

Computational study of Sonic Hedgehog proteins: insights into the role of metal ions

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

Dr. rer. nat.

der Fakultät für

Biologie

an der

Universität Duisburg–Essen

vorgelegt von

Rocio Rebolledo–Rios

aus Cienfuegos, Kuba

Februar 2014

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden am Zentrum für Medizinische Biotechnologie (ZMB) in der Abteilung für Bioinformatik der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Daniel Hoffmann

2. Gutachter: Prof. Dr. Andrea Vortkamp

Vorsitzender des Prüfungsausschusses: Prof. Dr. Jens Boenigk

Tag der mündlichen Prüfung: 03. Juli 2014

Contents

	Page
Abbreviations	III
List of Figures	VII
List of Tables	IX
Abstract	XI
Zusammenfassung	XII
Acknowledgements	XIII
1 General introduction to Hedgehog morphogen family	1
1.1 Discovery, evolution and function	1
1.2 Hedgehog signaling transduction pathway	3
1.2.1 From processing to reception	3
1.2.2 Mechanisms of release and long-range signaling	5
1.2.3 Role of dual lipidations in the amino-terminal domain	7
1.2.4 Protein modulators of the signaling	8
1.2.5 Dysregulation of the pathway	10
1.3 Distinctive structural features of Hedgehog proteins	12
1.3.1 Zinc binding site	12
1.3.2 Calcium binding site	15
1.4 General remarks on the Hedgehog signaling	16
1.5 Motivation and aims of the current research	16
2 Application of Molecular Dynamics methods in the study of biomolecules	19
2.1 Brief overview of computational methods	19
2.2 Molecular dynamics theory	20
2.2.1 Empirical force field models	21
2.2.2 Optimized Lennard-Jones parameters for Ca ²⁺ interactions	23
2.2.3 Periodic boundary conditions	23
2.2.4 Explicit solvent model	25
2.2.5 Implicit solvent model	25
2.2.6 Replication of experimental conditions	27

2.3	Brownian Dynamics	28
2.3.1	Biomolecular association reactions	28
2.3.2	BrownDye package for Brownian Dynamics	29
2.4	Analysis of trajectories	30
2.4.1	Root mean square deviation	31
2.4.2	Root mean square fluctuation	31
2.4.3	Temperature factors derived from RMSF	31
2.4.4	Cluster analysis	32
2.4.5	Principal Components Analysis	32
2.4.6	Buried surface area calculation	33
2.4.7	Hydrogen bond calculations	33
3	Calcium binding regulates putative Sonic Hedgehog peptidase	35
3.1	Introduction	35
3.2	Computational Methods	36
3.3	Results and Discussion	38
3.4	Conclusions	58
4	Molecular insights into Sonic Hedgehog dimerization	61
4.1	Introduction	61
4.2	Computational Methods	62
4.3	Results and Discussion	64
4.4	Conclusions	81
5	Outlook and remaining challenges	83
	Appendices	85
	References	97
	List of Publications	117
	Declarations	119

Abbreviations

ADAM	A Disintegrin And Metalloprotease family
AMBER	Assisted Model Building with Energy Refinement
BDA1	Brachydactyly type A-1
BD	Brownian Dynamics
Boc	Brother of Cdo
Boi	Brother of Ihog
BSA	Buried Surface Area
Ca0, Ca1, Ca2	Different calcium states respect to Ca ²⁺ binding
Cdo	Cell adhesion molecule-related/downregulated by oncogenes
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CPU	Central Processing Unit
CS	Chondroitin Sulfate
CW	Cardin-Weintraub motif
Dhh	Desert Hedgehog
Disp	Dispatched, the transmembrane transporter-like protein for release of Hh from secreting cells
DNT	ShhN dimers without the N-terminal extension, residues G25 to K38
DT	ShhN dimers with the N-terminal extension, residues G25 to K38
ECM	Extracellular Matrix
EGF	Epidermal Growth Factor
ER	Endoplasmic Reticulum
FF	Force Field
FNIII	Extracellular Fibronectin type III domain
FRET	Fluorescence Resonance Energy Transfer
Fz	N-terminal Frizzled-like cysteine-rich domain
GAG	Glycosaminoglycan. Linear, sulfated, negatively charged polysaccharides such as: heparin and heparan sulfate

Gas1	Growth-arrest-specific gene 1
GFR	Glial cell-derived neurotrophic factor family receptor
GPI	Glycosylphosphatidylinositol, which anchors glypicans to cell surface
Hhat	Hedgehog acyltransferase
HhC	C-terminal domain of Hedgehog protein
Hhip	Hedgehog interacting protein
HhN	N-terminal domain of Hedgehog protein
Hh	Hedgehog
HPE	Holoprosencephaly
HSPGs	Heparan Sulfate ProteoGlycans, in which two or three heparan sulfate chains are attached in close proximity to cell surface or extracellular matrix proteins
HS	Heparan Sulfate
Ig	Immunoglobulin
Ihh	Indian Hedgehog
Ihog	Interference hedgehog protein
LAS	Lysostaphin-type enzymes, D-Ala-D-Ala metallopeptidases and Sonic hedgehog arrangement
MD	Molecular Dynamics
MM	Molecular Mechanics
OMIM	Online Mendelian Inheritance in Man. An open public database of human genes and genetic disorders
OPLS	Optimized Potentials for Liquid Simulations
PBC	Periodic Boundary Conditions
PCA	Principal Components Analysis
PDB	Protein Data Bank
PME	Particle Mesh Ewald
Ptc	Patched, a twelve-pass integral membrane protein
QM	Quantum Mechanics
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SASA	Solvent Accessible Surface Area
ShhN	N-terminal domain of Sonic Hedgehog protein
Shh	Sonic Hedgehog
Ski	Skinny Hedgehog protein, member of the membrane-bound O-acyltransferase

	(MBOAT) protein family
Smo	Smoothed, a seven-pass integral membrane protein member of the G protein-coupled receptor superfamily
SPCE	Simple Point Charge Extended
SPC	Simple Point Charge
SRR	Sterol Recognition Region
SS	Signal peptide for Hh protein export
Wnt	Secreted lipid-modified signaling glycoproteins. The name Wnt is resultant from a fusion of the name of the <i>Drosophila</i> segment polarity gene <i>wingless</i> and the name of the vertebrate homolog, <i>integrated</i> or <i>int-1</i>

List of Figures

	Page
1.1 Evolutionary relationship between members of the Hh family	2
1.2 Simplified model of Hh release, reception and signal.	4
1.3 Hh processing reaction and lipid modifications.	6
1.4 Hh binding partners characterized to date	9
1.5 BDA type A1 phenotype of an affected individual.	11
1.6 Zinc and calcium binding sites highly conserved in Hh family	13
1.7 Schematic 2-D representation of metal centers in murine ShhN	14
1.8 Proximity of residues accounting for BDA1 and HPE diseases.	15
2.1 Balls and spring model	20
2.2 The leap-frog integration method	21
2.3 Periodic boundary conditions	24
2.4 The SPC/E explicit water model	26
2.5 Schematic representation of BD simulations.	30
2.6 Solvent-accessible surface area of a molecule.	33
3.1 RMSF of ShhN backbone atoms	39
3.2 PCA plot comparing conformational dynamics of calcium states	40
3.3 Effect of Ca ²⁺ binding on ShhN structure	41
3.4 Time evolution of RMSDs backbone atoms from murine ShhN	42
3.5 Effect of zinc ion on the stability of Shh	43
3.6 Dependence of ionic strength on calcium association rate	44
3.7 Zinc centers of different peptidases and ShhN	45
3.8 RMSDs between ShhN and LAS enzymes	46
3.9 Switch mechanism triggered by Ca ²⁺ calcium ion	48
3.10 Effect of calcium binding on the ShhN geometry	49
3.11 Distance between zinc ion and oxygen of water molecule	50
3.12 Angles between oxygen water, zinc ion and zinc ligands	51
3.13 Phylogenetic tree of Hedgehog proteins	53
3.14 Electrostatic potential on SAS of ShhN in three calcium binding states	54
3.15 Ligand variations of E91K mutant in the first coordination shell of calcium ions	58
4.1 ShhN _{dimer} model used	62

4.2	Thermodynamic cycle of the binding free energy calculation	64
4.3	Buried surface area of ShhN _{dimers}	65
4.4	Stable hydrogen bonds at ShhN _{dimer} interfaces	66
4.5	Electrostatics of ShhN _{dimers} with 0, 1 and 2 calcium ions.	67
4.6	Electrostatic contribution to the binding free energies of ShhN _{dimers}	68
4.7	Comparison of ShhN _{dimers} in DNT0 and DNT2 complexes	69
4.8	Location of residues I48 and V166 at the ShhN dimeric interface	70
4.9	Comparison of mutant _{dimers} I48A and V166A regarding to DT0 complex	71
4.10	Conservation of H-bonds in MD simulations of DT0, I48A and V166A	72
4.11	Aligned clustered structures taken from trajectories of DT0, I48A and V166A respectively.	73
4.12	RMSDs between ShhN and calcium binding proteins	74
4.13	Structural analysis of ShhN-GAG interactions	76
4.14	Analysis of the crystal lattice contacts of X-ray structure 4C4m	78
4.15	Schematic representation of the proposed model for ShhN multimerization	80
C1	B-factors from MD simulations of ShhN	88
C2	Docking of cholesterol to ShhN (1vhh)	90
D3	Correlation between the number of residues at the interfaces and the BSA	93

List of Tables

	Page
3.1	Estimated second-order rate constants of calcium ions 43
3.2	Estimated second-order rate constants between ShhN and regulators 56
4.1	BSA of residues I48 and V166 70
A1	Chemical structure of amino acids 85
B1	Crystal structures of Hh proteins available in the PDB 87
C1	Mean values of RMSFs and p-values from Wilcoxon rank sum tests 88
C2	P-values of RMSD comparisons between zinc center of LAS enzyme X-ray structure 1lbu and zinc centers of ShhN. 89
C3	As Table C2, but referring to X-ray structure of LAS enzyme 2v09. 89
C4	As Table C2, but referring to X-ray structure of LAS enzyme 1u10. 89
C5	As Table C2, but referring to X-ray structure of LAS enzyme 1r44. 89
C6	As Table C2, but referring to X-ray structure of LAS enzyme 2bop. 89
C7	P-values of RMSD comparisons between zinc center of LAS enzyme structure 1lbu and zinc centers of ShhN 89
C8	As Table C7, but for comparison of LAS enzyme 2v09 and ShhN. 89
C9	Distribution of distances between carboxylate-O of E127 and imidazole proton of H135 90
C10	Distribution of distances between side chains of H135 and E177 90
C11	Summary of docking results of cholesterol to the surface of ShhN 91
C12	Summary of docking results of cholesterol to the surface of ShhN including a water molecule 92
C13	Predicted pK_a values for calcium ligands 92
D1	P-values of BSA comparisons between calcium states of DNT _{dimers} 93
D2	As Table D1, but referring to ShhN _{dimers} with tail. 93
D3	P-values of binding free energy comparisons between calcium states of DNT _{dimers} 93
D4	As Table D3, but referring to ShhN _{dimers} with tail. 93
D5	P-values of BSA comparisons between DT0 _{dimer} and I48A, V166A mutants . 94
D6	P-values of RMSD comparisons between DT0 _{dimer} and I48A, V166A mutants 94
D7	P-values of RMSF comparisons between DT0 _{dimer} and I48A, V166A mutants 94
D8	P-values of binding free energy comparisons between DT0 _{dimer} and I48A, V166A mutants 94

D9	P-values of a H-bond comparison (K178-E167) between DT0 _{dimer} and I48A, V166A mutants	94
D10	As Table D9, but referring to a H-bond between Y44 H-E167 OE1.	94
D11	As Table D9, but referring to a H-bond between N50 OD1-R153 NH2.	94
D12	As Table D9, but referring to a H-bond between N50 NH-D152 O.	94
D13	As Table D9, but referring to a H-bond between F170 O-R155 NH.	95
D14	As Table D9, but referring to a H-bond between E167 OE1-S177 OH.	95

Abstract

The present work studies the effect of metal ions on the structure, dynamics and interactions of the protein ShhN, the N-terminal signaling domain of Sonic Hedgehog. To accomplish this task, molecular structures of ShhN proteins were analyzed with a set of computational methods, revealing new features of ShhN proteins.

