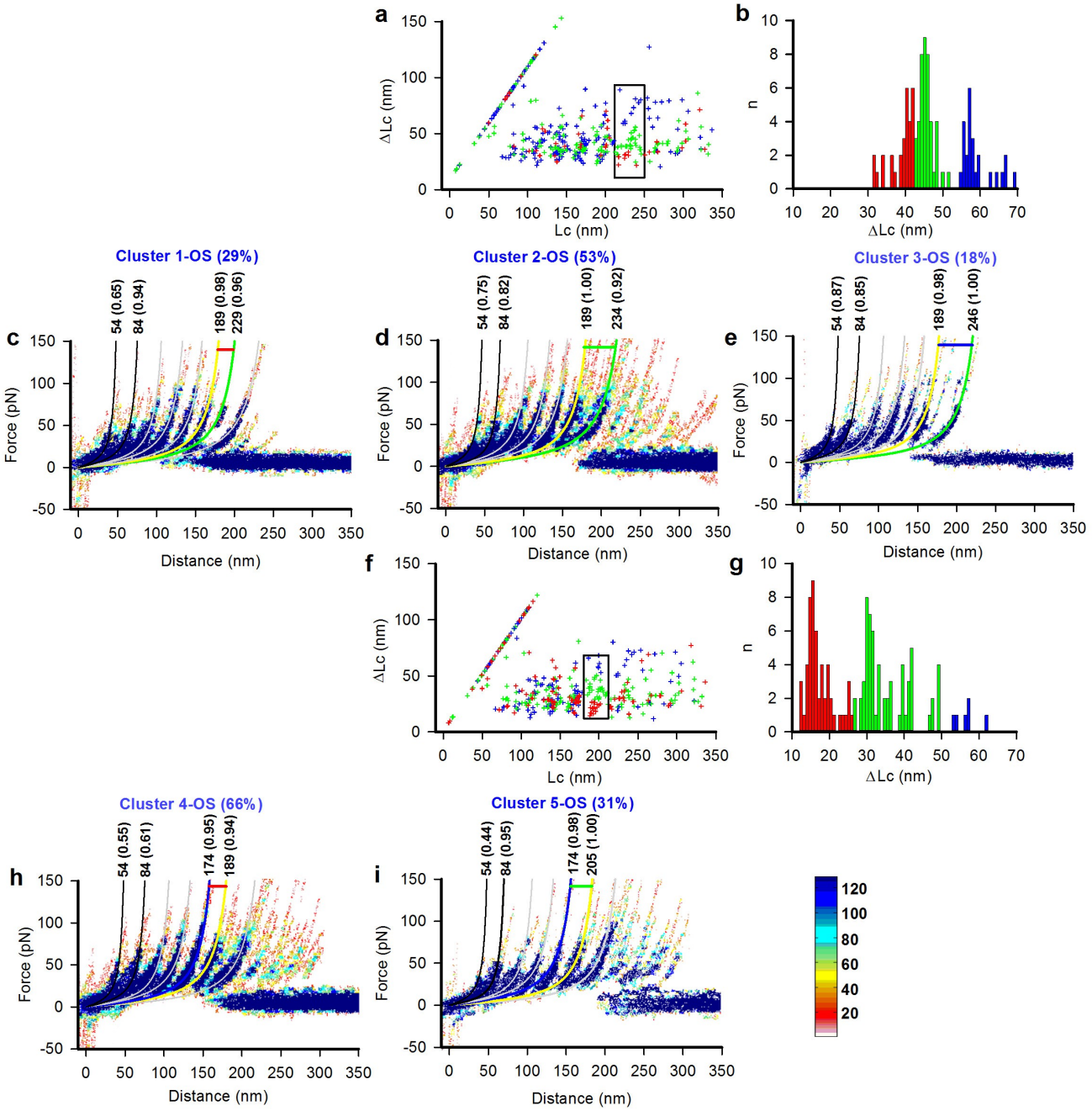


Fig.6: Variability of the unfolding of the transmembrane domain in the open state. a: *the clusterization around 234 nm (black box) in Lc-ΔLc plane;* ***b:*** *the ΔLc histogram (bin 2 nm) from all the points inside the box from (a), different colours are for corresponding clusters from (a);* ***c-e:*** *the density plot of F-D curves of those three cluster from (a), showing three different unfolding pathways of the transmembrane domain.* ***c:*** *cluster 1-OS, has a peak at Lc of 229± 2.33 nm with ΔLc of 40 nm has a frequency of 29%;* ***d:*** *cluster 2-OS with force peak at 234± 5.75 nm of Lc with ΔLc of 45 nm, with a frequency of 45%;* ***e:*** *cluster 3-OS has the peak at Lc value of 246±2.89 nm with a ΔLc value of 57 nm has frequency of 18%; all these clusters have common force peaks at a Lc values around 54, 84, 116, 144, 174 and 189 nm.* ***f:*** *the clusterization around 189 nm (black box) in Lc-ΔLc plane;* ***g:*** *the ΔLc histogram (bin 2 nm) from all the points inside the box from (f), different colours are for corresponding clusters from (f);* ***h-i:*** *the density plot of F-D curves of those major two clusters from (f), showing two different unfolding pathways of the transmembrane domain.* ***h:*** *cluster 4-OS, has a peak at Lc of 189±4.41 nm with ΔLc of 15 nm has a frequency of 66%;* ***i:*** *cluster 5-OS with force peak at 205±5.88 nm of Lc with ΔLc of 30 nm, appearing with a*

frequency of 31%; All these clusters have common force peaks at a Lc values around 54, 84, 116, 144, 174 and 234 nm. All the density plot of each cluster have fitted with WLC, where colour chains represent the peak under clusterization; while the black chain represents the peak appearing in the cytoplasmic domains (point size 1unit). The numbers corresponding to the WLC chain are the Lc value with probability.

As shown in Fig.6 , in the open state, the clusterization of the values of ΔLc associated to the values of Lc around 234 nm provides three main clusters (Fig.6A) with ΔLc values of 40, 45 and 57 nm, but F-D curves corresponding to these three clusters (Fig.6B-E) have common force peaks in the cytoplasmic domain both around 54 and 84 nm. Also the clusterization of the values around 189 nm provides two main clusters (Fig.6F-I) with ΔLc values of 15 and 30 nm, there are some traces which shows a ΔLc values above 50 nm (blue points in box Fig.6F) but are very few in number and too diverse to form a cluster. F-D curves corresponding to these two clusters (Fig.6H and I) have force peaks in the cytoplasmic domain both around 54 and 84 nm.

The effect of different pulling speeds

SMFS experiments described in Figures 1-6 were obtained at a pulling speed of 500 nm/sec. In order to verify and test whether the value of Lc of observed force peaks could be ascribed to the unfolding of specific stretches of aas, we have performed identical SMFS also when the pulling speed was varied from 200 nm/s up to 1500 nm/s (Fig 7 A-D). At the lower pulling speed of 200 nm/s, the magnitude of force peaks was slightly lower and on average was 45±5 pN, and the observed force peaks had values of Lc of 110± 8.61, 163± 4.96, 195± 7.38, 234± 6.52, 276± 7.45 nm respectively, i.e. very similar to those seen at the usual speed of 500 nm/s. At a pulling speed of 1μm/s, the magnitude of the average force peaks was 102± 20 pN with values of Lc equal to 98± 4.98, 118± 5.64, 161± 6.62, 189± 6.58, 234± 5.76 and 268± 9.96 nm. Using a speed of 1500 nm /s, the peaks are 96± 6.10, 125± 7.83, 159± 4.58, 196± 7.56, 240± 9.10 and 274± 10.66 nm. We tried to go beyond 1500 nm/s, but rate of success is significantly low, as we expect; this effect can be due to the long chain of CNGA1 channels and the limitation of nonspecific bond strength of the molecule to the AFM tip. We have observed for each pulling speed, the unfolding pathway for CNGA1 channels conserved by 80-95% to what we have observed at a pulling speed of 500 nm/s. Only difference we have seen is the breaking force for each peak. There is a significant increase in breaking force on increasing of the pulling speed (Fig7 E,F), as we can expect from previous work of Muller (Muller et al., 2002).

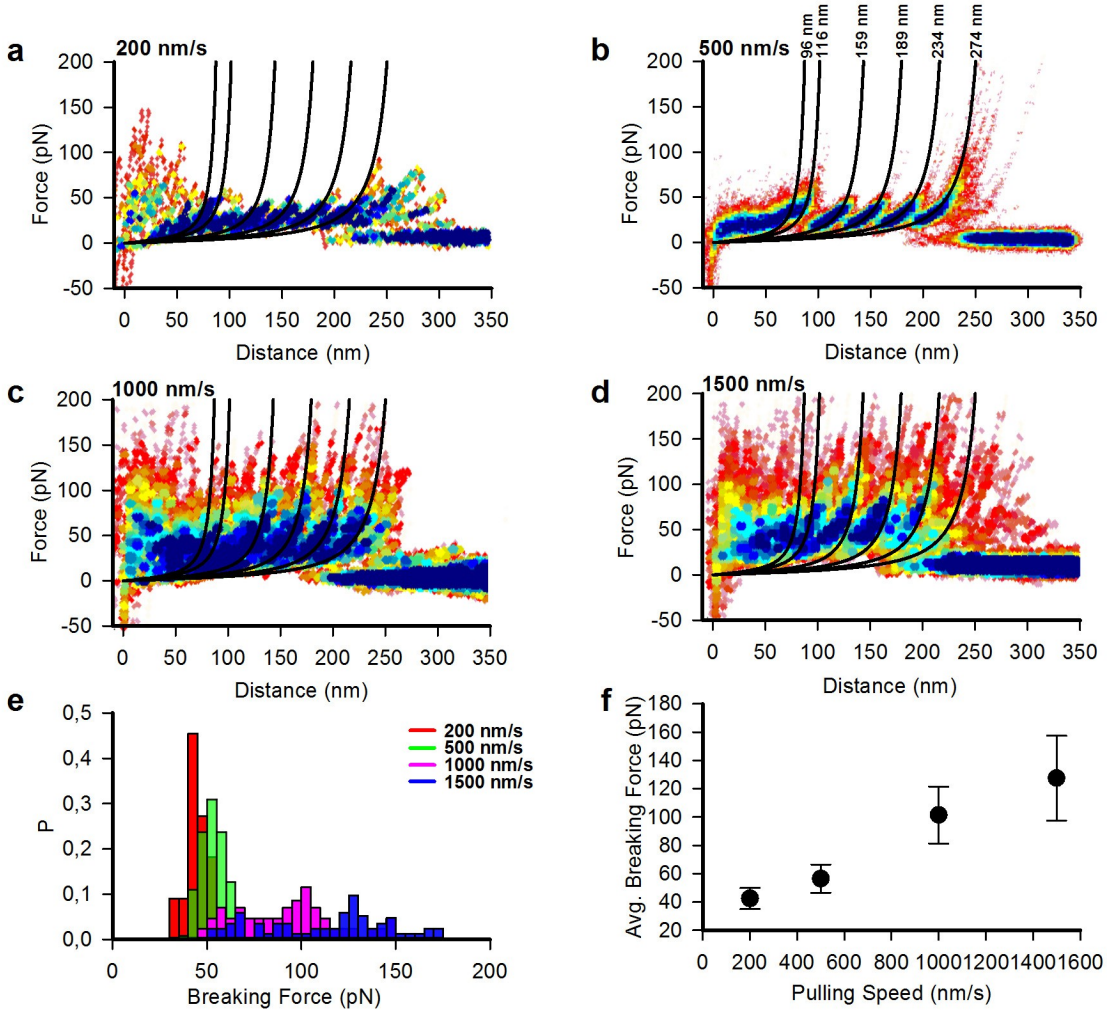


Fig.7: Pulling speed dependency of CNGA1 unfolding. a-d: *Density plot of F-D curves from CNGA1channels unfolding with a pulling speed of 200 nm/s, 500 nm/s, 1000 nm/s and 1500 nm/s respectively (point size 1.5 units); the black lines represents the WLC fitting with Lc values of 96, 116, 159, 189, 234 and 274.* ***e:*** *comparison breaking force distribution for each pulling speed. The red, green, pink and blue color represent the pulling speed of 200, 500, 1000, 1500 nm/s respectively.* ***f:*** *the plot of average breaking force vs pulling speed.*

DISCUSSION

SMFS has been extensively used to characterize the mechanical unfolding of many membrane and water-soluble proteins (Kedrov et al., 2007; Engel and Gaub, 2008; Hensen and Muller, 2013). Despite membrane proteins and water-soluble proteins have similar tertiary structures, their mechanical properties are significantly different (Bosshart et al., 2012; Dietz and Rief, 2004; 2006; Sapra et al., 2009; Thoma et al., 2012, Hensen and Muller, 2013). The common observation is that membrane proteins have multistep folding-unfoldindg, where each unfolding step is formed by a major unfolding barrier. In contrast, water-soluble proteins usually unfold in one major step, which suggests that only one major unfolding barrier (Hensen and Muller, 2013).