The results suggest that, ShhN is an enzyme with a zinc catalytic center that is regulated by the binding of the calcium ions. Explicitly, the binding of the second calcium ion involves a conformational change that is accompanied by a significant perturbation of the putative catalytic center, possibly affecting substrate stabilization. The dragging of E127 towards the calcium center implies the pulling of H135 with it and the disruption of the hydrogen bond between G128 and H141. Besides, the distance between residues E177 and H135 increases and therefore, the well-defined position of the catalytic water molecule is lost destabilizing the zinc environment. Electrostatic potential differences among calcium states suggest the possible binding of nonpolar substrates. One of the predictions is that ShhN autodegradates tuning its own concentration gradient. This possibility does not rule out, of course, the existence of other mechanisms that govern ShhN concentration gradient. The novel switching mechanism proposed could have many implications in the biological function of HhN proteins, but these are not well understood and require further research.

Both ShhN monomers and dimers show a flexibility pattern that strongly depends on the number of calcium ions. Specially, calcium binding loops reflect this behavior. The Cardin-Weintraub motif located within the N-terminal of ShhN proteins together with buried hydrophobic residues at the interface lead to an stable complex that enhances ShhN dimerization. The lower degree of conservation of I48 in vertebrate homologs might indicate that this is a hot spot residue with an important role in ShhN oligomerization. The presence of the calcium ions at the dimeric interface can promote ShhN-proteoglycan interactions providing a large positively charged region which is the ideal scenario for the binding of these molecules.

Taken all together, a multimerization model where different levels of interaction can control the way that ShhN multimers form is proposed. However, this model has yet to be tested.

Zusammenfassung

Die vorliegende Arbeit untersucht die Frage: Welche Wirkung haben Metall-Ionen auf die Struktur, Dynamik und Wechselwirkungen des Proteins ShhN, der N-terminalen Signal-Domäne von Sonic Hedgehog? Zur Beantwortung wurden molekulare Strukturen von ShhN mit verschiedenen Computer-gestützten Methoden analysiert. Die Analysen sagen einige überraschende Eigenschaften von ShhN vorher.

Zuerst legen die Analysen nahe, dass ShhN eine Zink-Protease ist, deren katalytisches Zentrum durch die Bindung von Calcium-Ionen reguliert wird. Speziell ändert die Bindung des zweiten Calcium-Ions die Konformation des Zink-Zentrums und beeinflusst so vermutlich die Bindung des Substrats. Diese Konformationsänderung kaskadiert vom Calcium-Ion zum katalytisch Zink-Zentrum über mehrere Residuen und destabilisiert die Substrat-Bindestelle und das Zink-gebundene katalytische Wasser-Molekül. Änderungen des elektrostatischen Potenzials um das Zink-Zentrum in Abhängigkeit von der Calcium-Bindung weisen hin auf eher unpolare Substrate, z.B. den C-Terminus von ShhN. ShhN könnte also ShhN-Moleküle abbauen, und auf diese Weise seinen Konzentrationsgradienten ändern. Die Implikationen des beschriebenen neuartigen Mechanismus der Schaltung von ShhN über Calcium-Ionen für die biologische Funktion des Proteins sind noch weitgehend unverstanden, und erfordern weitere Forschung.

Die Flexibilität von ShhN-Monomeren und -Dimeren hängt stark ab von der Zahl der gebundenen Calcium-Ionen. Im Besonderen sind die Calcium-bindenden Schleifen von ShhN davon betroffen. ShhN-Dimere werden stabilisiert durch das Cardin-Weintraub-Motiv am N-Terminus und durch Wechselwirkungen mit hydrophoben Residuen an der Schnittstelle zwischen den Monomeren. Die schwächere Konservierung von I48 in Wirbeltieren könnte ein Hinweis sein auf eine wichtige Rolle dieses Residuums in der Oligomerisierung von ShhN. Die Nähe von Calcium-Ionen zur Dimerisierungs-Region sorgt für ein positives elektrostatisches Potenzial und könnte so die Bindung von Proteoglykanen fördern. Insgesamt wird ein Modell der Multimerisierung von ShhN vorgeschlagen, das über mehrere Kontrollmechanismen verfügt, die experimentell getestet werden sollten.

Acknowledgements

It has been a long way (literally there is a long way from Cuba to Germany, about 8.000 km) since the first time I was inside these walls. I have experienced many things: I have laughed, smiled, got angry, fell down, cried, and I have missed many people, but in the end, I am very pleased with the result. To be honest, it is still hard to believe that I am sitting here writing the last two pages of this thesis. **THE DAY HAS COME!** However, it is always a challenge for me to express in black and white all my gratitude for everyone who contributed to making this possible.

I cannot express enough thanks to my supervisor [Daniel Hoffmann](#) for his continued support and encouragement. For giving me the chance to be part of this group, to learn from him every day (from his endless knowledge) and for giving me the opportunity to realize, once again, how much I love science. I hope I was up to the task. Thank you very much Daniel!

I would like to thank [Andrea Vortkamp](#) for all the useful scientific discussions and for agreeing to review this thesis.

My sincere thanks to a very important member of the Bioinformatics Department, to you [Claudia](#). Thank you for being so kind to me, so patient, and always answer all my questions and requests. I really appreciate all that you do for me. I do think that, all research groups should have someone like you.

Many thanks to all my former and newer [colleagues](#) from the Bioinformatics Department for all seminars, ideas, cookies, birthday's cakes and, for sure, all coffees shared. But especially, to [Niko](#) and [Dominik](#). Thank you for your friendship, for all the advices, beers, barbecues, corrections (Dominik) and many other things we have shared. I hope to see you both very soon!

Thanks to [Velina](#) for the scientific and non scientific talks. Be patient, your time is coming soon. Keep going! I almost forget it... If you still want to go to Cuba, remember that, one of the most beautiful islands in the Caribbean is waiting to share all the colors, music and fun with you.

To my dear friends, [Eileem](#), [Gretel](#), [Vane](#), and [Yiri](#) (in alphabetic order): thank you so much girls! Eileem and Yiri: thanks for the countless time we have spent over the phone (just the German Telephone Company really knows this) talking about whatever and making me feel much better every time I have a problem. Eileem: I will miss the "Habana Lliure" but, I miss much more to go out with you. So, the next time in Barcelona will be unforgettable. Gretel: Lia is beautiful! I am looking forward to meet her (perhaps this year). A big hug for you three! Vane: the distance does not count in more than 15 years of friendship. I always feel like it was yesterday the last time I saw you. We just need a few hours to catch up and laugh like it is the last day on earth. Thanks to [Raúl](#) for his share in this adventure and for all the years he has been carrying with me. See you both soon! Yiri: now you are so far away, [Gio](#) took you there and I am angry at him but, since I have the condition of "second wife", I have the privilege to go there as much as I want and you both have to deal with it. Thank you both for that lovely wedding and

have had the opportunity to share it with you. Eu adoro a música brasileira!

Thanks to my friend [Kenia](#) for all the good and not that good moments we have shared. What does not kill you makes you stronger, right? Thanks for writing me over these years even when I do not reply. Sofia is growing so fast, probably she does not recognize me next time. We will see!

I wish to acknowledge the help provided by [Roxana](#). Thanks for correcting part of this thesis, for all the advices and talks. We have to celebrate!

Thanks to my “Ausländer” friends, [Dani](#), [Jose](#) and [Manlio](#) for all dinners, games, interesting talks and tv shows (impossible to forget the last one).

Many thanks to [Jörn](#) (my tandem partner). I am pretty sure you will die laughing when you read this. If it is completely true, no doubts, my german should be amazing, isn't it? Danke für alles. Nos vemos en Mallorca?

Thanks to [Teodora](#) and [Jack](#) as well, for all the laughs, thoughts and for the great time we spent together. We have to meet more frequently!

Almost last but not least, to my beloved and unforgettable grandparents [Elsa](#) and [Anibal](#). In simple words: this is your legacy. Unfortunately, I cannot share this moment with you but, I hope you both are proud of me. Perhaps I can not live enough to say thanks to my parents [Elsita](#) and [Chinea](#). Thanks for always being there for me, unconditionally! I know mom, I am not that daughter that is always writing or calling but believe me when I say that, there is not a single day that goes by that I do not think of you. Te quiero y te extraño un montón! Thanks to [Mara](#) for being my little sister, it is good to have close!

Finally, to my extraordinarily patient and loving [Juanma](#). For being also a part of this. Gracias por todo Gordi, Te quiero!

“Our greatest weakness lies in giving up. The

most certain way to **succeed** is always to
try just **one** more time.”

– Thomas A Edison –

“Begin **at the
beginning,” the King
said gravely, “and
go on **till you
come to the **end:**
then stop.”****

– Lewis Carroll –

CHAPTER 1

GENERAL INTRODUCTION TO HEDGEHOG MORPHOGEN FAMILY

1.1 Discovery, evolution and function

In the late 1970s, Hedgehog (Hh) gene was first discovered by two afterwards Nobel laureates: Christiane Nüsslein-Volhard and Eric Wieschaus [1]. Through genetic screens they were able to identify genes that control the segmentation pattern in *Drosophila melanogaster* embryos. Mutations in the Hh gene cause a phenotype that resembles a hedgehog spines, fact to which the gene owes its name. Later on, in the early 1990s, the isolation of the Hh gene revealed that it encodes for a secreted protein expressed in segmentally repeated stripes, consistent with its role in specifying segmental pattern [2–4].

Hh proteins are synthesized as precursor proteins (about 400-460 amino acids long) and are composed of a signal peptide for protein export (SS) and two domains: an amino-terminal "Hedge" domain (HhN) which has the biological signal activity, and a carboxy-terminal autocatalytic "Hog" domain (HhC) with not known signaling role. HhC binds cholesterol in the sterol recognition region (SRR) [5] and cleaves Hh into two parts in an intramolecular reaction which also adds a cholesterol moiety to HhN. The determination of the X-ray structure of the *Drosophila* HhC (see Appendix B) confirmed that HhC has sequence similarity to the self-splicing inteins [6] and the shared region was termed Hint module. Inteins are protein sequences that autocatalytically splice themselves out of longer protein precursors (analogous to introns) and ligate the flanking regions into a functional protein. Although Hh proteins are the only proteins known to be conjugated to cholesterol in this way, the widespread distribution of the Hog domain indicates that its role in processing secreted proteins must be an evolutionary ancient one. Among the organism where Hog domain proteins have been identified are red algae, dinoflagellates and mosses, as well as throughout the *Metazoa* (or *Animalia* kingdom) [8]. However, the highly conserved Hedge domain, seems to be of more recent origin [9]. A protein with

a similar domain organization is encoded by the so-called Hedgling gene in the basal *Metazoan*. The *Nematostella vectensis* (the sea anenome) which belongs to the phylum *Cnidarian* (animals found exclusively in aquatic and mostly marine environments) also has a Hedgling gene. Likely, Hh proteins first arose in the common ancestor of the *Cnidarians* and the *Bilateria* (all animals having a bilateral symmetry) more than 650 million years ago through the combination of domains that were present in pre-existing proteins. In general, the Hh gene family is present throughout the *Eumetazoa* (comprises almost all major animals), although it has been lost in some nematodes [10].

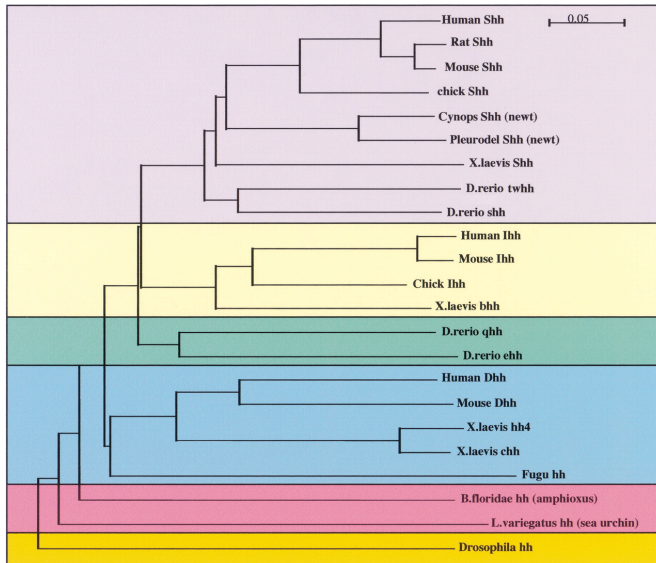


Figure 1.1: Evolutionary relationship between members of the Hh family from different vertebrate and invertebrate species. Phylogenetic tree taken from Ingham & McMahon [7]. The single Hh gene present in *Drosophila* is colored in yellow and other invertebrates such as sea urchin and amphioxus are colored in pink. The three Hh genes identified in vertebrates Shh, Ihh and Dhh are colored in lilac, light yellow and blue respectively. The zebrafish (*Danio rerio*) has two Shh-related genes, shh and twhh (tiggy-winkle hedgehog) and two other Hh-related genes, ehh (echina hedgehog) and qhh (qiqihar hedgehog) which are divergent members of the Ihh subgroup (green). This suggests that further duplication events occurred within the Shh and Ihh classes. One Dhh gene appears to be more closely related to amphioxus and *Fugu hh* than to other Dhh-related genes. Additional terms; bhh: banded hedgehog and chh: cephalic hedgehog.

and metamorphosis [16]. In vertebrates, Shh, Dhh, and Ihh mediate many developmental patterning events and have partially redundant functions. Shh is the most widely expressed of the three and regulates development from embryo to adult [17]. It is first expressed in the notochord, and later in the floor plate of the neural tube, where produces a gradient of activity in the ventral neural tube. Shh is also expressed in the zone of polarizing activity of the limb buds and is important for limb and digit formation. Other roles of Shh include inner ear, eye, taste bud, and hair follicle development. Ihh is required for bone growth and pancreas development [7, 18]. It has been implicated in

Shortly after the Hh gene was cloned, vertebrates homologs were reported. Unlike the fly, which has a single Hh gene, there are several related genes in vertebrate species. Three genes were identified in mouse, called Sonic Hedgehog (Shh), Indian Hedgehog (Ihh), and Desert Hedgehog (Dhh) [11–14], which are similarly processed, modified and released. They share about 60% amino acid identity of the full length sequences, and 90% amino acid identity in the HhN domain. Figure 1.1 shows the evolutionary relationship in the Hh family and it can be appreciated that Dhh is most closely related to *Drosophila* Hh while Ihh and Shh are more related to each other. This is probably the consequence of a more recent gene duplication event [7, 15].

In invertebrates animals e.g., fruit flies of the genus *Drosophila*, Hh is important in segmentation polarity and it is also involved in later stages of embryogenesis

modulating chondrogenesis in the appendicular skeleton and acts as a negative regulator of the differentiation of proliferating chondrocytes [19]. Both Shh and Ihh play roles in cardiovascular development [20]. Dhh expression is limited to the peripheral nervous system and reproductive organs; Dhh mutant males are sterile [21].

The major components of the Hh pathway are conserved in diverse *Eumetazoa*, including the non-bilaterian *Cnidaria*. Over the next sections, general and relevant aspects of the Hh signaling are discussed.

1.2 Hedgehog signaling transduction pathway

Hh molecules are morphogens that spread from localized sites of production to trigger differential cellular responses dependent on morphogen concentration [22]. The range of a signal is the domain over which it exerts its effects. In other words, the signaling range is the distance from the source at which a response is observed [23, 24]. Shh is the best established example of a morphogen as defined by Lewis Wolpert's French flag model [25]. As a result of its widespread expression, much of what is known about vertebrate Hh signaling stems from work on Shh. The signaling activity of Hh ligands is intimately linked to several post-translational modifications and cleavage events that modify their activity and regulate their spread from producing cells through tissues. The tight control of secretion is a key step in the regulation of signaling activity [26].

The mechanism of Hh signaling is complex and remains incompletely understood despite, overall it is conserved from fly to human [27]. The Hh pathway has been and perhaps remains the most enigmatic one. To provide some context, let's start by citing the unusual biochemistry of the Hh family e.g., the two lipid modifications of the HhN, the role of Patched (Ptc) as a receptor and the lack of direct interaction among Hh or Ptc and Smoothed (Smo). These facts together with the difficulty to reproduce *in vitro* some physiologically relevant interactions of the Hh signaling components have been impediments to progress. Nevertheless, new evidences have emerged in the last two decades, and many questions have been answered. A simplified model of the Hh signaling based on Shh is described bellow (subsection 1.2.1) and showed in Figure 1.2. For a better understanding of the whole text bellow, please see Appendix A where a list of all amino acids with one and three letter code is given.

1.2.1 From processing to reception

Hh proteins are synthesized as precursors of about 45 kDa. After signal sequence cleavage, it undergoes an autocatalytic processing reaction to yield a ~19 kDa HhN domain with a cholesterol modification and a ~25 kDa HhC domain (Figure 1.3). The cleavage reaction takes place in the endoplasmic reticulum (ER) and can be divided in two steps: In the first step, the free thiol of C198 (human Shh) acts as a nucleophile, attacking the carbonyl carbon of the preceding Gly residue and generating a thioester intermediate. In the second step, this thioester intermediate is subjected to nucleophilic attack by the 3 β -hydroxyl group of cholesterol, resulting in the covalent attachment of cholesterol to HhN and releasing the HhC domain [28]. Although C198 has long been recognized for

its role in autocatalytic cleavage, a second conserved cysteine C363 is also required for cleavage, forming a disulfide bond with C198 that likely facilitates protein folding and the reduction of which generates the reactive thiol required for cleavage. As such, mutating either cysteine residue prevents autoproteolysis of Hh precursors [29]. *In vitro* evidences suggest that the HhC domain undergoes rapid degradation, a process that requires its translocation out of the ER to the cytosol and subsequently to the proteasome [30].

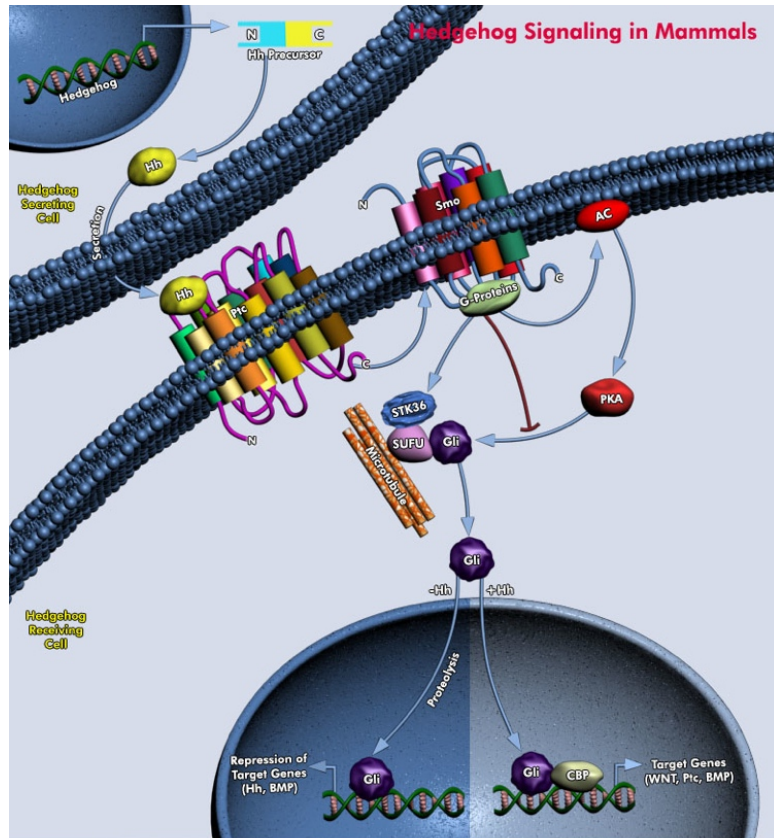


Figure 1.2: Simplified model of Hh release, reception and signal in mammals. Hh protein undergoes a series of processing steps before it is secreted from the cell. The HhN domain is dually lipidated with a cholesterol molecule added to the C-terminus and a palmitate added to the N-terminus. Once secreted, Hh contact Ptc (the major receptor for Hh proteins). In presence of Hh ligands, Ptc is unable to affect SMO signaling and SMO somehow can promote formation of activated forms of Gli, resulting in upregulation of Hh target genes. Courtesy: QIAGEN Sample & Assay Technologies.

Cholesterol-modified HhN domain enters the secretory pathway where it is subjected to a second covalent modification. A transmembrane acyltransferase named Skinny Hedgehog (Ski) in *Drosophila* [31–33] and Hedgehog acyltransferase (Hhat) in mammals [34] catalyses this event. Hhat modifies HhN by the attachment of palmitate (via an amide bond) to the free amino group of the N-terminal cysteine [35]. The subcellular compartment in which palmitoylation occurs has yet to be identified. Dually lipid-modified HhN is associated with the lipid bilayer at the outer surface of the plasma membrane, until it is released. Different models have been proposed to explain the release and spread of Hh proteins [36–38] (see subsection

1.2.2 for further details).

Once HhN reaches its target cell, its key function as an extracellular signal is to bind the Ptc receptor. The complex is internalized while Smo translocates to the primary cilia (in vertebrates) [39]. Localization of Smo to the primary cilia is a fundamental requirement for the pathway to be active. In the absence of HhN, Ptc inhibits this localization [40]. How exactly Ptc inhibits Smo and how Hh relieves this inhibition, is still not clear, although the ability of Ptc to inhibit a large stoichiometric excess of Smo suggests that this

inhibition does not occur through direct contact between Ptc and Smo [41]. Activated Smo is phosphorylated and signals via a cascade of microtubule-associated proteins to the nucleus, where zinc-finger proteins of the Gli¹ family of transcription factors, activate or repress target genes [42]. Among the many target genes regulated by mammalian Gli1 are those for Ptc and Gli1 themselves. This results in feedback loops in which upregulation of Ptc leads to negative feedback, whereas upregulation of Gli1 leads to positive feedback.

1.2.2 Mechanisms of release and long-range signaling

Native HhN is modified by cholesterol and palmitate, as described previously. Both hydrophobic modifications are involved on the ability of HhN to associate with cellular membranes, a property that initially appeared inconsistent with its ability to act far from its site of synthesis. HhN was originally considered to be mainly responsible for short-range signaling [43]. Evidence of freely diffusible HhN forms that travel a distance up to 50 μm in the imaginal wing disc of *D. melanogaster* and up to 300 μm in the limb bud of vertebrates [26] raise critical questions about the way HhN is secreted and how it can reach cells located at a distance from the signaling cell. Although it is now clear that Hh family members are capable of acting directly in long-range signaling; there is no agreement about the existence of a unique mechanism for the release and spread. Several models to release HhN from cellular membranes with its lipid moieties intact are discussed below.

One of the reported models is the HhN-oligomers formation. Monomeric HhN can multimerize with itself to form large soluble multimers that are released from the membrane [36]. The lipid moieties are thought to be embedded in the core of these complexes [44]. This soluble multimer would then diffuse far from its site of synthesis to engage directly in long-range signaling. The exact number of HhN molecules in the multimers is unknown, as is the requirement, if any, of additional factors that may act as intermediate between the various HhN molecules. It has been suggested that molecular weight of HhN complexes is ranging from approximately 100 up to 700 kDa [45]. Nevertheless, how the multimers form remains an unsolved question. Several studies [46–49] have reported the implication of specific residues in HhN multimerization. Among them the highly conserved N-terminal of HhN, E71, R72 and K74, as well as residues R123, D152 and R155 (human nomenclature) have been claimed to play an important role. Evidence support the existence of oligomers *in vivo*. Fluorescence resonance energy transfer (FRET) microscopy analysis in *Drosophila* revealed the formation of HhN oligomers grouped into visible clusters [47]. The authors suggest that the nanoscale organization of HhN at the cell surface is specified by electrostatic interactions between subunits and it is relevant for the generation of visible clusters and, ultimately, HhN *in vivo* activity [50]. Additional experiments need to be performed to characterize these complexes anyhow.

¹So-called because the gene encoding the founding member GLI1 is frequently amplified in glioblastoma cells.