We have used SMFS to perform experiments aiming to recover structural information on CNGA1 channels in a physiological-like environment, thus avoiding purification and incorporation in an artificial lipid membrane. F-D curves obtained from the unfolding of CNGA1 channels expressed in *Xenopus laevis* oocytes were identified by using an informatics analysis validated by the use of a variety of finger-prints (MS1). Our results show a specific degree of unfolding variability when the cytoplasmic and the transmembrane domains are unfolded. The unfolding variability, observed in our experiments, is reminiscent of what previously observed during the unfolding of the lisozyme T4L* (Peng & Li PNAS 2008) and the LBP (Kotamarthi et al., 2013) in which protein unfolding occurred either in a single or in two steps. Both these proteins are multi-domain proteins composed by α-helices and β-strands. CNGA1 channels have a large cytoplamic domain (Zagotta et al., 2003; Mari et al., 2011) of about 290 aa (from N400 to D690) - organized by α-helices (named A'-F') of the C-linker (from N400 to E482) and three α-helices (named A-C) and eight stranded anti-parallel β-rolls (named 1-8) of the CNB domain (from A483 to N610)- as well as a transmembrane domain composed by 6 α-helices (named S1-S6) and it is not surprising that its unfolding is more complex than the unfolding of rhodospin (Sapra et al., 2006; Kawamura et al., 2013), composed primarily by seven transmembrane α-helices with very short unfolded N- and C-terminals and of purely soluble proteins, such as the lysozyme T4L* and LBP. Our observation shows clearly that CNGA1 channel shows different unfolding pathway in both the water-soluble part, i.e. the cytoplasmic domain, as well as in the transmembrane domain. CNGA1 channels, despite many attempts and efforts, have never been crystallized and a possible explanation for their resistance to form well ordered crystals able to diffract X-rays is the heterogeneity of their native conformations. This observation suggests that the unfolding variability of CNGA1 channels, here described, could be caused by the combination of several distinct native conformations and a ragged free-energy landscape.

Unfolding variability of the cytoplasmic domain of CNGA1 channels

In the open state, unfolding forces are usually larger by about 10-20 pN and the cytoplasmic domain is unfolded in a single step (Fig.3O) in 24% of F-D curves and in two (Fig.3M) and three steps (Fig.3K) in 33 and 43% of F-D curves, respectively. In the closed state, in about 8% of F-D curves the cytoplasmic domain unfolds in a single step (Fig.3I) and in about 28% of F-D curves the unfolding of the cytoplasmic domain occurs in two steps (Fig.3E) and in the remaining 64% of F-D curves more than 3 steps are observed (Fig.3C and Fig.3G). In the closed state several F-D curves have three force peaks at Lc 17± 2.59, 35± 2.46 and 54± 4.18 nm with low magnitude of 25-35 pN, the corresponding mapping of Lc to the aa number are 43, 88 and 135 respectively and consistent with the unfolding of short α-helices or partially unfolded (MS1). Most of F-D curves have a force peak at 54± 4 nm with an amplitude magnitude ~55 pN. These experimental evidences obtained from SMFS are consistent with previous electrophysiological observations suggesting that in the absence of cGMP, the C-helix of CNB domain does not have a well defined conformation (Mazzolini et al., 2002), which become properly folded following cGMP. Therefore we propose that in the closed state, when the CNGA1 channel has a partial folding can be unfolded in several steps, in which the force peaks of about 30 pN at 17 and 31 nm correspond to the complete unfolding of C-terminal up to the C-helix (from D690 to D588). The force peaks with an Lc of 54, 84 and 96 nm are likely to correspond to the partial unfolding of the β-roll forming the CNB domain and of the C-linker. When the cytoplasmic domain is properly folded its unfolding is expected to occur in one or two steps as for several soluble proteins such in the lysozyme T4L* and LBP.

As shown in Fig.5, the unfolding of the cytoplasmic domain is somehow coupled to the unfolding of the transmembrane domain. The peak at 234 nm (585 aa) shows a significant variability ranging from 224 nm to 241 nm, indicating that the complete unfolding of S1and S2 can occur at an amino acid varying from E130 (aa690-560) to N87 (aa690-603) at the N-terminal . This observation suggests that either N-terminal have some secondary structure in a mechanical continuity with S1 or S1 can have varying number of aas folded as an α-helix (Fig8 e). Similarly, the force peak with an average value of Lc of about 189 nm (473 aa), the measured values of Lc varied from 182 nm to 208 nm (460 aa to 520 aa), suggesting a possible variability of coupling of S3 and S4 (Fig.8C,D) in which when S2 coupled to S1, the unfolding of the two α-helices S3 and S4 occurs with a value of Lc ranging from 184 to189 nm. But when S2 is mechanically more coupled to S3, the unfolding force has the corresponding value of Lc at 208 nm and in this case the unfolding of S3 and S4 occurs with a concomitant partial unfolding of S2 .In both these cases, when ΔLc has a higher value (Fig.5E,J), the frequency of appearance of peaks at 35± 1.95, 84± 2.10 and 96± 2.58 nm, increases.

This observation hints to a possible influence of how the transmembrane domain is folded - and consequently unfolded - on how the cytolasmic domain is folded and unfolded.

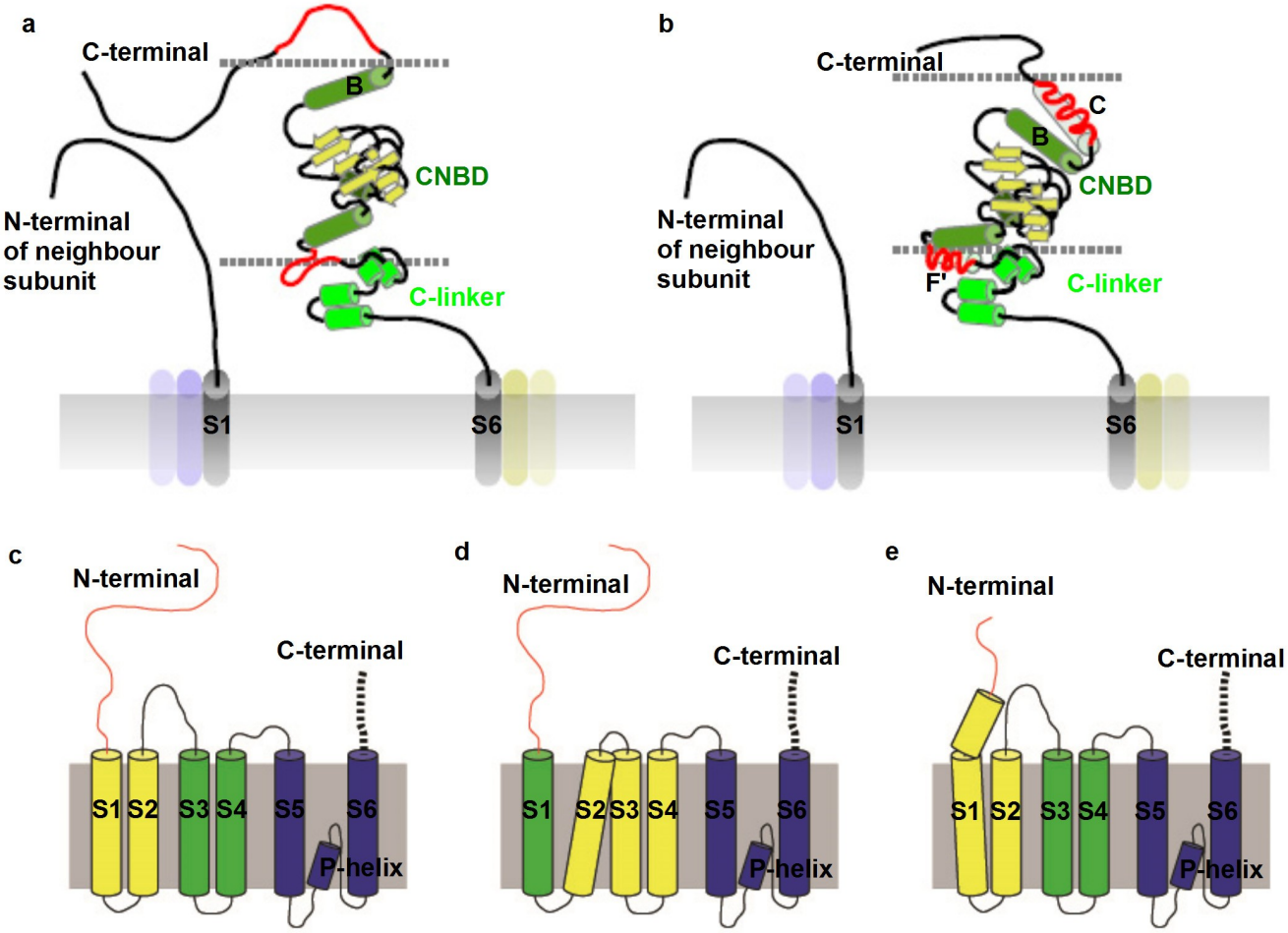


Fig.8: cartoon representation to summarize the experimental observations. a,b: *represents the possible conformational changes in the cytoplasmic domain inhibit the N-C terminal interactions;* ***a:*** *showing a free chain movement can increase the possibility for C-terminal of one subunit to interact with a N-terminal of other subunit. While in* ***(b)*** *the possibility decreases with a partial or complete folding of the c-terminal.* ***c-e:*** *a cartoon representation to explain the possible conformations that lead to different unfolding pathways; different colors for the transmembrane site represent different unfolding steps.*

Unfolding variability of the detachment

Approximately 12% of F-D curves obtained from the unfolding of CNGA1 channels in the closed state have the detachment occurring concomitantly with the unfolding of the S2-S1 transmembrane segments (Fig.4A), occurring with a value of Lc of around 247± 5.63 nm. The remaining F-D curves have the detachment after the unfolding of the S2-S1 segments suggesting that the N-termi-

nal of the unfolded CNGA1 subunit interacts at some extent with other domains of neighboring CNGA1 subunits. As shown in Fig.4a, F-D curves with the detachment with a value of Lc equal to 54± 1.95 and 84± 2.58 nm. In the open state, as shown in Fig.4D, 21% of F-D curves appear to be duplicated, suggesting a strong interaction between the N and C-terminal, in agreement with previous electrophysiological investigations (Gordon et al., 1997; Rosenbaum and Gordon, 2002). F-D curves in which the detachment occurs simultaneously with the unfolding of the S2-S1 transmembrane segments (Fig.4C) have force peak with a value of Lc equal to 54 nm.