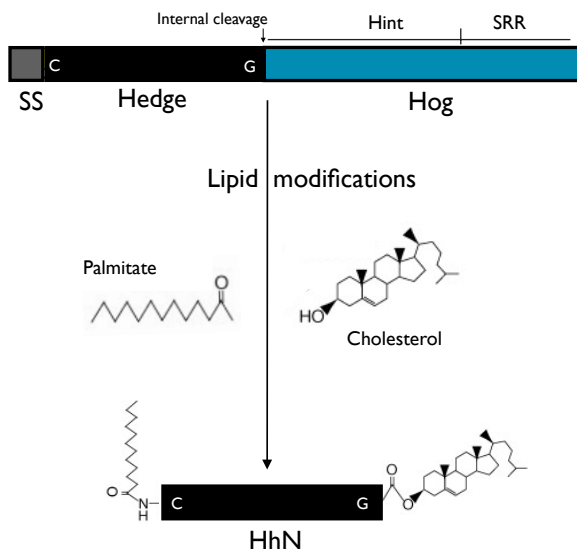


Figure 1.3: Hh processing reaction and lipid modifications. Hh is synthesized as a precursor protein that undergoes a series of posttranslational modifications before it is secreted from the cell. A short signal peptide sequence (SS) is responsible for the translocation of Hh into the ER. Once there, an autocatalytic processing reaction takes place to generate both the Hedge and Hog domains. The cleavage reaction is mediated by the Hog domain (HhC), which also couples cholesterol covalently to the C-terminus of the Hedge domain (HhN). A second lipid modification that incorporates palmitic acid occurs at the N-terminus cysteine that is exposed after signal peptide cleavage.

onic tissues, and thus whether they play a role in Hh signaling at that stage.

The Dispatched (Disp) [55, 56] and the Scube family of proteins seem to be also required for HhN long-range signaling. Disp is a multispanning membrane protein, belongs to the RND family of transporters [57] and is required for the secretion of cholesterol-modified HhN proteins from vertebrate cell surfaces. The Scube family [58] consists of the secreted proteins Scube1, Scube2, and Scube3. Scube2 may be involved in the transport or stability of HhN ligand in the extracellular space [59, 60]. Disp and Scube recognize different structural aspects of cholesterol, suggesting a handoff mechanism for transferring HhN from Disp to Scube. Thus, they cooperate to promote the release of HhN from the cell surface [61, 38]. It is unclear in which cellular compartment Disp is required, but it is tempting to speculate that Disp transfers cholesterol-modified HhN to Scube2 at the plasma membrane, so that the cholesterol anchor of HhN is shielded from the aqueous environment. How the hydrophobic character of dual lipid adducts is accommodated in the soluble form of HhN released by Disp/Scube remains to be determined. One possibility is the formation of a micelle-like complex with a hydrophobic core of interacting lipids, as previously suggested by Ref. [36, 44]. Also, the mechanism by which the Disp and Scube proteins promote long-range Hh signaling is unknown.

The association of HhN with lipoproteins (the major lipid carriers in circulation) have been proposed as another mechanism for the release of dually lipidated HhN in vertebrates and invertebrates [37, 51, 52]. HhN oligomers can interact with heparan sulfate (HS) proteoglycans (HSPGs) which enables them to assemble into lipoprotein particles. Release of these particles from the cell surface might be mediated by the phospholipase C-like protein Notum, which cleaves the glycosylphosphatidylinositol (GPI) anchors from the glypicans (a type of HSPGs) [53], acting as vehicles for long-range transport [54].

Lipoprotein-associated forms of HhN and ShhN have been observed in systemic circulation of fruit flies and humans [52]. It is unclear, however, whether lipoproteins are present in the extracellular space of early embry-

Besides the soluble multimers, lipoproteins and Disp/Scube proteins models some reports have suggested that the release and activity of modified HhN proteins is dependent on sheddase-mediated removal of N-terminal amino acids and palmitate [62–64] by ADAM proteins [65]. The authors propose a model where specifically ShhN multimers are produced as biological inactive forms with the zinc center mostly blocked by the N-terminal. Shedding of both N and C-terminals results in the mobilization and activation of ShhN multimers. Alternatively, HhN may be transported in extracellular vesicular particles called exovesicles [66, 67]. ShhN has also been observed in exovesicles derived from apical microvilli budding at the surface of the mouse ventral node during embryonic development.

So far, all models described above are based in the association of HhN either with itself or other proteins. It is worth mentioning that monomeric forms of HhN have been detected in the extracellular matrix (ECM) in presence or absence of N and C-terminal lipids [68], but how a doubly lipid-modified protein could remain soluble as a monomer has not been explained to date.

The apparent diversity in the release of HhN might reflect the several functions of this morphogen. Obviously, the number of potential interactants in the ECM is vast and any changes therein could influence the way that HhN propagates. How the HhN morphogen travels from the signaling cells to the target cells requires further investigation. Besides, additional studies are required to determine whether similar or distinct mechanisms are used in different tissues and whether this reflects the distance that HhN needs to travel or a specific tissue substrate through which HhN needs to act.

1.2.3 Role of dual lipidations in the amino-terminal domain

The connection among Hh signaling and lipid molecules attached to the HhN domain was recognized since a few years ago. However, the roles of both cholesterol and palmitic acid in the Hh pathway are still under debate. An important issue that currently is not totally clarified is whether various lipid modifications forms of HhN proteins possess different biological activities *in vivo*.

In general, the formation of the highest molecular-weight HhN complexes depends on both palmitoylation of the N-terminus and cholesterol modification of the C-terminus [69]. Consistent with this, the biochemical analysis of vertebrate ShhN expressed in tissue culture cells indicates that lipid modification promotes the formation of freely diffusible multimeric complexes [36, 46]. HhN proteins lacking these two lipid modifications result in a deficient secretion and spreading of the morphogen throughout the ECM [46, 70–73].

Cholesterol-mediated clustering may also promote interactions between HhN and other membrane-associated molecules such as HSPGs [69]. While the role of cholesterol in processing and trafficking of HhN proteins is well established [45, 74], a recent study suggests that it may not be absolutely required for activity [75].

Mutation of the N-terminal cysteine to serine (C24S in human ShhN) results in mutant forms that do not undergo palmitoylation and show reduced patterning activity in the mouse limb and neural tube [44, 76, 77].

1.2.4 Protein modulators of the signaling

Besides Ptc, many other proteins that bind HhN in the ECM or on the cell surface have been identified. Once released into the ECM, HhN may interact with a number of different proteins that modulate the signaling: the interference hedgehog protein family (Ihog in fly and its vertebrate homolog (Cdo) Cell adhesion molecule related/downregulated by oncogenes) [78], HSPGs [79], Growth-arrest-specific gene 1 (Gas1) [80, 81] and the hedgehog-interacting protein (Hhip) [82]. Ihog/Cdo and HSPGs act as crucial modulators in both vertebrates and invertebrates, while Gas1 and Hhip are vertebrate-specific. The extracellular regulation of signaling is more complex in vertebrates than in invertebrates. A summary of all proteins implicated in receiving or modulating Hh signaling is shown in Figure 1.4.

Ihog/Cdo-HhN: Ihog/Cdo proteins function as co-receptors with Ptc positively regulating Hh signaling [83–85]. Ihog encodes a protein with four extracellular Immunoglobulin (Ig) domains, two extracellular fibronectin type III (FNIII) domains, a transmembrane domain, and a C-terminal tail with no significant homology to other proteins. The *Drosophila* genome contains a related gene, named brother of Ihog (Boi). Ihog localizes to the cell surface, and both Ihog and Boi bind HhN through the first (Fn1) and second (Fn2) FNIII domain in a heparin-dependent manner [84, 86]. Nevertheless, the mode that Ihog binds HhN is not conserved across different phyla [87], suggesting that the Ihog-like binding mode was lost in mammals [88]. The closest vertebrate homologs of Ihog are Cdo and Brother of Cdo (Boc) and both bind vertebrate Hh proteins in the same manner [89]. Cdo is composed of five Ig domains followed by three FNIII repeats (an additional FNIII repeat relative to Ihog and Boi), a membrane-spanning region, and a cytoplasmic tail [90]. Sequence conservation clearly indicates that the second and third FNIII repeats of Cdo/Boc (Fn2 and Fn3) are homologous to the two Ihog/Boi FNIII repeats (Fn1 and Fn2, respectively) [91], but the third FNIII repeat (CdoFn3) appears to be most important for binding ShhN. The interaction between ShhN and CdoFn3 was revealed by McLellan *et al.* [87]. Such interaction requires calcium and is weakened at low pH [88].

Hhip-HhN: Hhip is a highly conserved, vertebrate-specific inhibitor of HhN signaling. It is a 700-residue protein that contains a predicted N-terminal Frizzled-like cysteine-rich domain (Fz) followed by a β -propeller region, two epidermal growth factor (EGF) repeats, and it is attached to the cell surface by a 22 amino acid C-terminal hydrophobic region. Hhip binds all three vertebrate Hh ligands (Shh, Ihh and Dhh) with nanomolar affinity [92]. In order to interact with Hh proteins Hhip adopts a lollipop-like shape with the β -propeller disc attached to an EGF repeat stem; where the binding site is localized at the periphery of the β -propeller domain, distant from the EGF stalk. In contrast to the Cdo–HhN complex, the Hhip–HhN interaction is centered on the zinc binding site, rather than the calcium binding site. More specifically, the zinc binding site of HhN proteins is occupied by a loop from the inhibitor, that is the most energetically important of the interacting loops [93, 94]. This interaction is mediated by D383 of Hhip, which explains the significant loss of binding affinity and decreased activity upon chelation of Zn^{2+} or mutation of D383. Examination of sequence similarities between Ptc and Hhip reveals

a common motif in the second ectodomain of Ptc which is similar to the loop in Hhip that contacts the zinc center. It has been proposed that Hhip inhibits Hh signaling by competing directly with Ptc for HhN binding. Cdo was also shown to compete with Hhip [87]. Although Hhip and Cdo have very different orientations relative to the Hh ligand, there is a considerable overlap of the binding surfaces on HhN that makes plausible the previous finding.

Gas1-HhN: Gas1 gene encodes a 45kDa GPI-anchored cell surface protein that is structurally related to the glial cell-derived neurotrophic factor family receptor (GFR) and binds ShhN with high affinity ($K_d \sim 6nM$). It was initially described as a binding factor that antagonizes HhN signaling when it is overexpressed [95]. However, the most recent work clearly implicates Gas1 as a positive regulator of the Hh signaling in multiple developmental contexts by acting cooperatively with Ptc and/or Cdo for HhN binding [81, 96]. These studies presented a model in which Gas1 helps to transform the Hh protein gradient into the observed activity gradient and suggested that Gas1 is an evolutionary novel vertebrate-specific Hh pathway regulator. In addition, McLellan *et al.* [87] suggest that an intact and accessible calcium binding site on ShhN is required to mediate interactions between ShhN and Gas1.

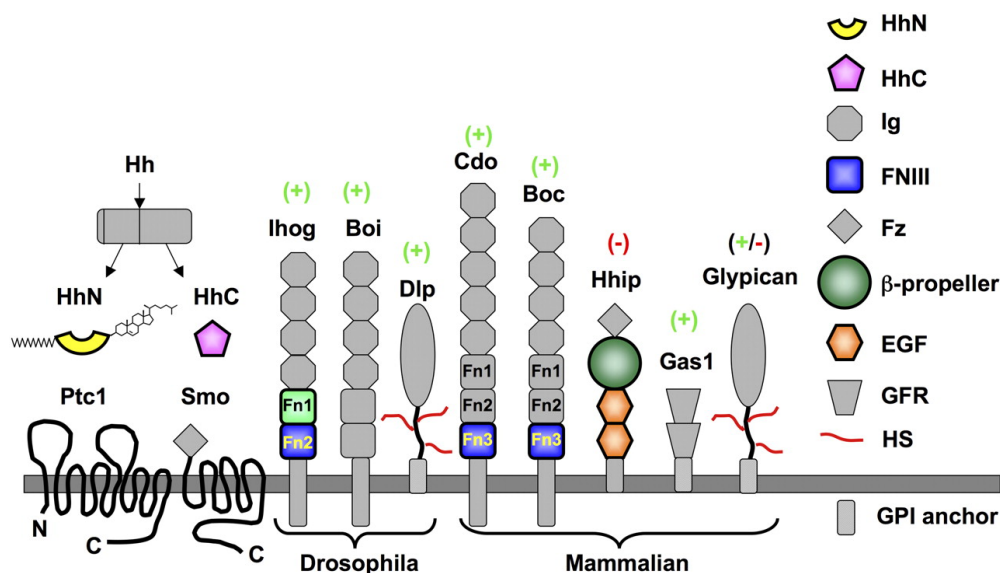


Figure 1.4: Hh binding partners which have been characterized to date. Multiple cell surface proteins have been implicated in receiving or modulating responses to Hh signals; proteins that are positive or negative regulators of Hh pathway activity are indicated by (+) or (-), respectively. In color are shown, domains for which high-resolution structures have been determined. DIP (Dally-like protein), note that all other acronyms are listed in the beginning of the document. Source: Beachy *et al.* [89].

HSPGs-HhN: Heparan sulfate polysaccharides are glycosaminoglycan (GAG) characterized by alternating D-glucuronic acid or L-iduronic acid and D-glucosamine units, which may be sulfated at different positions [97]. As a result, a simple octasaccharide can have over one million different sulfation sequences [98]. In nature, GAGs chains rarely occur as free entities but rather form covalent complexes with specific “core” proteins [99, 100].

The composite glycoproteins constitute the superfamily of HSPGs and they can be found on the cell surface and in the ECM [101, 102]. HSPGs are the major component of the ECM in mammals [103] and although heparin (its degree of sulfation is higher than HS) is commonly used to study interactions between signaling molecules and GAGs, it is worth mentioning that HS is the most likely physiological partner [104].

Studies in *Drosophila* and vertebrates have demonstrated the crucial roles of HSPGs in these signaling pathways during development [105, 106]. HSPGs are involved in the stabilization of HhN oligomers, and this stabilization is linked to the long-range spread of Hh [53]. Protecting HhN oligomers from degradation is likely to be required for the accumulation of HhN at the surface of producing cells, which is sufficient to yield an accurate gradient. If this threshold is not reached, the slope of the gradient will be affected, and the distance of target gene expression reduced. HSPGs also appear to be involved in the planar spread of HhN, as HSPGs keep HhN at the cell surface, preventing its dispersal into the extracellular space. Concentration of these signals near the cell surface might be especially important in developing epithelial tissues to prevent the release of ligand from the epithelial surface into the lumen [107]. Alternatively, HSPGs may participate in releasing Hh ligand from producing cells (see subsection 1.2.2). The effect of HSPGs on Hh signaling may be positive or negative and may affect either responsiveness to Hh or the tissue distribution of Hh [108].

HhN proteins interact with HSPGs through a cluster of positively charged residues within the N-terminal region [109] that allows for electrostatic interactions between the positive charges on the protein and the negatively charged sulfates of HSPGs. This cluster of residues composes the called Cardin-Weintraub motif (CW) proposed as BBBXBB, where B is a lysine or arginine (with a very rare occurrence of histidine) and X is a hydrophobic residue [110]. In the case of human ShhN, this region consists of residues K32-K38. Mutations of amino acids within this highly conserved motif eliminates the high affinity interaction of HhN with various HSPGs but do not abolish completely the binding [46, 48, 109] suggesting that other motifs may mediate the interaction between Hh and HSPGs in addition to this Cardin-Weintraub sequence. A recent study performed by Whalen *et al.* [111] reveals another GAG binding common to all HhN proteins that comprises residues K87, R123, R153, R155 and K178 (human ShhN nomenclature). These results suggest a two-site model for HhN-GAGs interactions. In addition, solution competition studies on ShhN showed that the minimum size of a heparin oligosaccharide capable to interact with ShhN is larger than an octasaccharide [112].

1.2.5 Dysregulation of the pathway

The role of Hh signaling in various developing tissues explains why genetic defects in the pathway lead to several congenital abnormalities [42]. For instance, the aberrant activation of the pathway can have important roles in the initiation and growth of cancer [113, 114] while decreased activity can result in developmental malformations. It has been estimated that 25% of all human tumors require Hh signaling to maintain tumor cell viability [115]. Aberrations in the Gli protein family can cause the Pallister–Hall syndrome. This complex genetic disease involves polydactyly (additional fingers or toes) and other abnormalities such as a short nose with flat nasal bridge, and cleft palate [116, 117].

Genetic mutations represent the major source of Hh signaling abnormalities causing congenital dimorphism in humans. Missense mutations in the *Shh* gene have been associated with *Holoprosencephaly* (HPE) while mutations in *Ihh* with *Brachydactyly type A-1* (BDA1). These mutations can be classified as alterations in the amino-terminal domain, which affect either the secretion or activity of the signaling domain.

Holoprosencephaly (OMIM entry: 236100): Is an etiologically heterogenous condition with teratogenic and genetic factors, it occurs with a frequency of 1 in 16,000 live births and 1 in every 250 during early embryogenesis [118, 119]. HPE is caused by impaired midline cleavage of the embryonic forebrain and is the most common defect of the forebrain and midface in human [120, 121]. The most severe form is cyclopia, and the mildest phenotype is a single upper central incisor. However, the disease is highly heterogeneous and can also be associated with abnormalities such as microcephaly, mental retardation and epilepsy [122].

Shh was the first gene identified to cause HPE in mice and humans from nonsense mutations or deletions that result in loss of function [123, 124]. Mutations on *Shh* chromosome 7q36 account for a significant proportion of autosomal dominant HPE [125]. In this regard, various missense mutations in human *Shh* (G132R, D88V, N115K, W117G, W117R) have been reported and linked to HPE [126, 127].

Brachydactyly type A-1 (OMIM entry: 112500): The term *brachydactyly* is derived from the ancient Greek and means "short digits". It forms part of the group of limb malformations characterized by bone dysostosis and can occur either as an isolated malformation or as a part of a complex malformation syndrome [128]. To date, many different forms of brachydactyly have been identified e.g., A1, A2, A3, A4, B, C, D [129]. Particularly, BDA1 was the first disorder to be described as an autosomal dominant Mendelian trait [130] in humans and in 1951 was classified by Bell [131] as type A1.

In BDA1, middle phalanges of all digits are variably short or rudimentary and occasionally, are fused with terminal phalanges (Figure 1.5). The proximal phalanges of the thumbs and big toes are short; affected individuals tend to be of shorter stature in adulthood [133]. BDA1 can be caused by mutations in the *Ihh* gene located on different loci that affect the capacity and range of the signaling; this is consistent with the fact that *Ihh* is involved in the development of the endochondral skeleton of vertebrates [134]. The proximity of all reported mutations in human *Ihh*: E95K, E95G, D100E, D100N, R128Q, T130N, E131K, T154I and R158C [132, 135–140] suggests that they may reside in a region of particular importance for the normal functioning of the growth plate during development. Multiple mutations in codon 95 and 100 indicate that these codons may be mutational hot spots [133].

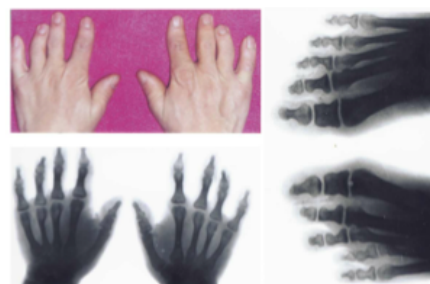


Figure 1.5: BDA type A1 phenotype of an affected individual. Middle phalanges of all digits are variably short or rudimentary and are occasionally fused with terminal phalanges. The proximal phalanges of the thumbs and big toes are short. Figure adapted from Ref. [132].

1.3 Distinctive structural features of Hedgehog proteins

The first crystal structure of the N-terminal domain of murine ShhN (without lipid modifications) was determined in 1995 [141]. It shows a relatively globular domain of $\alpha+\beta$ sandwich core composed of two α -helices and six β -strands decorated by extensive loop regions. The presence of a zinc ion tetrahedrally coordinated bound at the bottom of a solvent-accessible cleft was also shown. The arrangement of Zn^{2+} is very similar to that of zinc hydrolases such as thermolysin and carboxypeptidase A [142]. However, about 13 years later, McLellan *et al.* revealed (in addition to the zinc center) the presence of a double calcium binding site buried at the interface with CDO [87]. The two metal centers are separated by a distance of about 1.1 nm and connected by a long loop between two antiparallel β -sheets. Although in the beginning, the zinc center was found to be a potential catalytic site no enzymatic activity has been uncovered so far. Besides, the role of calcium ions is not fully understood yet. A list of all X-ray structures available in the Protein Data Bank (PDB) [143] is shown in Appendix B.

In this section, a detailed description of the peculiarities of both zinc and calcium binding sites is given.

1.3.1 Zinc binding site

Zinc unlike other first row transition metals does not participate in redox reactions which makes it a stable ion in a biological medium. In proteins, zinc can play a catalytic or structural role. In catalytic sites, Zn^{2+} acts as a Lewis acid and is found bound usually to water and preferentially to His, followed by Glu/Asp and then Cys. In structural sites Zn^{2+} can stabilize or induce the folding of proteins and is bound preferentially to Cys rather than to His [144].

As is shown in Figure 1.7A, Zn^{2+} in murine ShhN is coordinated by H141, D148, and H183 with a water molecule acting as the fourth ligand. Based on the previous comments, the zinc arrangement is more similar to catalytic sites rather than structural ones. The question of whether Zn^{2+} played a pivotal catalytic role in the Hh signaling pathway was addressed by different authors [145, 146] about two decades ago. Mutagenesis studies ruled out this possibility proving that Hh vertebrate homologs act as a ligand for Ptc containing a pseudo-active site and not as a proteolytic enzyme with respect to Hh signaling. Moreover, loss of the zinc decreases Hh stability.

Despite no enzymatic activity has been ascribed to date to Hh morphogen, the zinc arrangement is known to occur in other zinc proteases. All proteins that share such a zinc arrangement have been grouped into a new family called LAS (from Lysostaphin-type enzymes, D-Ala-D-Ala metallopeptidases and Sonic hedgehog) [147]. The LAS group comprises six structurally characterized protein families with no significant overall sequence similarity [148–153]. The three amino acid ligands of the zinc ion occur in the order His, Asp and His having a large gap in sequence between the second and third Zn^{2+} ligand.

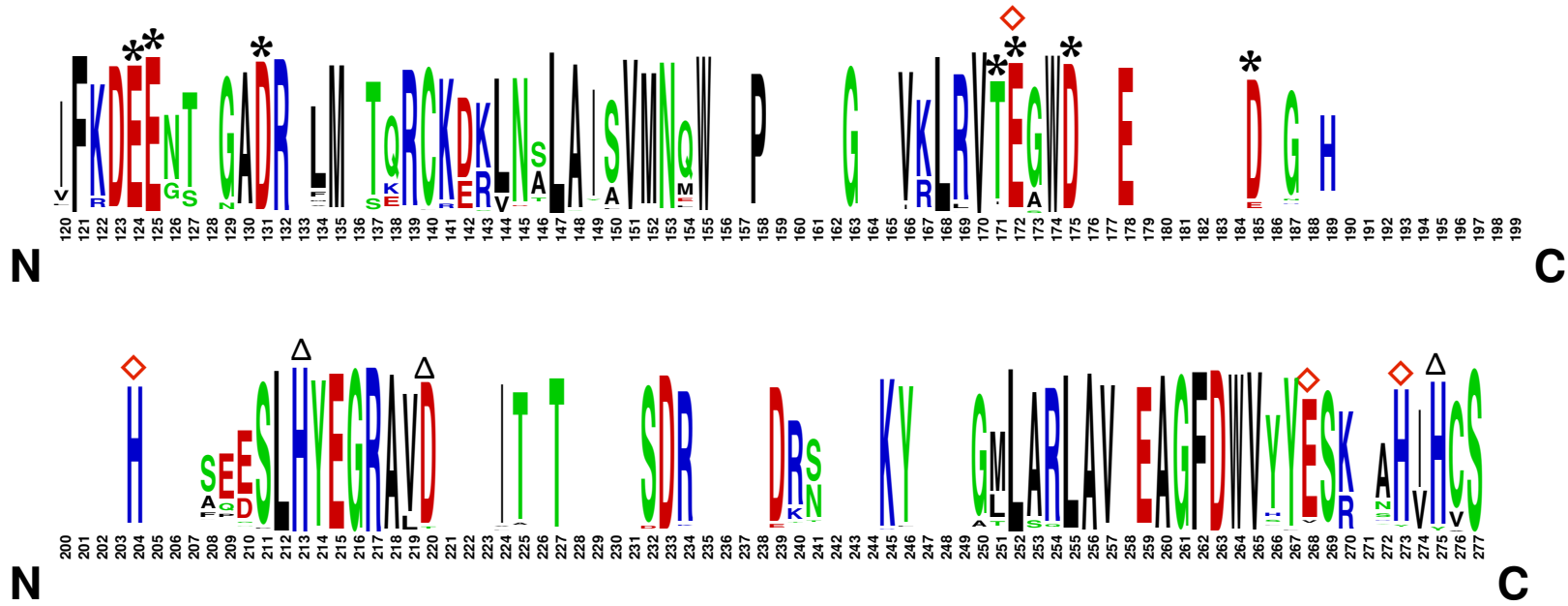


Figure 1.6: Zinc and calcium binding sites are highly conserved along the Hh family. Residue conservation of HhN is shown after the alignment of 383 sequences (Pfam ID:PF01085). Calcium and zinc ligands are highlighted by black stars and triangles respectively. Residues proposed to be involved in the putative catalytic mechanism of Hh proteins are marked as red diamonds, note that residue E127 coordinated the calcium ions and might act in the catalysis too. Amino acids are colored according to their chemical properties: polar (green), basic (blue), acidic (red) and hydrophobic (black). Sequence logo was generated with WebLogo [154].