This - rather unexpected - coupling between force peaks in the cytoplasmic domain and the detachment can be rationalized by assuming that the cytoplasmic domain - in the closed state - has different conformations determining whether the N-terminal can interact with neighboring subunits or not. We speculate that if the cytoplasmic domain has a folded structure requiring for its unfolding force peaks with a value of Lc equal to 54 and/or 84 nm the N-terminal structurally more constrained and cannot easily interact with the neighboring subunits (Fig.8A). Fig.8 summarizes possible correlations between the unfolding of the cytoplasmic domain and the folding of the other regions of the CNGA1 channel.

If the C-terminal of a given subunit is unfolded or partially unfolded they can more easily interact with the N-terminal of a neighboring subunit (Fig.8A and B). Similarly if the N-terminal is more folded (Fig.8E) than the establishment of interactions with the C-terminal of an other subunit will be more difficult.

Physical interpretation of the variability of the unfolding

The unfolding variability, here observed, is very similar to what previously observed during the unfolding of the lisozyme T4L* (Peng & Li PNAS 2008) where this variability was described by a kinetic partitioning. These authors propose two possible scenarios consistent with the kinetic partitioning mechanism: in one scenario CNGA1 channels have different initial or native conformations and in the second scenario the variability of the unfolding is caused by the stochastic nature of the unfolding process in ragged free-energy landscape. If a protein has different stable conformations it will be more flexible and easily will explore different conformations. Our observation shows clearly that CNGA1 shows different unfolding pathway in both the water-soluble part, i.e. the cytoplasmic domain as well as in the transmembrane domain.

CNGA1 channels, despite many attempts and efforts, have never been crystallized and a possible explanation for their resistance to form well ordered crystals ale to diffract X-rays is the heterogeneity of their native conformations. This observation suggests that the unfolding variability of

CNGA1 channels, here described, could be caused by the combination of several distinct native conformations and a ragged free-energy landscape.

REFERENCES

Anderson, P.A. & Greenberg, R.M. Phylogeny of ion channels: clues to structure and function. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.***129,** 17-28 (2001).

Becchetti, A., Gamel, K. &Torre, V. Cyclic nucleotide-gated channels. Pore topology studied through the accessibility of reporter cysteines. *J. Gen. Physiol.***114,** 377-392 (1999).

Bosshart, P.D. *et al.*The transmembrane protein KpOmpA anchoring the outer membrane of Kleb siellapneumoniae unfolds and refolds in response to tensile load. *Structure***20,** 121-127 (2012).

Bustamante C. Probing biological surfaces.Science. 1994 Apr 8;264(5156):296.

Contreras, J.E., Srikumar, D. &Holmgren, M. Gating at the selectivity filter in cyclic nucleotide-gated channels. *Proc. Natl. Acad. Sci. USA***105,** 3310-3314 (2008).

Craven, K.B. &Zagotta, W.N. CNG and HCN channels: two peas, one pod. *Annu.Rev. Physiol.***68,** 375-401 (2006).

Derebe, M.G., Zeng, W., Li, Y., Alam, A. & Jiang, Y. Structural studies of ion permeation and Ca2+ blockage of a bacterial channel mimicking the cyclic nucleotide-gated channel pore. *Proc. Natl. Acad. Sci. USA***108,** 592-597 (2011).

Dietz H[1], Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments.ProcNatlAcadSci U S A.2004 Nov 16;101(46):16192-7. Epub2004 Nov 5.

Engel, A. &Gaub, H.E. Structure and mechanics of membrane proteins.*Annu. Rev. Biochem.***77,** 127-148 (2008).

Flynn, G.E. &Zagotta, W,N. Conformational changes in S6 coupled to the opening of cyclic nucleotide-gated channels.*Neuron***30,** 689-698 (2001).

Flynn, G.E. &Zagotta, W.N.A cysteine scan of the inner vestibule of cyclic nucleotide-gated channels reveals architecture and rearrangement of the pore. *J. Gen. Physiol.* **121,** 563-582 (2003).

Ge, L., Perez, C., Waclawska, I., Ziegler, C. & Muller, D.J. Locating an extracellular K+-dependent

interaction site that modulates betaine-binding of the Na+-coupled betainesymporterBetP. *Proc. Natl. Acad. Sci. USA***108,** E890-898 (2011).

Gorostiza, P. *et al.*Molecular handles for the mechanical manipulation of single-membrane proteins in living cells.*IEEE Trans Nanobioscience***4,** 269-276 (2005).

Higgins, M.K., Weitz, D., Warne, T., Schertler, G.F.&Kaupp, U.B. Molecular architecture of a retinal cGMP-gated channel: the arrangement of the cytoplasmic domains. *EMBO J.***21,** 2087- 2094 (2002).

Hensen U, Müller DJ.Mechanistic explanation of different unfolding behaviors observed for transmembrane and soluble β-barrel proteins.Structure. 2013 Aug 6;21(8):1317-24.

Hoffmann, T. &Dougan, L. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.***41,** 4781-4796 (2012).

Kawamura, S. *et al.*Kinetic, energetic, and mechanical differences between dark-state rhodopsin and opsin. *Structure***21,** 426-37 (2013).

Kaupp, U.B. *et al.*Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel.*Nature***342,** 762-766 (1989).

Kaupp, U.B. & Seifert, R. Cyclic nucleotide-gated ion channels.*Physiol. Rev.***82,** 769-824 (2002).

Kedrov, A., Ziegler, C. & Muller, D.J. Differentiating ligand and inhibitor interactions of a single antiporter.*J. Mol Biol*. **362,** 925-932 (2006).

Kedrov, A., Janovjak, H., Sapra, K.T. & Müller, D.J. Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. Annu. Rev. Biophys. Biomol.Struct.**36,** 233-260 (2007).

Kotamarthi HC, Sharma R, Narayan S, Ray S, Ainavarapu SR. Multiple unfolding pathways of leucine binding protein (LBP) probed by single-molecule force spectroscopy (SMFS).J Am Chem Soc. 2013 Oct 2;135(39):14768-74.

Kotamarthi HC, Sharma R, Koti Ainavarapu SR. Single-molecule studies on PolySUMO proteins reveal their mechanical flexibility.Biophys J. 2013 May 21;104(10):2273-81.

Li, H *et al.* Reverse engineering of the giant muscle protein titin.*Nature* **418**, 998-1002 (2002).

Lolicato, M. *et al.*Tetramerization dynamics of C-terminal domain underlies isoform-specific cAMP gating in hyperpolarization-activated cyclic nucleotide-gated channels. *J. Biol. Chem.***286,** 44811-44820 (2011).

Marchesi, A., Mazzolini, M. & Torre, V. Gating of cyclic nucleotide-gated channels is voltage dependent. *Nat. Commun.* **3,** 973 (2012).doi: 10.1038/ncomms1972.

Mari, S.A. *et al.*Gating of the MlotiK1 potassium channel involves large rearrangements of the cyclic nucleotide-binding domains. *Proc. Natl. Acad. Sci. USA***108,** 20802-20807 (2011).

Mazzolini, M., Anselmi, C. & Torre, V. The analysis of desensitizing CNGA1 channels reveals molecular interactions essential for normal gating. *J. Gen. Physiol.***133,** 375-86 (2009).

Müller, D.J., Wu, N. &Palczewski, K. Vertebrate membrane proteins: structure, function, and insights from biophysical approaches. *Pharmacol.Rev.***60,** 43-78 (2008).

Nair, A.V., Nguyen, C.H. &Mazzolini, M. Conformational rearrangements in the S6 domain and C-linker during gating in CNGA1 channels.*Eur. Biophys. J.***38,** 993-1002 (2009).

Nair, A.V., Anselmi, C. &Mazzolini, M. Movements of native C505 during channel gating in CNGA1 channels.*Eur. Biophys. J.***38,** 465-478 (2009).

Oesterhelt, F. *et al.*Unfolding pathways of individual bacteriorhodopsins.*Science***288,** 143-146 (2000).

Orsini, F. *et al.*Intermittent contact mode AFM investigation of native plasma membrane of Xeno puslaevis oocyte.*Eur. Biophys. J.***38,** 903-910 (2009).

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. &Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by AFM.*Science***276**, 1109-1112 (1997).

Rief M, Pascual J, Saraste M, Gaub HE.Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles.J Mol Biol. 1999 Feb 19;286(2):553-61.

Santacroce, M. *et al.*Imaging of Xenopuslaevis oocyte plasma membrane in physiological-like conditions by atomic force microscopy.*MicroscMicroanal***19,** 1358-1363 (2013).

Sapra, T.K. et al.Detecting molecular interactions that stabilize native bovine rhodopsin. *J. Mol. Biol.***358,** 255-69 (2006).

Schünke, S., Stoldt, M., Lecher, J., Kaupp, U.B. &Willbold, D. Structural insights into conformational changes of a cyclic nucleotide-binding domain in solution from Mesorhizobium loti K1 channel. *Proc. Natl. Acad. Sci. USA***108,** 6121-6126 (2011).

Tetreault, M.L., Henry, D., Horrigan, D.M., Matthews, G. & Zimmerman, A.L. Characterization of a novel cyclic nucleotide-gated channel from zebrafish brain.*BiochemBiophys Res Commun.***348,** 441-449 (2006).

Zagotta, W.N. *et al.*Structural basis for modulation and agonist specificity of HCN pacemaker channels.*Nature***425,** 200-205 (2003).

Zocher, M., Zhang, C., Rasmussen, S.G., Kobilka, B.K. & Müller, D,J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor.*Proc. Natl. Acad. Sci. USA*. **109,** E3463-472 (2012).

Manuscript 3

Under preparation

Single molecule force spectroscopy of rhodopsin and CNG channels from Rod outer segment plasma membrane from Xenopus laevis

Sourav Maity and Vincent Torre*

Author contributions:

Experiment designing, performance, data analysis and figures preparation was done by S.M.

S.M. And V.T. Wrote the Manuscript.

V.T. Supervised the experiment.

I am thankful to Monica Mazzolini for teaching me the extraction of Rod from Frog Retina.

Single molecule force spectroscopy of rhodopsin and CNG channels from Rod outer segment plasma membrane from Xenopus laevis

Sourav Maity and Vincent Torre*

1 - International School for Advanced Studies (SISSA) Neuroscience Area via Bonomea 265, 34136 Trieste (Italy)

Rod cell outer segments (ROSs) harbor the phototransduction cascade, which involves membrane bound and cytosolic proteins. Membrane bound proteins are distributed on the membrane of disks and on the plasma membrane of ROS. The composition of these two membranes is different in their lipid and protein concentration and distribution. The most populated protein in both membranes is the G-protein coupled receptor rhodopsin. Another important protein playing an essential role in photodranduction cascade is the cyclic nucleotide gated (CNG) channel, which is exclusively found in the ROS plasma membrane. Here we used Single molecule force spectroscopy (SMFS) to unfold membrane proteins from membrane patches of rod outer segments from *Xenopus laevis* retinas. We found different clusters/ensembles of force-distance (F-D) curves which can be associated with unfolding of different native proteins from ROS plasma membrane. The most conserved ensembles were identified to be unfolding of rhodopsin and the unfolding of CNGA1 and CNGB1 single subunits. Both the unfolding of rhodopsin and CNG channels shows a good agreement to what observed previously, but with some significant differences in their unfolding pathway. The unfolding of rhodopsin from the plasma-membrane of ROS shows a pattern of force peaks consistent with the unfolding of single individual transmembrane segments at each peak. In contrast, the unfolding of rhodopsin from disks is obtained from a pairwise unfolding of the two transmembrane segments at each force peak. The unfolding of native subunits of CNG channels is similar to that observed for CNGA1 channels expressed in oocytes, but with some noticeable differences.

INTRODUCTION

Atomic force microscopy (AFM) based single molecule force spectroscopy (SMFS) consists in the application of the force (F) required to unfold a polymer while the distance (D) between the AFM cantilever tip and the sample is measured as the coordinate reaction. In a force-distance (F-D) curve, characterizing the stretching of a protein, each force peak represents the unfolding of a structural segment of the molecule. The obtained sequence of the unfolding force peaks and of their height allows the identification of folded and unfolded regions and to gain indirectly insight into the secondary structure of the protein under investigation (Reif et al., 1997, Engel & Gaub, 2008, Muller et al. 2008). SMFS has been used to identify conformational changes of membrane proteins belonging to the rhodopsin family (Oesterhelt et al., 2000; Sapra et al., 2006; Kawamura et al., 2013) and of other proteins like the Na^+/H^+ antiporter, BetP symporter, KpOmpA transmembrane protein and β2-adrenergic receptor (Kedrov et al., 2006; Ge et al., 2011; Bosshart et al., 2012; Zocher et al., 2012) . SMFS has otherwise been used only rarely to identify conformational changes of a protein in its natural environment.

The outer segments (OSs) of vertebrate photoreceptors are specialized biological structures harboring the phototrnasduction cascade, i.e the biochemical machinery transforming the absorption of a photon into a detectable electrical signal. The outer segment of rod photoreceptors (ROS) is stacked with thousands of lipid discs containing rhodopsin molecules that absorb photons (Reif et al., 1997; Engel & Gaub, 2008; Muller et al., 2008; Oesterhelt et al., 2000). Within 1 s, each excited rhodopsin activates tens of G-proteins, each of which activates one phosphodiesterase molecule (PDE)(Sapra et al., 2006; Kawamura et al., 2013; Kedrov et al., 2006). Activated PDEs rapidly hydrolyse cytoplasmic cGMP thereby closing cyclic nucleotide-gated (CNG) channels (Ge et al., 2011; Bosshart et al., 2012). In darkness, a current carried by Na^+, K^+ and Ca^{2+} ions, which is known as the photo-current, enters via CNG channels into the ROS and is pumped out by Na^+/K^+ ATPase, located in the inner segment (IS)(Zocher et al., 2012). This current can be recorded using suction electrodes (Engel & Gaub, 2008; Jan et al., 1990). The membrane of vertebrate ROS is covered by rhodospin molecules, which are present at a high density and host all the proteins involved in the phototransduction machinery, such as the light sensitive channels, identified as the CNG channels. There is a significant difference in protein-lipid composition of ROS plasma-membrane and disc membrane, which appease to differentiate their activities (Molday &.Molday, 1987; Spencer et al.,1988) in either cases. The ROS plasma membrane contains approximately 3-4 times cholesterol than that found in discs (Kathleen and Schimmel, 1997). The concentration of cholesterol appears to

be an important modulation factor for rhodopsin activity (Kathleen and Albert, 1990;.Molday et al., 1987).

The superfamily of voltage-gated ion channels comprises Na^+, K^+ and Ca^{2+} channels, whose gating (transitions between the open and closed conformation) depends on the membrane voltage. This superfamily also includes cyclic nucleotide-gated (CNG) channels[15-19] that are voltage-dependent(Marchesi et al., 2012) but are opened by the binding of cyclic nucleotides of the cyclic nucleotide-binding (CNB) domain (Kaupp et al., 1989; Kaup & Seifert, 2002; Matulef & Zagotta, 2002). CNG channels are open by the binding of cyclic nucleotides. In vertebrates, seven members of the CNG channel gene family have been identified (Kaupp & Seifert, 2002; Tetreault, 2006) and are grouped into two subtypes, CNGA (CNGA1-CNGA5) and CNGB (CNGB1 and CNGB3). CNGA1, CNGA2, CNGA3, and CNGA5 (but not CNGA4), can form cyclic nucleotide-activated homotetrameric channels, while CNGB1 and CNGB3 are modulatory subunits that cannot form functional homomeric channels.

Hydropathicity and biochemical analyses (Kaupp et al., 1989) of a single subunit of CNGA1 channel (690 amino acid residues -aa- long) have identified six transmembrane α-helices that are known as S1, S2, S3, S4, S5 and S6 (see Supplementary 1) and that are separated by unfolded loops with cytoplasmic N- and C- terminals. There is a pore region between S5 and S6 where ion permeation occurs, and electrophysiological experiments have identified 20 amino acids that form the P-helix (V348-P358) and the selectivity filter (T359-P367) (Becchetti et al., 1999; Flynn & Zagotta, 2001, 2003; Nair et al., 2009; Mazzolini et al., 2002). The C-terminal (N400-D690) is a large cytoplasmic domain composed of the C-linker (N400-E482) and the CNB domain (A483-N610) (Nair et al., 2009; Mazzolini et al., 2002). The CNB domain shares 20% sequence identity with other cyclic nucleotide-binding proteins such as the CNB domain of HCN channels (Zagotta et al., 2003a) and of MlotiK1 channels (Mari, 2011) and consists of three α-helices and eight stranded anti-parallel β-rolls. The functional properties of CNG channels have been investigated extensively (Kaupp & Seifert, 2002; Craven & Zagotta, 2006; Mazzolini et al., 2010), and low-resolution architecture (Higgins et al., 2002), partial crystal structures of the CNB domain (Zagotta et al., 2003; Scunke et al., 2011; Lolicato et al, 2011) and mimics of the pore (Derebe et al., 2011) are available. However, the full-length channel has never been crystallized, and the conformational changes that are associated with gating are not poorly understood.

In a recent investigation (MS1) we have used SMFS to examine gating of CNGA1 channels that

are expressed in *Xenopus laevis* oocytes, i.e. almost *in situ*. We identify F-D curves by an informatics analysis and by expressing fusion proteins that are composed of CNGA1 channels that are linked at their C-terminal to marker proteins whose unfolding is known. We identify those domains that change their conformation upon gating. In the present manuscript, we attempt to unfold native CNG channels from membrane patches of ROS plasma membrane. We use the F-D traces obtained from the unfolding of CNGA1 channels, as a template, to recognize and identify F-D traces from the unfolding of native CNG channels. In the present MS we describe SMFS experiments from *Xenopus laevis* ROS, in which we characterize native CNG channels, rhodopsin and other – at the moment not identified – membrane proteins, possibly involved in the phototransduction cascade.

METHODS

Experimental setup

A simple but effective experimental design has been introduced for the preparation of ROS plasma membrane for AFM experiments. The AFM (JPK NanoWizard 3) was mounted on a Olympus IX71 inverted microscope (fig.1 a). The stage of the microscope contained a liquid chamber which was connected through the inlet-outlet channels to a liquid pump. The pump had a range of operation of 0.02 to 3 ml/min. A low and high flow of liquid can be maintained inside the chamber. The illumination for the microscope was done under dim red light. The hole setup is isolated by an acoustic hood to maintain the dark and increase the stability.

Mechanical Dissociation of Photoreceptor

We have used *Xenopus laevis* ("Xenopus express" Ancienne Ecole de Vernassal, Le Bourg 43270, Vernassal, Haute Loire, France) rod cells to investigate the structural properties of ROS CNG channels. The frog was dark adapted 3 hours prior to the experiment. Sample preparation and AFM experiments were done in presence of dim red light. The eyes of Xenopus laevis were extracted surgically. The eye balls were preserved in ringer solution (110 mM NaCl, 2.5 mM Kcl, 1 mM $CaCl_2$,, 1.6 mM $MgCl_2$, 3 mM Hepes, 0.01 mM EDTA and 10 mM sucrose, pH 7.8). One eye ball was then enucleated and hemisected in coronal section posterior to the ciliary body. The eyecups were then quartered in the ringer solution. One of the slice were then transfer to absorption buffer

(150 mM Kcl, 25 mM $MgCl_2$ and 10 mM Trizma base, pH 7.5). Intact rods were obtained by mechanical peeling from the retinal pigment epithelium, intact rods were then collected with 200 µl pipette and deposited on a freshly cleaved muscovite mica, placed in the liquid chamber mounted on AFM stage.

Preparation of Rod plasma membrane

Incubated rods were maintained for 20 minutes to be adsorbed by the negatively charged surface. The surface adhesion increases in the presence of high concentration of Mg^{2+} ions in the absorption buffer. The surface were monitored by the inverted microscope (fig.1 b). Then the surface were cleaned properly with absorption buffer for 2 minutes at a flow rate of 0.5 ml/min. The surface were monitored again (fig.1 c). In that step, all the rod cells which were loosely bounded to the surface were removed, leaving all the cells strongly attached to the surface. The AFM head was mounted placing the tip on top of a healthy rod outer segment (ROS), but far from the surface (~500 µm). An enzyme mixture of 100 µM Hyaluronidase and 1 uN/ml Neuroaminidase was injected into the chamber using the pump for 5 minutes. After the enzymatic treatment the surface was cleaned immediately with recording solution (150 mM Kcl, 10 mM tris Hcl, pH 7.5) at a liquid flow rate of 1 to 2 ml/min. the removal of the targeted ROS can be observed through the camera. After the removal of the target cell (usually after 5-8 minutes) the surface was cleaned gently (flow rate of 0.06 ml/min) first with absorption buffer, to remove different cytoplasmic molecules which can aggregate on top of the attached plasma membrane ; and then with recording solution. At that stage the surface was ready for AFM applications.

Imaging and identification of ROS plasma membrane

After the sample was ready, the AFM tip (HYDRA 50NGG from App-Nano, spring constant ~0.08 N/m) was brought back to the surface. The AFM image was taken using tapping mode at a operating frequency of ~14 kHz and in presence of recording solution, on a surface area of 50 μm^2. About 80 % attempt was successful to localized a long patch of protrusion with a height profile of ~5 nm and diameter of ~5 µm. Which then can be considered confirmed to be a ROS plasma membrane. Further higher resolution image was taken on top of the membrane in an area of 1-2 μm^2. Then the surface was taken for the SMFS experiments.

SMFS on ROS plasma membrane

Single molecule force spectroscopy was performed using the same tip as for imaging, the cantilever was calibrated by thermal fluctuation method (Hutter et al., 1993) each time of the experiment. Each approach of the cantilever was maintained at a constant speed of 1 μm/s. Then the tip was pushed to the surface with a force of 1 nN for 0.5 s, then the tip was retracted with a constant speed of 500 nm/s up to a distance of 1 μm. This processes approach and retraction goes for several thousands of time for different point for each scanning area, while the deflection of the cantilever was recorded as force vs distance (F-D) curve. F-D curves were saved automatically. After acquisition of ~200000 F-D curves, we performed the bio-informatics analysis as introduced in our previous work (MS1). The clusterization was done for a distance in between 10 nm to 750 nm to characterize all possible clusters of F-D curves that may come from the unfolding of the variety of proteins that are present in ROS plasma membrane. Then each cluster was then separated according to there unfolded length (Lc nm). And different clusters were then evaluated separately. For extraction of molecular information from the F-D curves we have used Worm like chain (WLC) model to fitt individual F-D curves with a persistence length of 0.4nm, and the contour length (Lc) was calculated. For mapping the Lc value with corresponding amino acid sequence, we have used 0.4 nm to be an average length of a single amino acid residue. These values were maintained through out the experiments.