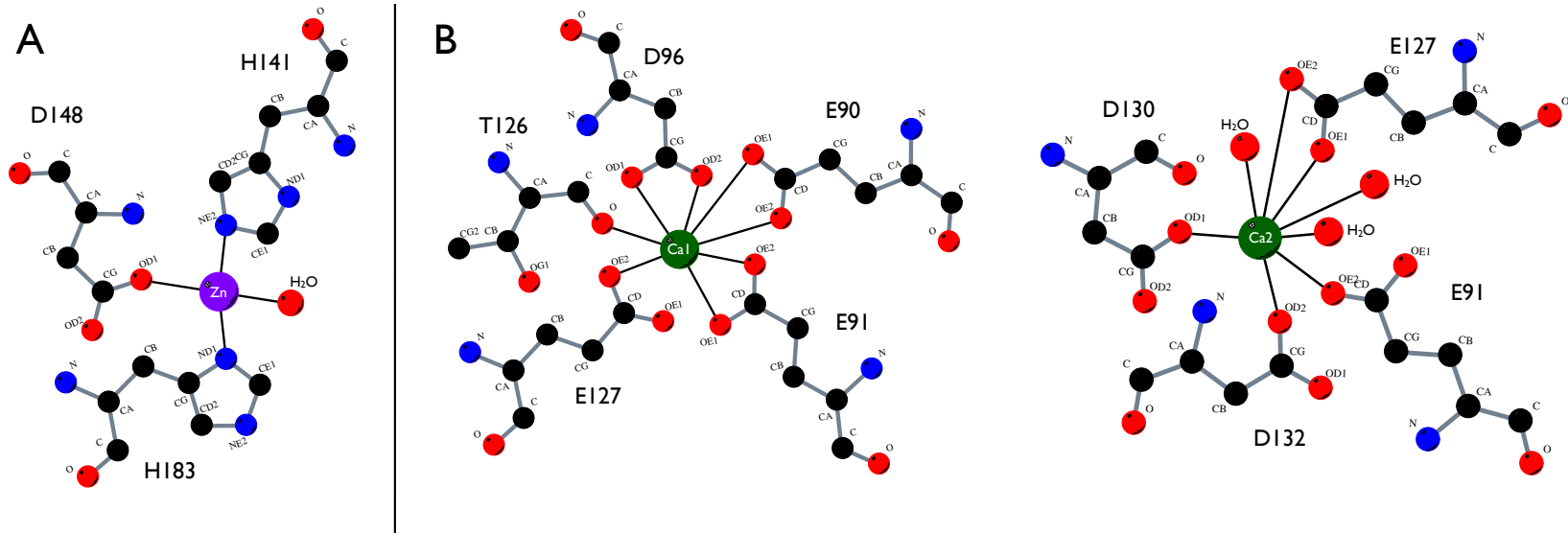


Figure 1.7: Schematic 2-D representation of the zinc (A) and the double calcium (B) binding sites in murine ShhN. The illustration is based on the X-ray structure, 3d1m. Shown are residues that directly coordinate the metals, as well as, water molecules. Oxygen atoms are in red, nitrogen atoms in blue, and carbon atoms in grey. LigPlot⁺ program [155, 156] was used to generate the plot.

The structural similarity between the Hh zinc binding site and the active site of zinc hydrolases strongly suggests that these classes of proteins are evolutionarily related [94]. Nevertheless, the fact that this site is conserved in all known Hh proteins (see Figure 1.6) except in *Drosophila* (in which two of the three zinc ligands are not conserved) have caused a great controversy. A plausible explanation regarding the Hh function might be that the presence of this zinc binding site in vertebrates can introduce an additional level of regulation in the signaling pathway, e.g., the interaction between Hhip inhibitor and Hh that is missing in *Drosophila* (subsection 1.2.4).

1.3.2 Calcium binding site

The double calcium binding site from murine ShhN protein is shown in Figure 1.7B. Coordination of both calcium ions is dominated by side chain carboxylate oxygens from Asp and Glu residues (E90, E91, D96, E127, D130, D132) but main chain carbonyl group of Thr (T126) and water oxygens are also present. The level of conservation of the seven calcium coordinating residues in Hh sequences suggests that calcium binding is a common property of Hh proteins (Figure 1.6).

Ca1 (left side in Figure 1.7B) exhibits an almost perfect pentagonal bipyramid geometry formed by eight ligands all of which are protein oxygen atoms. Moreover, E91 and E96 occupy positions as bidentate ligands with Ca-O-C angles of about 90° . Structural similarities between Ca1 binding site and the canonical EF-hand motif can be appreciated if the composition of the first one is analyzed according to the description of calcium binding proteins provided by Kirberger *et al.* [157]. Some criteria as ligand preference, geometry, coordination number and angles are in agreement. Nevertheless, unlike the EF-hand motif which is a continuous binding site, the residue ligands in Ca1 site are distant from each other within the sequence. In contrast to Ca1, in Ca2 (right side in Figure 1.7B) three water oxygen ligands coordinate the metal ion, and is more exposed to the solvent. The distance between the two calcium ions is about 0.4 nm.

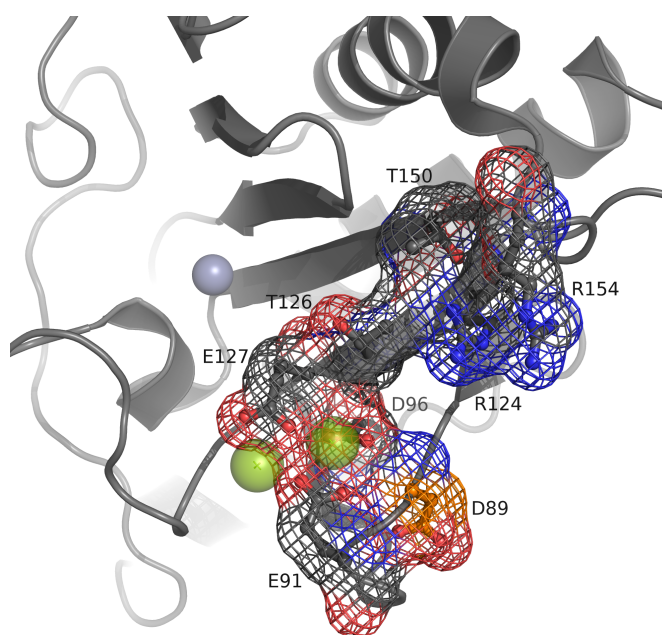


Figure 1.8: Proximity of residues accounting for BDA1 and HPE diseases. In spite of, not all residues contact directly the calcium ions they reside in a region nearby. Amino acids carrying mutations that have been reported to cause BDA1 are in gray color (carbon atoms) while other causing HPE are depicted in orange. Oxygen and nitrogen atoms are colored in red and blue respectively. Residues are labeled according to mouse ShhN nomenclature and shown in stick and mesh representations. Metal ions are visible as zinc (slate) and calciums (green) spheres.

Despite the weak affinity ($K_d \geq 100 \mu\text{M}$) [87] of the calcium binding site, several studies have shown that Ca^{2+} is important for optimal interactions between HhN and different binding partners [87, 88, 93, 94, 158].

The role of Hh signaling in various developing tissues explains why genetic defects in the pathway lead to several severe congenital abnormalities. Interestingly, most of the missense mutations that cause HPE and BDA1 map to this calcium binding site (see Figure 1.8). The introduction of D88V (D89V in mouse) mutation into ShhN abolishes binding to Cdo and decreases binding to Hhip, Gas1, and Ptc [122]. Some BDA1 mutations in residues that directly coordinated the Ca^{2+} in Ihh or their equivalents in Shh also eliminates binding to Cdo and decreases binding to Hhip and Ptc [87]. Taking into account the multiple partners in the Hh signaling and their role as positive or negative regulators, the phenotypic effects caused by HPE and BDA1 mutations can be a complex mix of gain and loss of function.

1.4 General remarks on the Hedgehog signaling

The conserved but unusual way in which Hh ligands are generated and released from cells; the differences observed in the binding modes of vertebrate and invertebrate Hh proteins and their homologous coreceptors, Cdo and Ihog; the discovery in the beginning of a zinc site in Hh proteins with striking similarity to zinc hydrolases implying a hydrolytic activity as a Hh function that was discounted later on by mutagenesis studies; the surprising identification of a conserved double-calcium center just 1.1 nm away from the zinc site; the regulation of Smo by Ptc and the existence of multiple cofactors capable of modulating responses to Hh are just a few of many issues to keep in mind when about Hh signaling is argued.

Addressing these issues have represented a challenge for the scientific community (in terms of knowledge) in the last twenty years. Despite the advances, however, there are still large gaps in our comprehension of the Hh signaling and many unsolved questions. Given the number and complexity of Hh pathway components, and the intertwined processes of modulating Hh distribution and responsiveness, so far is difficult to develop a complete molecular view of the nature of interactions between Hh proteins and their partners.

Understanding this important signaling pathway is crucial for developing new therapeutic strategies to target the medical consequences of its dysregulation and for characterize new interactions. Clearly, there is a long way to go.

1.5 Motivation and aims of the current research

The development and application of new experimental techniques in the last years make possible that genetic data and biochemical studies highlighted the importance of both, zinc and calcium metal binding sites on Hh family proteins (ShhN vertebrate homolog is

one of the most studied). Nevertheless, so far the function of these metal ions is not fully understood and is widely debated. For instance, the following basic questions remain unclear: *What role do calcium ions play in maintaining the ShhN structure?, Is the zinc ion only implicated in the stability of ShhN despite this metal center resembles a catalytic site?, It would be possible that ShhN is a Ca^{2+} activated LAS peptidase?.*

The importance of addressing these questions might be relevant to shed light on the rules governing Hh signaling pathway. Therefore, the goal of the current work was to derive theoretical insights into these questions, giving guidance for the development of new experimental studies. To accomplish such a task, different computational tools, in particular, molecular dynamics approaches were applied.

“After all, when **you** are in **love**, you want to tell the **world**. The **idea** that **scientists** should **not** talk about their science to the public **seems** to me **bizarre.**”

– Carl Sagan –

CHAPTER 2

APPLICATION OF MOLECULAR DYNAMICS METHODS IN THE STUDY OF BIOMOLECULES

2.1 Brief overview of computational methods

One of the major biological and biophysical interests is to investigate and understand the basis of protein structure and function whereas protein function is intimately linked to its dynamic nature. The use of theoretical methods or computational approaches has become a valuable tool to provide accurate answers in this regard, often more easily than experiments [159]. Computational approaches to calculate potential energy may be divided into two broad categories:

Quantum Mechanics (QM) approaches [160] are a very precise way to get a physical description of your system. QM explicitly represents the electrons in a calculation and thus is specially useful to investigate chemical reactions in which bonds are formed and broken or any molecular system where electronic effects are dominant. Nowadays, they are considered as the highest level of theory that can be achieved. Nevertheless, here we face a problem: the size of the biomolecules (e.g., proteins). Typical biomolecular systems are often too large to be treated with QM. Even when some electrons could be ignored still a large number of particles must be considered and therefore the calculation is time-consuming. The next question arises then: How do we handle, from a theoretical point of view, a system containing a significant number of atoms, such as proteins? The answer brings us to the second category.