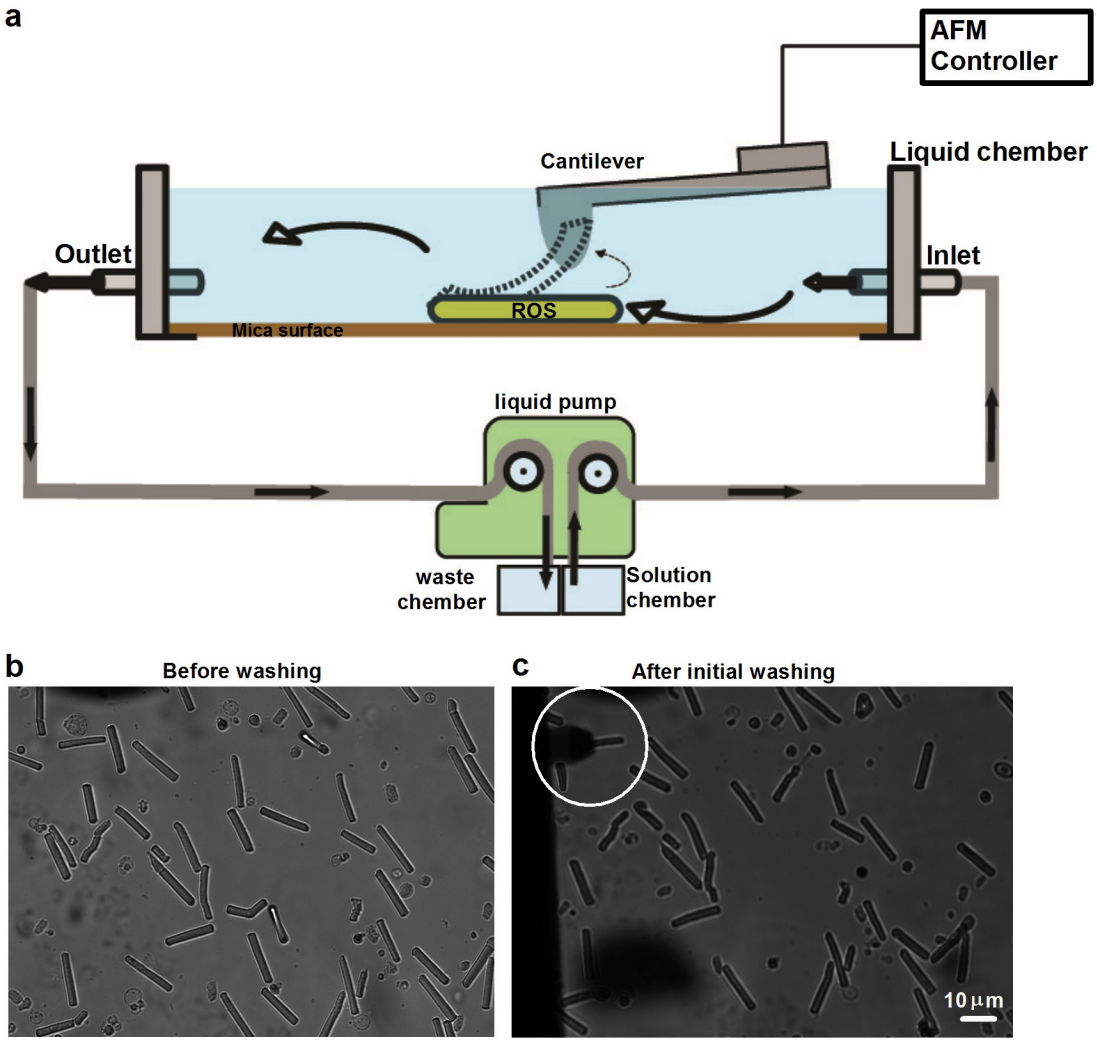


Figure 1: *Experimental setup for sample preparation: a: the schematic diagram for preparation of ROS plasma membrane attached on the mica surface. The setup contains a liquid chamber mounted on AFM sample holder; a liquid pump connected to the chamber which can control a liquid flow at a rate from 0.02 to 3 ml/min. b: image of Rod cells after incubation on mica surface. C: image of sample surface after washing with primary buffer.*

RESULTS

AFM images of membrane patches from rod outer segments

Rod outer segments (ROS) were deposited on a mica surface and by following the procedure described in the Methods section we removed the cytoplasm leaving exposed the intracellular side of the membrane. These membrane patches had the extracellular side adhering to the mica surface and the intracellular – or cytoplasmic – facing the cantilever tip of the AFM. These membrane patches were imaged by AFM in tapping mode. Images of these membrane patches at different spatial resolution are shown in Fig.3a,c and e.

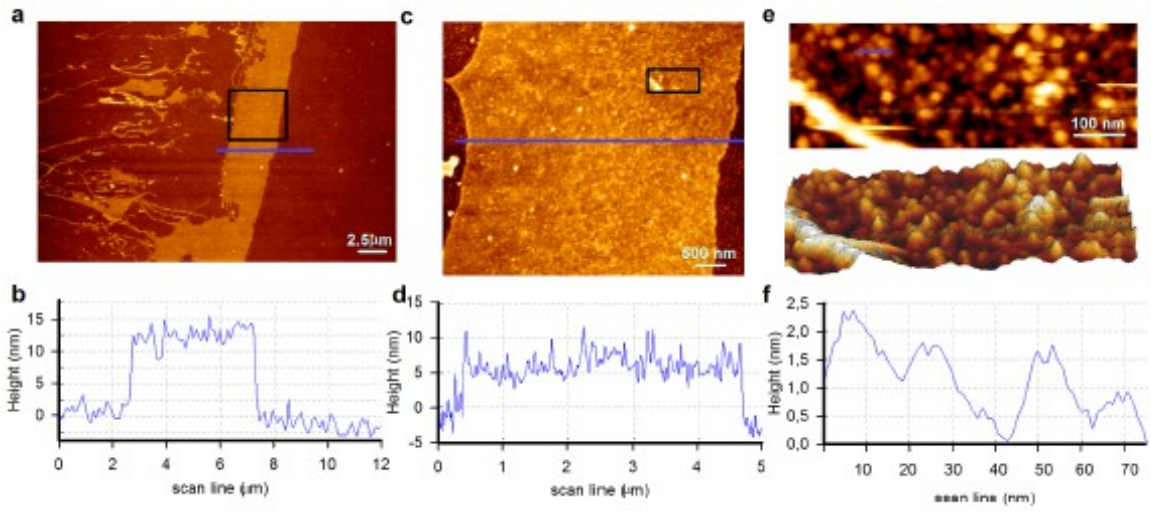


Figure 2 AFM image of membrane patches from ROS plasma membrane deposited on mica surface. ***a:*** *the localization of the membrane patch from ROS. Clearly visible a long patch with diameter of 5 µm and a protrusion of height ~5 nm. b: the height profile of the surface, in the cross section denoted by blue solid line in a. c, e: image taken in the same patch but increase in magnifications; d,f represent their corresponding height profile respectively.*

The height of these membrane patches is 5-10 nm (Fig.3band d) and - at the highest resolution - protrusions with a height of 1-2 nm and a width of about 10 nm. These protrusions are likely to be membrane proteins belonging to the ROS plasma membrane. These proteins could be involved in the usual cell metabolism and homeostasis but could be also involved in phototransduction. A higher resolution AFM imaging of these proteins could identify their structural arrangements (Muller et al., 1995) .

SMFS on ROS plasma membrane

After the localization of the plasma membrane, we performed SMFS experiments. All F-D curves were passed through an initial filtering to discard invalid F-D curves (i.e. those coming from tip sample or mechanical artifacts, multi-molecule unfolding, unspecific adhesion,….). Then we performed clusterization on all F-D curves, that passed through the initial filtering, as described in method (MS1). We obtained ~250000 F-D curves and about 4% of them passed the initial filtering step. We applied the clusterization procedure (see Methods) and we obtained some tens of clusters with significantly different unfolding spectra. Each of those cluster were then scrutinized rigorously to be sure that those clusters could be ascribed to the unfolding of different molecules. This was done by comparing the total unfolded length (Lc) and the increase in contour length (ΔLc) histogram. After this validation step we identified 10 different clusters that were ascribed to the unfolding of 10 different molecules. In the present manuscript, we have report data only for the unfolding of rhodopsin molecules and CNG channels (A1 and B1), while all other clusters will be described in a future manuscript.

Single molecule unfolding of rhodopsin

The unfolding of rhodopsin molecules has been very extensively investigated in isolated discs or reconstituted nano-discs from bovine ROS (Muller et al., 2000, Kessler & Gaubb 2006). In these experiments the cantilever tip approached the rhodopsin molecule so to attach to either C or N-terminal end of the protein and the total length of the unfolded protein was about 70-80 nm (219 aa) and 80-90 nm respectively. Those F-D traces have three reproducible force peaks with a value of Lc of approximately 32 (88aminoacids aa), 53 (148 aa) and 78 nm (219 aa) for N-terminal pulling and 31 (86 aa), 53 (147 aa) and 87 nm (242 aa) for C-terminal pulling.

Instead of bovine disks – as in those previous investigations – we used ROS from the *Xenopus* retinas. Rhodopsin from frog retinas is 354 aa long (Pittler et al., 1992), while the bovine rhodopsin is 348 aa long (Kedrov & Muller et al.,Rev. 2007). Overall, frog and bovine rhodopsin display 85% identity and another 10% similarity at the amino acid level. In our experiments, when the extracellular side of the membrane patch is attached to the mica surface, the cantilever tip will approach rhodopsin from its C-terminal and not from the N-terminal as in experiments with isolated discs. As a consequence of these differences we obtained several hundreds of F-D traces similar to those obtained for the unfolding of bovine rhodopsin (Kedrov et al., 2007, Engel &

Gaubb 2008) but with some noticeable differences. SMFS experiments in ROS plasma-membrane provided the largest cluster with a frequency of ~22% over all valid F-D curves, which had a complete unfolding length of 109 nm (~302 aa) of Lc (fig.4 a). This cluster corresponds to the unfolding of single molecule of rhodopsin. The Lc histogram shows force peaks at 15±5.65, 25±4.35, 41±2.81, 56±5.15, 74±4.76, 92±4.21 and 109±6.53 nm of Lc. The most reproducible peaks are at 41±2.81, 56±5.15, 74±4.76, 92±4.21 and 109±6.53 nm, while the peaks at 15±5.65 and 25±4.35 nm appears with lower probabilities. A mapping of corresponding force peaks to the secondary structure of rhodopsin is shown in figure 4 c. Rhodopsin from discs, has an unfolding pathway that has three reproducible force peaks which correspond to pairwise unfolding of the transmembrane domains (Kessler & Gaub,2006); i.e. a pairwise unfolding of H7-H6, H5-H4 and H3-H2 Pulling from C-terminal) were observed most frequently. In the present SMFS experiments, the unfolding of a single rhodopsin molecule shows an unfolding pathway with five instead of three force peaks. The first two least probable force peaks are around ~15 nm (42 aa) and ~25 nm (69 aa) of Lc and correspond to the partial unfolding of cytoplasmic helix H8, and a complete unfolding of H7 respectively. The force peaks around ~41 nm (112aa) and ~56 nm (157 aa) of Lc can be described as the stepwise unfolding of H6 and H5 helices respectively. Then the complete unfolding of H4 occurs at a Lc value of ~74 nm (207 aa) following the unfolding of H3 with an Lc value of ~92 nm (254 aa). The final detachment with a complete unfolding of H2 and a partial unfolding of H1 is observed at a Lc value of 109 nm (302 aa). This difference in observation between the unfolding pathway of rhodopsin from disc membrane to plasma membrane can be explain from the molecular composition of plasma and discs membrane. It is now well understood that cell membrane lipid composition modulates the function of membrane protein at a large scale. Previous finding suggests the higher concentration of cholesterol in the ROS plasma membrane caused the inhibition of rhodopsin functions (Boesze-Battaglia et al., 1990 a,b). ROS plasma-membrane contain three fold higher concentration of cholesterol than the disc membrane. Also cholesterol is known to hold a key contribution to cell rigidity (Inbar & Shinitzky, 1974; Yeagle P. L., 1987), where higher concentration of cholesterol decreases the lipid-protein mobility, thus increasing rigidity of the membrane. We propose that higher concentration of cholesterol in the ROS plasma membrane decreases the mobility of rhodopsin by increasing the mechanical stress on the lipid-protein interface. Each transmembrane helices of rhodopsin form an individual non-interactive potential barrier which is observed when and external perturbation is applied.

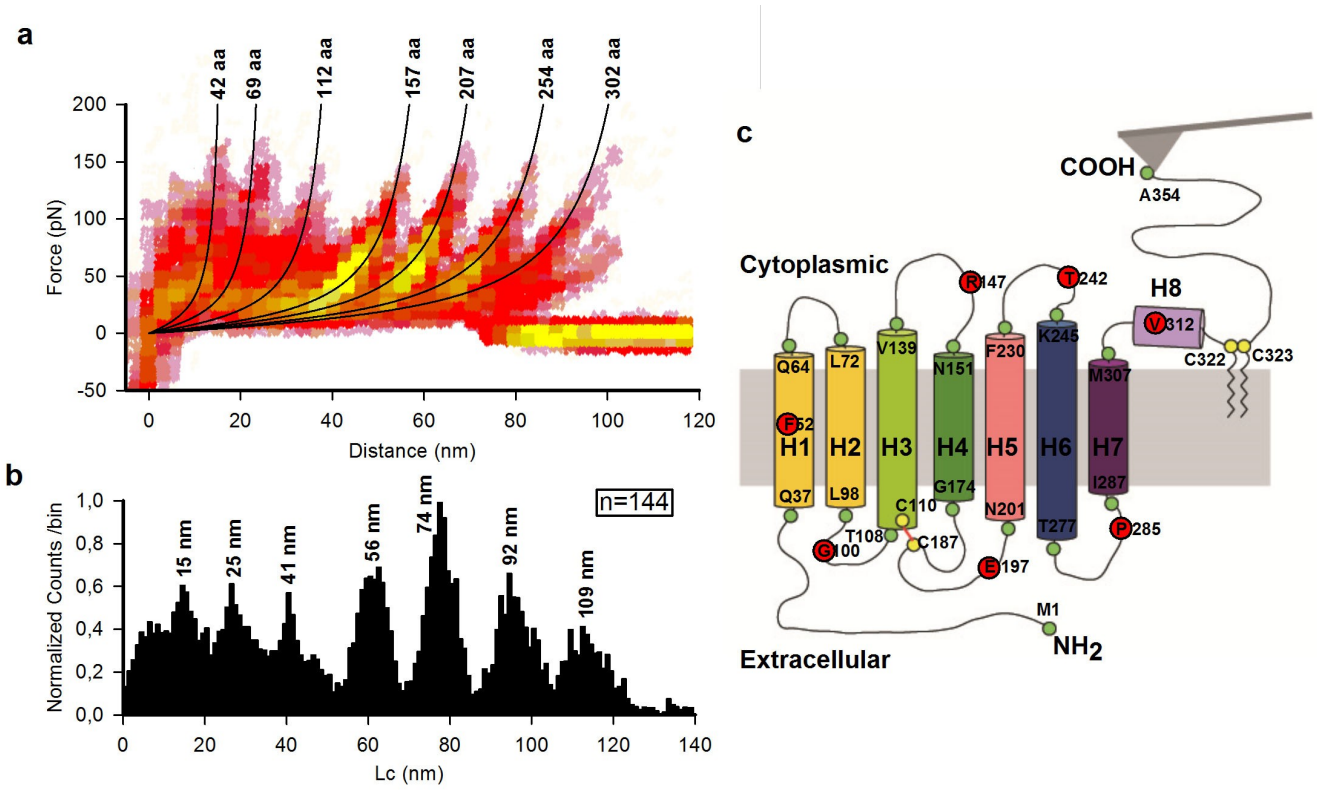


Figure 3. unfolding pathways of native rhodopsin from Frog ROS plasma membrane. a: *a density plot of superposition of F-D curves obtained from unfolding of a single rhodopsin from native ROS plasma membrane, numbers represent the amino acid (aa) number of corresponding force peak position.* ***b:*** *Lc histogram obtained from all curves in panel a (bin 1 nm), n represent the total number of F-D-curves.* ***c:*** *cartoon representation of the frog rhodopsin. Green dots represent the amino acid that define the structural transitions. Red circles represent the force peak positions of rhodopsin molecule unfolded from C-terminal end.*

Unfolding of native CNG channels from ROS

Native CNG channels are heterotetramers composed by three CNGA1 subunits and one CNGB1 subunit (Weitz et al., 2002, Zheng et al. 2002, Zhong et al., 2002). There are experimental evidence stating that the CNG channel is directly connected to the disc rims via a direct interaction between the glutamic acid rich domain (GARP) of CNGB1 subunit and the disc rim specific scaffolding protein tetraspanin peripherin-2 (Molday et al., 2001; Ritter et al., 2011). A recent finding by Becirovic & Nguyen et al. (2014) reveals that this interaction by peripherin-2 links the CNGB1 to rhodopsin.

Rhodopsin is by far the most abundant membrane proteins in discs and ROS plasma membranes (Molday & Molday, 1987; Spencer et al., 1988) and it is expected that SMFS will harvest a large

number of rhodopsin molecules and indeed we obtained a cluster with 47% frequency that can be ascribed to the complete unfolding of single rhodopsin molecule. We then looked for F-D traces presumably obtained from the unfolding of the alpha subunit from native CNG channels from ROS membranes. The density of native CNG channels is <20% compared to the rhodopsin density of >60% depending upon the age of the cell. The CNGB1 subunit is very similar to the CNGA1 subunit but has a much longer N-terminal harboring the GARP domain, (Fig.4 c and f). Therefore F-D curves obtained from the unfolding of the CNGB1 subunit from the C-terminal end is expected to be very similar to the unfolding of the CNGA1 subunit from the C-terminal but to span for longer values of distance (or tip sample separation), as shown in Fig.4 d.

In the absence of cGMP, clusterization provided us two cluster with about 10-15 % frequency of occurrence, which have a total length of ~278 nm (fig4 a-b) and ~302nm (fig4 d-e) of Lc. The cluster with total length of ~278 nm shows >90 % similar unfolding pathway to what we have observed from CNGA1 homomeric channel expressed in oocyte plasma membrane. The appearance of force peaks are 60±4.41 (150aa), 97±5.01 (242 aa), 118±4.85 (295 aa), 158±3.21 (39 aa), 173±2.81 (432 aa), 205±6.21 (512 aa), 251±6.45 (627 aa) and 274±3.55 nm (685 aa) of Lc values, with unfolding forces between 45 to 85 pN. The peaks at Lc values of 60±4.41 and 173±2.81 nm appear with a lower frequency, i.e. approximately 80% less than of the other force peaks. A mapping of this force peak to the amino acid sequence of CNGA1 suggests that, unfolding of entire C-terminal occurs in three steps at Lc values around 60, 97, 118 nm while the peaks from 158 to 251 nm of Lc can be considered to be the complete unfolding of the entire transmembrane domain, also the S6 and S5 are unfolding together with the force peak around 158 nm, while sometimes S4 unfolds concomitantly at a single step around 173 nm of Lc. Then S3, S2 are unfolded together at Lc value around 205 nm. The force peak at 251 nm can be ascribed as the unfolding of S1 with some part of the cytoplasmic site of amino terminal. The detachment in this case are also occurring with an additional force peak around 278 nm, as we have shown in our previous work suggesting a moderate intra-subunit interaction from N to C terminal.

The second cluster with a full length Lc value of 302 nm shows some similarity in the unfolding pathway as in CNGA1. This unfolding pathway is tentatively ascribed to the unfolding of single CNGB1 subunit, pulled from C-terminal. We could not find any cluster which could be ascribed as the pulling of CNGB1 from N-terminal, which should be approximately 400nm (984 aa) long. This might be because of the interactive nature of GARP domain in the N-terminal, which makes the loop unfavorable to attach to the AFM tip. The unfolding of CNGB1 occurring at several force

peak positions (fig4d) at Lc values of 46±3.15 (115 aa), 78±3.85 (195 aa), 107±4.55 (267 aa), 140± 4.75 (350 aa), 171±3.65 (427 aa), 206±3.85 (515 aa), 254±3.22 (635 aa), 281±4.31 (702 aa) and 302±3.99 nm (755 aa) (fig4e). A mapping of Lc value to the amino acid sequence of CNGB1 (fig4f) implies that the cytoplasmic domain of CNGB1 is unfolded in three steps, as in case of CNGA1, but with shorter Lc values around 46, 78 and 107 nm. While the trans-membrane domains unfolds in five steps around 140 nm for unfolding of S6 domain, 171 nm for single step unfolding of S5 domain, 206 nm for a paired unfolding of S4 and S3 domains, and finally S2 and S1 unfold together at a Lc value around 254 nm. After the unfolding of all the trans-membrane domains the unfolding of interactive amino-terminal occurs in two steps at Lc values of 281 and 302 nm. The main difference in the unfolding of CNGB1 to CNGA1 is the S5 domain, where in CNGA1 S5 always coupled either with S6 (in closed state) or with S4 (in open state), but in CNGB1, S5 unfolds in an individual step. This different behavior is interesting in respect to the fact that CNGB1 alone can not compose a functional channel, thus needed to be considered for further verifications.