Molecular Mechanics (MM) approaches [161] have been successfully applied to model large biomolecules. In general terms, the MM model consists of spherical atoms connected by springs which represent bonds (Figure 2.1). It uses simple potential-energy functions (e.g., harmonic oscillator or Coulombic potentials) which are described at the nuclear level by the Born-Oppenheimer approximation. Nuclei and electrons are lumped together and treated as unified atom-like, in other words; the electron motion is averaged out

during nuclear motion, and thus electrons are not explicitly represented.

The object of MM is to predict the energy associated with a given conformation of a molecule. However, MM energies have no meaning as absolute quantities. Only differences in energy between two or more conformations have meaning. The parameters that are used in the energy calculations usually are derived from databases of structures that have been developed by experimental and/or QM methods. MM approaches can be applied to different scenarios, such as: molecular structure refinement, molecular dynamics, Monte Carlo simulations and ligand-docking simulations. The compromise among speed, accuracy and precision makes MM approach computationally feasible for routine usage.



Figure 2.1: Balls and spring model. In MM methods the molecules are represented as systems of solid spheres connected via springs.

In addition, to study the electronic behavior in biomolecules, QM and MM are combined into one calculation (Hybrid QM/MM methods) [162–164] that models a large molecule (e.g., enzyme) using MM and one crucial section of the molecule (e.g., active site) with QM. This calculation is designed to be efficient in terms of accuracy and speed since only a certain region needs to be modeled quantum mechanically.

In the present study, MM approaches and, in particular, molecular dynamics (MD) methods [165] have been used. In systems where the motions and their development with time are the primary interest, MD can provide the necessary information. To achieve such a task GROMACS package (<http://www.gromacs.org>) version 4.5 [166] and BrownDye software (<http://browndye.ucsd.edu>) [167] were utilized. In the next section, concepts upon which MD methods are justified and implemented are discussed.

2.2 Molecular dynamics theory

A MD simulation is based on Newton's equations of motion, allowing structural fluctuations to be observed with respect to time. For a system of N interacting atoms:

$$F_i = m_i a_i, \quad i = 1 \dots N \quad (2.1)$$

where F_i is the force exerted on particle i , m_i and a_i are the mass and the acceleration of particle i respectively. The forces can also be expressed as the negative derivatives of a potential function V with respect to the position of particle i (r_i):

$$-\frac{dV}{dr_i} = F_i \quad (2.2)$$

Combining these two equations yields:

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (2.3)$$

The evolution of the simulated structure with respect to time is called trajectory and it is calculated in small time steps through the integration of the equation 2.3. In practice, due to the lack of analytical solutions, the trajectories are not directly obtained from Newton's equations. They must be solved numerically by different integration methods. GROMACS utilizes by default the *leap-frog* algorithm [168] for this purpose (Figure 2.2). First the velocities v are calculated at time $t - \frac{1}{2}\Delta T$ and these velocities are used to calculate the positions r at time t . Positions and velocities are updated using the forces $F(t)$ determined by the positions at time t . Steps can be summarized as follows:

$$\begin{aligned} v\left(t + \frac{1}{2}\Delta t\right) &= v\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}F(t) \\ r(t + \Delta t) &= r(t) + \Delta t v\left(t + \frac{1}{2}\Delta t\right) \end{aligned} \quad (2.4)$$

The advantages of this algorithm is that the velocities are explicitly calculated and it does not require the calculation of the differences of large numbers. However, the disadvantage is that velocities are not calculated at the same time as the positions. The equations of motion are deterministic, which means that the positions and the velocities at time zero determine the positions and velocities at all other times.

Once the trajectory is ready, the average values of many macroscopic properties can be extracted. It is important to note that MD simulations generate information at the microscopic level. The conversion of this microscopic information to macroscopic observables (e.g., pressure, energy, heat capacities) is made via statistical mechanics. It provides the rigorous mathematical expressions that relate the distribution and motion of the atoms and molecules to macroscopic properties.

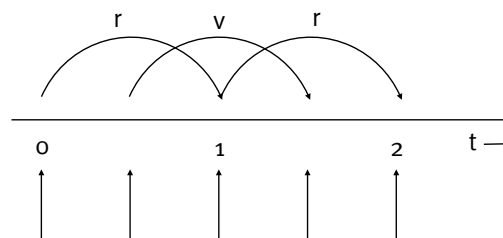


Figure 2.2: The leap-frog integration method. In this way, the velocities (v) *leap* over the positions (r), then the positions *leap* over the velocities.

2.2.1 Empirical force field models

In the context of molecular modeling, the potential function form V and the parameters used to describe intermolecular and intramolecular interactions are termed *empirical force field* (FF). The term *empirical* refers to the number of substantial approximations that are applied. Empirical FFs are the main tool of MD simulations. Most FFs are similar in the types of terms they contain, however, the direct transfer of parameters from one FF to another is generally not valid. The main differences between the various FFs result from the diverse approach used to derive the individual parameters [169]. One important feature is the transferability of a FF. It enables to transfer parameters developed from data on small molecules to a series of related molecules or much larger molecules rather than,

having to define a new set of parameters for each individual molecule [170]. Examples of FFs developed primarily for MD of macromolecules are: AMBER [171], CHARMM [172], OPLS-AA [173, 174] and GROMOS [175]. In this study, the GROMOS-96 FF [176, 177] has been used to perform the MD simulations. This is a united atom FF, i.e. without explicit representation of non-polar hydrogens.

The potential energy function takes the form:

$$\begin{aligned}
 V(r) = & \sum_{bonds} \frac{1}{4} K_b (b^2 - b_0^2)^2 \\
 & + \sum_{angles} \frac{1}{2} K_\theta (\cos \theta - \cos \theta_0)^2 \\
 & + \sum_{har} \frac{1}{2} K_\xi (\xi - \xi_0)^2 \\
 & + \sum_{torsions} K_\phi (1 + \cos(\delta) \cos(m\phi)) \\
 & + \sum_{nonb} \left[\frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^6}{r_{ij}^6} \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \quad (2.5)
 \end{aligned}$$

where \mathbf{r} is the (3N-dimensional) position vector of the N atoms of the system. The first four summations stand for bonded interactions based on a fixed list of atoms. They are not exclusively pair interactions, but include 3 and 4-body interactions as well. A brief description is given below.

The first summation models all covalent bonds in the molecular system not treated as constraints where K_b is the force constant, and b and b_0 are the instantaneous and ideal lengths of bond n , respectively. Although the form is not harmonic, the force constants are mostly derived from the harmonic ones used in GROMACS-87 [178]. The second summation is a simplified function used to calculate bond angles in the molecular system where K_θ is the force constant, and θ and θ_0 are the instantaneous and ideal values of the bond angle n , respectively. The third summation stands for a special type of dihedral interaction called improper dihedral. It is used to force atoms to remain in a plane or to prevent transition to a configuration of opposite chirality (a mirror image). In equation 2.5 K_ξ is the force constant, and ξ and ξ_0 are the instantaneous and ideal values of the improper dihedral angle n , respectively. Improper dihedrals are commonly used in united atom FFs and the simplest improper dihedral potential is a harmonic potential. Since the potential is harmonic it is discontinuous but, since the discontinuity is chosen at 180° distance from ξ_0 this will never cause problems. The fourth summation runs over a set of torsional dihedral angles (for alkanes all) and its modeled by a simple periodic function. K_ϕ is the force constant, δ is the phase shift, and m is the multiplicity.

Pairwise interactions are calculated in the final sum representing the non-bonded term. The nonbonded interactions are computed on the basis of a neighbor list (a list of non-bonded atoms within a certain radius), in which exclusions are already removed. In principle, it models all atoms pairs, excluding atom pairs separated by one or two covalent bonds and atom pairs separated by a distance beyond a given cut-off. The non-bonded

energy accounts for van der Waals interactions (first term) and electrostatic interactions (second term). Van der Waals interactions are modeled using the 6-12-Lennard-Jones potential, where the repulsion and dispersion term are combined. The parameters C_{ij}^{12} and C_{ij}^6 determine the van der Waals interaction between atoms i and j . The electrostatic contribution is modeled using a Coulombic potential depending on partial atomic charges (q_i and q_j are the partial charges of atoms i and j). ϵ_0 is the dielectric permittivity of the vacuum, ϵ_1 is the relative dielectric permittivity within the sphere of a cutoff radius centered in atom i . In general, van der Waals interactions occur at short-range while electrostatic interactions dominate at long-range.

A detailed description of the parameterization aspects of the GROMOS-96 FF can be found in Ref [177].

2.2.2 Optimized Lennard-Jones parameters for Ca^{2+} interactions

As was mentioned in subsection 2.2.1, FFs are based on numerous approximations and derived from different types of theoretical and experimental data. In classical MM, FF parameters for ions and proteins are usually constructed to reproduce thermodynamic properties such as enthalpies or solvation rather than kinetic properties. Often, for specific ion-ligand interaction, FFs use heuristic combination rules which may lead to inaccurate values [179].

Project and collaborators [180], have demonstrated by MD simulations of systems containing Ca^{2+} ions and proteins, that inconsistencies in different FFs are due to improper parameterization of the ion-carboxylate interactions. Although, a detailed description of the referenced work is beyond the scope of this study, we note that they have tested and further ruled out the possibility that solvent- Ca^{2+} interactions or protein-solvent interactions were responsible for the significant differences observed among the FFs. Hence, they were focused on selecting better parameters for the direct reaction between the calcium ion and the carboxylate oxygens. For GROMOS-96 FF the set ($C_6 = 2.0 \cdot 10^{-3} \text{ kJ/mol nm}^6$ and $C_{12} = 1.67 \cdot 10^{-6} \text{ kJ/mol nm}^{12}$) was recommended. The new set of Lennard-Jones parameters successfully reproduces experimental pK_a values without affecting any other aspect of the simulation. Therefore, in the present study this set instead of the sets obtained from the standard combination rules has been used to model the Ca^{2+} -protein interactions.

2.2.3 Periodic boundary conditions

MD is typically applied to systems containing thousands of atoms. The classical manner to minimize edge effects in a finite system is to apply *periodic boundary conditions* (PBC). In this way, it ensure that all simulated atoms are surrounded by neighboring atoms, whether those neighbors are images or not.

The system is assumed to be composed of the central cell surrounded by its exact replicas to model a macroscopic sample. The image cells not only have the same size and shape as the central cell but also contain atoms that are simply images of the atoms simulated in that cell (see Figure 2.3). Atoms can freely enter or leave any cell, thus, there are no

boundaries of the system. When atoms leave the cell, their images simultaneously enter the cell through the opposite face. Therefore the shape of the cells must be space-filling. The most common shape used to replicate the central cell is a cubic lattice. However, PBC are not restricted to cubic systems, other geometries are available such as: the rhombic dodecahedron and truncated octahedron. The former is the smallest space-filling representation of a unit cell. Each of the 12 image cells is at the same distance. The volume is 71% of the volume of a cube having the same image distance. This saves about 29% of CPU-time when the user wants to simulate a spherical or flexible molecule in solvent [181].

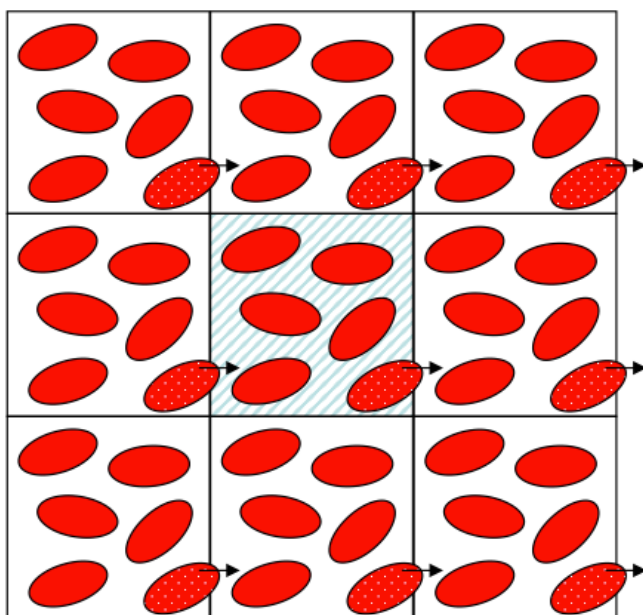


Figure 2.3: Periodic boundary conditions in two dimensions. The primary cell is surrounded by its image cells. Molecules that leave the cell will be replaced by their images entering the cell from the opposite side.