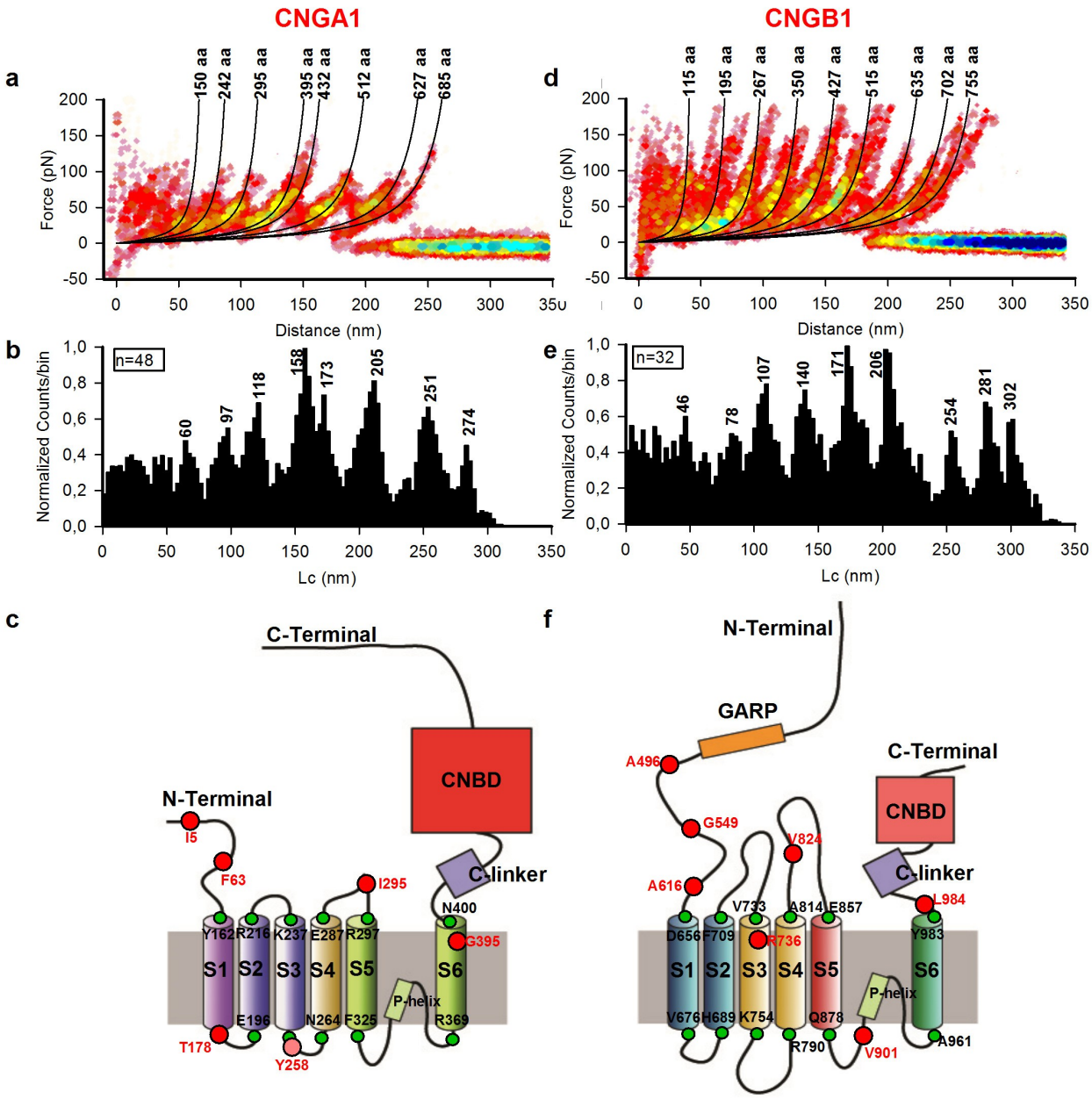


Figure 4. Single molecule unfolding of CNG channels in the closed state. ***a:*** *density plot of superposition of F-D curves from unfolding of single subunit of CNGA1, solid line represents the WLC fitting of the individual force peak, the number represents the number of amino acids pulled in the corresponding force peaks.* ***b:*** *Lc histogram obtained from all the curves in panel a (bin-2 nm), number represents the corresponding Lc values in nm.* ***c:*** *a cartoon representation of CNGA1 subunit, green dots showing the structurally specific amino acids, and the red circles are representing an approximate position of the corresponding force peaks. The pink circle represents the least probable force peak position.* ***d, e, f :*** *representing the same as in a, b, c respectively but for CNGB1.*

Unfolding of native CNG channels in the open state

we performed SMFS experiment on the ROS plasma membrane in presence of 2 mM cGMP (saturating concentration). We obtained clusters representing the unfolding of CNGA1, where the individual force peak showing an increased unfolding force of 80-150 pN (fig5 a), in agreement to our previous observation. The force peaks are appearing at 55±4.95 (137 aa), 80±5.11 (200 aa), 112±4.41 (280 aa), 143±4.54 (357 aa), 176±5.11 (440 aa), 205±3.12 (512 aa), 248±6.56 nm (620 aa) and then the detachment at a Lc value of 278±6.88 nm (fig5b). A mapping of the Lc value to the structural segment of CNGA1 is shown in (figure5c). In presence of cGMP, CNGA1 unfolding shows some significant difference. In the open state the S6 helix unfold in a single step and then S5 and S4 unfold together, showing a coupling between the S4-S5 helices as reported before. The next force peak is corresponding to the paired unfolding of S3 and S2 helices, and then complete unfolding of S1 along with a part of cytoplasmic site appears together.

In presence of cGMP we could not found a cluster which can ascribed as complete unfolding of CNGB1 as observed in the closed state. The reasoning behind this can be given as: in open state CNG channels shows higher order of inter-subunit interactions (Trudeau & Zagotta, 2003), which may caused the C and N terminal of CNGB1 inaccessible to the AFM tip.

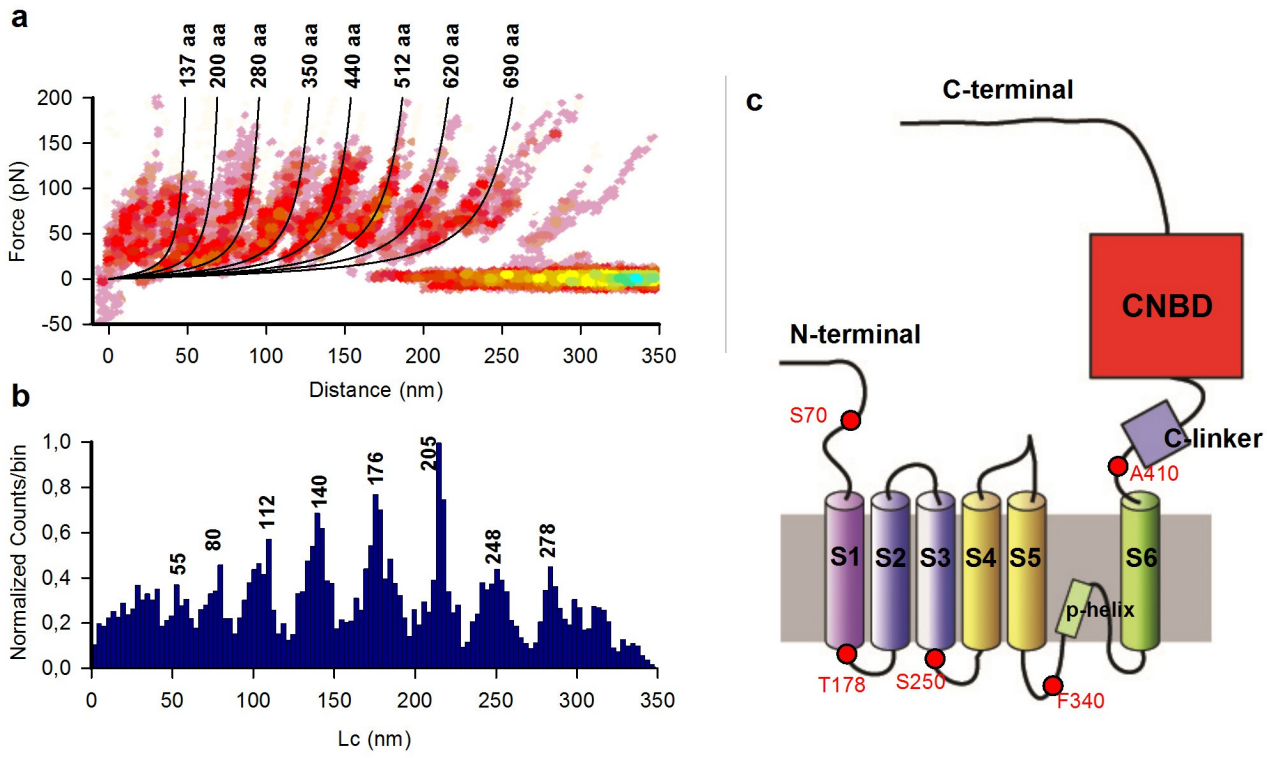


Figure 4 Single molecule unfolding of CNGA1 in the open state. a: *density plot of superposition of curves in presence of 2mM cGMP. Solid lines represent the WLC fitting and the number represents the corresponding Lc in term of amino acid (aa).* ***b:*** *Lc histogram obtained from all the traces in panel a (bin-3 nm), numbers represent the corresponding Lc values in nm.* ***c:*** *a cartoon representation of single CNGA1 subunit identifying different structural segments forming separated potential barrier, red dot and numbers represent the estimated force peak positions in different structural segment.*

DISCUSSION

In the present manuscript, we have successfully introduced a new method to perform SMFS experiments in situ, i.e. from ROS plasma-membrane patch from *Xenopus laevis* retina. This work extend the previous investigation of SMFS of CNGA1 channels expressed in *Xenopus laevis* oocytes (MS1,2). In the present investigation, the ROS membrane patch was attached to mica surface allowing AFM imaging and SMFS experiments to investigate the most important proteins in phototransduction. We have obtained reliable new data for the unfolding of native rhodopsin from ROS plasma mebrane and native CNG channels (A1, B1). Rhodopsin in the ROS plasma membrane showed different unfolding pathway than that observed from discs. The unfolding of native CNGA1 and CNGB1 from frogs shows very similar unfolding pathways as that seen in bovine CNGA1 channels expressed in *Xenopus laevis* oocytes. Some differences are observed for the S5 helix: in CNGB1 S5 unfolds alone independently from S6 and S4, where in case of

CNGA1, the unfolding of S5 is coupled with the unfolding of S6 in the closed state and with S4 in the open state.

Mechanical stability of a protein molecule depends on the spatial distribution of energy of their individual domains. Under an external perturbation these individual domains evolve through a number of intermediate transition levels (unfolding pathway) to the unfolded state. Our observation shows that the transmembrane domains of rhodopsin from the ROS plasma membrane has discrete potential barriers against the external perturbation – i.e. the external force exerted by the AFM cantilever - which generate and lead to discrete unfolding steps for each a-helix transmembrane. These data suggest that rhodopsin in the plasma membrane has a restricted number of energetically favorable conformations showing higher order of mechanical and energetic stability. Therefore, rhodopsin in the ROS plasma membrane shows a higher order of mechanical and energetic stability than that from the disc membrane. This difference in mechanical stability is likely to be a direct consequence of the cholesterol-lipid composition in those membranes (Boesze-Battaglia et al., 1990). We propose that, this increase in mechanical and energetic stability of rhodopsin in the ROS is directly responsible for the inhibition of their activity in the ROS plasma membrane.

Native CNGA1 channels from ROS plasma membrane show a similar unfolding pathway as the one we have observed previously when pulling experiments were performed from the plasma membrane of oocytes injected with the mRNA coding for the bovine CNGA1 subunit. The unfolding of the native CNGB1 subunit has a similar pathway as that for the CNGA1 subunit, but the S5 domain of CNGB1 shows different mechanical properties from CNGA1. In CNG channels, the trans-membrane domain S6 and S5 surround the pore region. In CNGA1 we have previously observed that the S5 is mechanically coupled to S6 in the closed state, while it is more coupled to the voltage sensor S4 in the open state Electrophysiological experiments suggested that the S4-S5 loop may go through a structural transition during the opening of the channel, thus may have a important role in the allosteric channel opening. But in case of CNGB1, the S5 domain unfolds always in an individual step, i.e. neither coupled to S6 nor to S4. Considering that CNGB1 fails to form a functional homomeric channel as CNGA1 and in contrast to what observed in CNGA1, we propose that this coupling between S5 and S4 may have a significant influence on the channel activity and should be further investigated.

REFERENCES

Anderson, P.A. & Greenberg, R.M. Phylogeny of ion channels: clues to structure and function. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.***129,** 17-28 (2001).

Becchetti, A., Gamel, K. &Torre, V. Cyclic nucleotide-gated channels. Pore topology studied through the accessibility of reporter cysteines. *J. Gen. Physiol.***114,** 377-392 (1999).

Bosshart, P.D. *et al.*The transmembrane protein KpOmpA anchoring the outer membrane of Kleb siellapneumoniae unfolds and refolds in response to tensile load. *Structure***20,** 121-127 (2012).

Boesze-Battaglia K , S J Fliesler and A D. Albert. Relationship of cholesterol content to spatial dis tribution and age of disc membranes in retinal rod outer segments. *J. Biol. Chem. 1990, 265:18867-18870.*

Boesze-Battaglia K. and Albert A.D. Cholesterol modulation of photoreceptor function in bovine retinal rod outer segments. J. Biol. Chem. 1990, 265:20727-20730

Contreras, J.E., Srikumar, D. &Holmgren, M. Gating at the selectivity filter in cyclic nucleotide-gated channels. *Proc. Natl. Acad. Sci. USA***105,** 3310-3314 (2008).

Craven, K.B. &Zagotta, W.N. CNG and HCN channels: two peas, one pod. *Annu.Rev. Physiol.***68,** 375-401 (2006).

Derebe, M.G., Zeng, W., Li, Y., Alam, A. & Jiang, Y. Structural studies of ion permeation and Ca2+ blockage of a bacterial channel mimicking the cyclic nucleotide-gated channel pore. *Proc. Natl. Acad. Sci. USA***108,** 592-597 (2011).

Engel, A. &Gaub, H.E. Structure and mechanics of membrane proteins.*Annu. Rev. Biochem.***77,** 127-148 (2008).

Flynn, G.E. &Zagotta, W,N. Conformational changes in S6 coupled to the opening of cyclic nucleotide-gated channels.*Neuron***30,** 689-698 (2001).

Flynn, G.E. &Zagotta, W.N.A cysteine scan of the inner vestibule of cyclic nucleotide-gated channels reveals architecture and rearrangement of the pore. *J. Gen. Physiol.* **121,** 563-582 (2003).

Ge, L., Perez, C., Waclawska, I., Ziegler, C. & Muller, D.J. Locating an extracellular K+-dependent interaction site that modulates betaine-binding of the Na+-coupled betainesymporterBetP. *Proc. Natl. Acad. Sci. USA***108,** E890-898 (2011).

Gorostiza, P. *et al.*Molecular handles for the mechanical manipulation of single-membrane proteins in living cells.*IEEE Trans Nanobioscience***4,** 269-276 (2005).

Higgins, M.K., Weitz, D., Warne, T., Schertler, G.F.&Kaupp, U.B. Molecular architecture of a retinal cGMP-gated channel: the arrangement of the cytoplasmic domains. *EMBO J.***21,**

2087-2094 (2002).

Hoffmann, T. &Dougan, L. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.***41,** 4781-4796 (2012).

Hutter, J. L., and J. Bechhoefer (1993).Calibration of Atomic-Force Microscope Tips.*Rev. Sci. Instrum.* **64**:3342-3342.

Inbar M, Shinitzky M. Increase of cholesterol level in the surface membrane of lymphoma cells and its inhibitory effect on ascites tumor development.Proc Natl Acad Sci U S A. 1974 May;71(5):2128-30

Jan, L.Y. & Jan, Y.N.A superfamily of ion channels.*Nature***345,** 672 (1990).

Kawamura, S. *et al.*Kinetic, energetic, and mechanical differences between dark-state rhodopsin and opsin. *Structure***21,** 426-37 (2013).

Kaupp, U.B. *et al.*Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel.*Nature***342,** 762-766 (1989).

Kaupp, U.B. & Seifert, R. Cyclic nucleotide-gated ion channels.*Physiol. Rev.***82,** 769-824 (2002).

Kedrov, A., Ziegler, C. & Muller, D.J. Differentiating ligand and inhibitor interactions of a single antiporter.*J. Mol Biol*. **362,** 925-932 (2006).

Kedrov, A., Janovjak, H., Sapra, K.T. & Müller, D.J. Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. Annu. Rev. Biophys. Biomol.Struct.**36,** 233-260 (2007).

Kessler M, Gaub HE. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM.Structure. 2006 Mar;14(3):521-7.

Lolicato, M. *et al.*Tetramerization dynamics of C-terminal domain underlies isoform-specific cAMP gating in hyperpolarization-activated cyclic nucleotide-gated channels. *J. Biol. Chem.***286,** 44811-44820 (2011).

Marchesi, A., Mazzolini, M. & Torre, V. Gating of cyclic nucleotide-gated channels is voltage dependent. *Nat. Commun.* **3,** 973 (2012).doi: 10.1038/ncomms1972.

Mari, S.A. *et al.*Gating of the MlotiK1 potassium channel involves large rearrangements of the cyclic nucleotide-binding domains. *Proc. Natl. Acad. Sci. USA***108,** 20802-20807 (2011).

Mazzolini, M., Anselmi, C. & Torre, V. The analysis of desensitizing CNGA1 channels reveals molecular interactions essential for normal gating. *J. Gen. Physiol.***133,** 375-86 (2009).

Müller, D.J., Wu, N. &Palczewski, K. Vertebrate membrane proteins: structure, function, and insights from biophysical approaches. *Pharmacol.Rev.***60,** 43-78 (2008).

Molday RS, Molday LL.Differences in the protein composition of bovine retinal rod outer

segment disk and plasma membranes isolated by a ricin-gold-dextran density perturbation method.J Cell Biol. 1987 Dec;105(6 Pt 1):2589-601.

Nair, A.V., Nguyen, C.H. &Mazzolini, M. Conformational rearrangements in the S6 domain and C-linker during gating in CNGA1 channels.*Eur. Biophys. J.***38,** 993-1002 (2009).

Nair, A.V., Anselmi, C. &Mazzolini, M. Movements of native C505 during channel gating in CNGA1 channels. *Eur. Biophys. J.***38,** 465-478 (2009).

Oesterhelt, F. *et al.*Unfolding pathways of individual bacteriorhodopsins.*Science***288,** 143-146 (2000).

Pittler SJ, Fliesler SJ, Baehr W.Primary structure of frog rhodopsin.FEBS Lett. 1992 Nov 23;313(2):103-8.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. &Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by AFM.*Science***276**, 1109-1112 (1997).

Santacroce, M. *et al.*Imaging of Xenopuslaevis oocyte plasma membrane in physiological-like conditions by atomic force microscopy.*MicroscMicroanal***19,** 1358-1363 (2013).

Sapra, T.K. et al.Detecting molecular interactions that stabilize native bovine rhodopsin. *J. Mol. Biol.***358,** 255-69 (2006).

Schünke, S., Stoldt, M., Lecher, J., Kaupp, U.B. &Willbold, D. Structural insights into conformational changes of a cyclic nucleotide-binding domain in solution from Mesorhizobium loti K1 channel. *Proc. Natl. Acad. Sci. USA***108,** 6121-6126 (2011).

Tetreault, M.L., Henry, D., Horrigan, D.M., Matthews, G. & Zimmerman, A.L. Characterization of a novel cyclic nucleotide-gated channel from zebrafish brain.*BiochemBiophys Res Commun.***348,** 441-449 (2006).

Trudeau MC[1], Zagotta WN. Calcium/calmodulin modulation of olfactory and rod cyclic nucleotide-gated ion channels.JBiol Chem.2003 May 23;278(21):18705-8. Epub2003 Mar 7.

Yeagle PL, Frye J.Effects of unsaturation on 2H-NMR quadrupole splittings and 13C-NMR relaxation in phospholipid bilayers.Biochim Biophys Acta. 1987 May 29;899(2):137-42.

Zagotta, W.N. *et al.*Structural basis for modulation and agonist specificity of HCN pacemaker channels.*Nature***425,** 200-205 (2003).

Zocher, M., Zhang, C., Rasmussen, S.G., Kobilka, B.K. & Müller, D,J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor.*Proc. Natl. Acad. Sci. USA*. **109,** E3463-472 (2012).

Conclusions and perspectives

The objectives of my PhD thesis were: to use atomic force microscopy (AFM) and single molecule force spectroscopy (SMFS) in CNGA1 ion channels and to gather structural and functional information of these channels in their physiological conditions, thus avoiding the problematic crystallization or plasma membrane reconstitution. The experimental work presented in this thesis shows the potential of SMFS when combined with bioinformatics, mutagenesis and electrophysiology to obtaining structural and functional information of the CNGA1 channel. Moreover, this combined method not only provides new and original information, but also opens the possibility to apply the same analysis to several other membrane proteins of biological interests. The conclusions drawn from performed experiments, and the perspectives offered by the present study, are summarized below.

I have successfully applied single molecule force spectroscopy in combination with bioinformatics, mutagenesis and electrophysiology to investigate the structural properties of CNGA1 channels from overexpressed oocyte plasma membrane and ROS plasma membrane from *Xenopus laevis*. The unfolding of the CNGA1 channels shows significantly different unfolding pathways between the open and closed state, thus providing information on changing conformations during the allosteric transition.

SMFS experiments provide a huge amount of stochastic data (force-distance curves or F-D curves). I have introduced a novel clusterization technique, where individual F-D curves are coded in a sequence of symbols. This analytical tool has been successfully tested in different experimental conditions and provided a good way to handle the data and spot the relevant information.

In the first part of my work, I have successfully identified the unfolding of a single subunit of CNGA1 channels from oocyte plasma membrane. The channel shows different unfolding pathways in the nucleotide binding domain as well as in the trans-membrane domain, suggesting the conformational changes in those domains. The channel in the open state shows significant increase in mechanical coupling between the S5 trans-membrane (TM) domain and voltage sensor S4, thus suggesting a favorable energetical stability between these two domains. SMFS in the closed state identified a fall in mechanical coupling at a target residue of Proline (293) in the S4-S5 loop. Further, electrophysiological recording on a mutant channel P293A (which is in the S4-S5 loop) shows channel deactivation in the presence of ethylammonium. All these results are suggesting a

structural transition of the S4-S5 loop during the channel opening.

The unfolding force of the CNGA1 channel in the open state is ~20-30 pN higher than that in the closed state. This increase in the unfolding force can be due to an increase in hydrophobic and/or electrostatic interactions between different TM domains. In both cases the trans-membrane domain has higher order of mechanical and energetical stability in the open state.

In the second phase of my work, I have introduced a new clustering technique, where in place of taking F-Lc as clustering coordinate, I placed it into Lc- ΔLc coordinated, where, distribution in ΔLc increases the efficiency to detect the rare events even in between two most reproducible force peaks. With the application of this clustering technique I have successfully identified different unfolding pathway for CNGA1 channel both for open and closed states.

In the final phase or my work, I have performed SMFS experiments on ROS plasma membrane, where the unfolding of rhodopsin, CNGA1 and CNGB1 channels. The rhodopsin from plasma membrane shows difference in the unfolding pathway than that observed for rhodopsin from disc membrane. My results suggest that rhodopsin in the plasma membrane has a higher mechanical stability than that in the disc membrane. As previously proposed (Kathleen & Albert, 1990 a,b), high cholesterol concentration in the plasma membrane is directly responsible for rhodopsin inhibition. On the basis of those observations, I propose that, a higher concentration of cholesterol in the ROS plasma membrane decreases the mobility of rhodopsin by increasing the mechanical stress on the lipid-protein interface.

Unfolding of the CNGA1 channel from ROS plasma membrane shows almost a similar pathway as I have observed for the CNGA1 channel expressed in oocyte plasma membrane. CNGB1 channel unfolds almost in a similar way of the CNGA1 channel, but with a different unfolding of S5 domain.

An important issue and improvement for SMFS is to couple it to direct chemical analysis such as Raman spectroscopy so to obtain a direct determination of the secondary structure of the unfolded domains. The combination of these two techniques will represent a major technical improvement for single molecule studies and I hope I will be involved in this adventure.

Bibliography

Boesze-Battaglia K , S J Fliesler and A D. Albert. Relationship of cholesterol content to spatial dis tribution and age of disc membranes in retinal rod outer segments. *J. Biol. Chem. 1990, 265:18867-18870.*

Boesze-Battaglia K. and Albert A.D. Cholesterol modulation of photoreceptor function in bovine retinal rod outer segments. J. Biol. Chem. 1990, 265:20727-20730