The dimensions of the box should be large enough to prevent that molecules in one box do not interact with their own images in the neighboring box. Ideally, the distance from the macromolecule periphery to the box edge should be no less than 1.0 nm for most systems.

The number of image cells needed in the system depends on the range of intermolecular forces. GROMACS uses PBC, combined with the *minimum image convention*: only one -the nearest- image of each atom is considered for short-range non-bonded interaction terms. For long-range electrostatic interactions this is not always accurate enough, and GROMACS therefore also incorporates lattice summation methods such as Ewald sum-

mation [182] and Particle-mesh Ewald (PME) [183, 184].

Electrostatics: Calculation of long-range electrostatic interactions is the most time consuming part of MD simulations. As the number of charges in a system increases, the number of Coulombic interactions will grow as the square of that number. Thus, it will result in a large number of interactions to evaluate. The application of truncation methods is not recommended as this can lead to unstable systems where the Coulombic forces acting on the atoms can be over or underestimated.

The PME method is one of the approaches to consider all electrostatic interactions over the complete system in a computationally tractable manner and is substantially faster than ordinary Ewald summation on medium to large systems. The algorithm employs interpolation of reciprocal space Ewald sums from a grid and computation of convolu-

tions using fast Fourier transforms. It gives more reliable energy estimates especially for systems where counterions like Na^+ , Cl^- , Ca^{2+} , etc. are used. This method offers a theoretically rigorous approach to the evaluation of electrostatic interactions in infinite periodic systems.

2.2.4 Explicit solvent model

In natural conditions, most proteins are immersed in aqueous solvent. Therefore, aqueous solvent is commonly selected as the environment for the vast majority of simulations. Due to the importance of solvent effects in the structure, stability, dynamics, and function of biomolecules, a proper representation of the water molecules should be considered. Explicit solvent models, which treat the solvent in atomic detail, are widely used in MD simulations. The simplest model considers a water molecule as a rigid body and rely on the non-bonded interactions.

Usually, FFs have been parametrized with one specific water model and it is recommend not to use any other. For all GROMOS FFs this water model is SPC [185]. Nevertheless, studies have revealed that the SPC/E water model [186] performs best in water box simulations [187] and it is better for use with the long-range electrostatics methods.

As can be seen in Figure 2.4, the SPC/E is a rigid three center point charge model with OH distance of 0.1 nm and H-O-H angle equal to 109.47° (the model assumes an ideal tetrahedral shape instead of the known geometry of the water molecule). Charges on the oxygen and hydrogen equal to -0.8476 and +0.4238 e, respectively and with Lennard-Jones parameters of oxygen-oxygen interaction according to equation 2.6: $A=0.37122$ (kJ/mol) $^{1/6} \cdot nm$ and $B=0.3428$ (kJ/mol) $^{1/12} \cdot nm$.

$$V_{LJ} = - \left(\frac{A}{r} \right)^6 + \left(\frac{B}{r} \right)^{12} \quad (2.6)$$

where r is the distance between two atoms.

The SPC/E model adds an average polarization correction term to the potential energy function:

$$E_{pol} = \frac{1}{2} \sum_i \frac{(\mu - \mu^0)^2}{\alpha_i} \quad (2.7)$$

where μ is the dipole moment of the effectively polarized water molecule (2.35 D), μ^0 is the dipole moment of an isolated molecule and α_i is an isotropic polarizability constant. The correction results in adding 5.22 kJ/mol to the total electrical energy. The SPC/E model results in a better density and diffusion constant than the SPC model.

2.2.5 Implicit solvent model

While explicit solvent models offer some of the highest levels of detail, they generally require extensive sampling to converge properties of interest. In practical situations, the evaluation of the solvent-solvent interactions consumes a big part of the computational

calculations, even though the primary goal is the solute behavior. A more common approach is therefore to use an implicit (continuum) model which generally replace the explicit solvent with a dielectric continuum. The continuum representation of solvent also significantly improves the computational speed and reduces errors in statistical averaging that arise from an incomplete sampling of solvent conformations. These type of solvent models have become popular for a variety of biomedical research problems, e.g., the estimation of free energies of solvation [188].

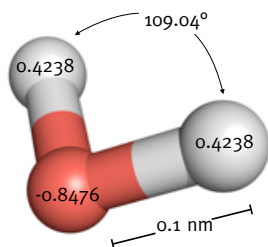


Figure 2.4: The SPC/E explicit water model.

This is a three-site model very popular because of its simplicity and computational efficiency. Geometry and parameters are shown.

Generally, these models are divided in two categories: (1) polar solvations in which Poisson–Boltzmann approaches [189] and Generalized Born models [190] are often used to approximate polar solute–solvent interactions by representing the solvent with a simple dielectric continuum model and, (2) nonpolar contributions [191] as attractive van der Waals interactions with the solvent. In these models, the nonpolar component of the solvation energy is derived from the cost to form a cavity in the solvent large enough to accommodate the solute and, from the hydrophobic effects associated with solvation of nonpolar solutes. Next, one of the most popular continuum model for describing electrostatic interactions between molecular solutes in nonhomogeneous media is described.

Poisson-Boltzmann model: This model is based on the nonlinear Poisson-Boltzmann equation. For systems where ions and water are both included in the solvent, it takes the form:

$$\vec{\nabla} \cdot [\epsilon(\vec{r}) \vec{\nabla} \psi(\vec{r})] = -4\pi \rho^f(\vec{r}) - 4\pi \sum_i c_i^\infty z_i q \lambda(\vec{r}) e^{-\frac{z_i q \psi(\vec{r})}{kT}} \quad (2.8)$$

where $\epsilon(\vec{r})$ represents the position dependent dielectric, $\psi(\vec{r})$ represents the electrostatic potential, $\rho^f(\vec{r})$ is the charge density of the solute, c_i^∞ stands for the concentration of the ion i at a distance of infinity from the solute, z_i is the valence of the ion, q is the charge of a proton, k is the Boltzmann constant, T is the temperature, and $\lambda(\vec{r})$ describes the accessibility to ions at points \vec{r} (often set to uniformly 1). In the Poisson–Boltzmann approach all solute atoms are often considered explicitly as particles with low dielectric constant with point partial charges at atomic positions. Typically, the dielectric constant of the solute (protein or a nucleic acid) is in the range 2–4 while the solvent surrounding the solute has a high dielectric constant (about 80) [192].

A number of numerical Poisson-Boltzmann equation solvers of varying generality and efficiency have been developed [193, 194]. In this study, the APBS software package [195], designed to efficiently evaluate electrostatic interactions of large biomolecules systems has been used.

2.2.6 Replication of experimental conditions

The incorporation of values that represent physical conditions such as pressure and temperature into MD, allows that more realistic and biologically relevant simulations can be performed. The systems can be described by thermodynamic ensembles that depend on a few observable parameters, and which are in statistical equilibrium. The term *ensemble* stands for a collection of points in phase space satisfying the conditions of a particular thermodynamic state. The most widely simulated ensembles are: the canonical (NVT) and isothermal-isobaric (NPT) ensembles. In the NVT ensemble, the number of particles (N), system volume (V) and temperature (T) are constant. Temperature is regulated via a thermostat; in terms of an average of the kinetic energy. In the NPT ensemble, the number of particles (N), system pressure (P) and temperature (T) are fixed while the volume is a variable. Most experimental conditions are of this type, hence the NPT ensemble is commonly used for biomolecules.

Temperature coupling: GROMACS can use either the weak-coupling scheme of Berendsen [196], the extended ensemble Nosé-Hoover scheme [197, 198], or a velocity-rescaling scheme [199] to simulate constant temperature. In the Berendsen weak-coupling method velocities are rescaled deterministically after each step so that the system is forced towards the desired temperature. This thermostat does not strictly fix the temperature, but leads to exponential relaxation of instantaneous temperatures to a target one. It is best applied during the early stages of equilibration, as it converges relatively quickly and can be useful for stabilizing systems that may be far from equilibrium. The Nosé-Hoover thermostat is an extended-system method for controlling the temperature of the simulated system. It allows temperatures to fluctuate about an average value, and uses a damping factor to control the oscillation of this temperature. This thermostat produces a correct kinetic ensemble. However is appropriate to use it once the system has reached the equilibrium. The velocity-rescaling thermostat is essentially a Berendsen thermostat with an additional stochastic term that ensures a correct kinetic energy distribution and it still has the advantage of the Berendsen thermostat: first order decay of temperature deviations and no oscillations.

Pressure coupling: GROMACS supports both the Berendsen algorithm and the extended-ensemble Parrinello-Rahman approach [200, 201] Berendsen barostat scales coordinates and box vectors every step. It is important to note that although the Berendsen pressure control algorithm yields a simulation with the correct average pressure, it does not yield the exact NPT ensemble¹. Parrinello-Rahman barostat is similar to the Nosé-Hoover temperature coupling, and in theory gives the true NPT ensemble. Berendsen and Parrinello-Rahman approaches can be combined with any of the temperature coupling methods above.

¹It is often advisable to perform equilibration using weak-coupling techniques for temperature and pressure, especially if the system is far from equilibrium.

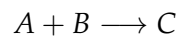
2.3 Brownian Dynamics

2.3.1 Biomolecular association reactions

Many binding events *in vivo* are governed by kinetic rather than equilibrium binding properties. For several binding partners, it is not sufficient to bind tightly, they must also bind quickly. Fast association may enhance binding affinity [202], especially for proteins involved in signaling pathways. The bimolecular binding process can be considered to consist of two steps: diffusional association to form a diffusional encounter complex (which is strongly affected by electrostatics interactions), followed by nondiffusional rearrangement to form the fully bound complex.

In order to characterize the kinetic behavior of a reaction, it is desirable to determine how the rate of reaction varies as the reaction progresses. The rate of reaction depends on the product of the concentration of both species involved, which makes bimolecular reactions second-order reactions. To better understand how the kinetic of two molecules occurs some basic concepts will be introduced below.

Given a hypothetical chemical reaction:



The general rate law is:

$$rate = k[A]^x[B]^y \quad (2.9)$$

where $[A]$ is the concentration of species A, x is the order with respect to species A. $[B]$ is the concentration of species B, y is the order with respect to species B, and k is the rate constant. The rate of reaction is the change in concentration over the change in time, *rate* refers to the speed at which a reactions occurs. The relationship between the concentrations of species to the rate of a reaction is the reaction order. Once the rate law of a reaction has been determined, that same law can be used to understand more fully the composition of the reaction mixture. More specifically, the reaction order is the power to which the concentration of that species is raised to. For the second-order reaction, the rate of reaction is directly proportional to the product of two reactant concentrations (or the square of one concentration).

$$rate = k[A][B] = k[A]^2 \quad (2.10)$$

As was mentioned above, many important biological reactions can be described using second order kinetics. This information is especially useful for determining how a reaction occurs. The observed rate constant of proteins associations comprise a wide range from 10^3 to $10^9 \text{ M}^{-1}\text{s}^{-1}$ [203].

In 1917, Smoluchowski [204] showed for the first time the theoretical result for the association rate constant of two spherical particles. Since that moment, different algorithms have been developed [205, 206] to compute association rates for biomolecules. Brownian dynamics (BD) is the primary method for the computational study of biomolecular diffu-

sional association [207, 208]. BD is a diffusional analogue of MD where particles move according to some equations of motion (just like MD). The interacting solute molecules move in a viscous continuum solvent that exerts stochastic forces on the solute molecules. This approach has been used by many groups [209–211].

2.3.2 BrownDye package for Brownian Dynamics

BrownDye algorithm computes the diffusional encounter of two large biomolecules. It can be used to estimate second-order rate constants of two rigid bodies and encounter probabilities. The solvent is modeled using continuum models (see subsection 2.2.5) along with a simplified description of the macromolecules. The Brownian motion of the macromolecules performs a random walk with no defined linear or angular velocities. The general equation describing the motion has the form [212, 213]:

$$d \begin{pmatrix} x_1 \\ \phi_1 \\ x_2 \\ \phi_2 \end{pmatrix} = \frac{dt}{k_b T} \mathbf{D} \cdot \begin{pmatrix} F_1 \\ T_1 \\ F_2 \\ T_2 \end{pmatrix} + \sqrt{2dt} \mathbf{S} \cdot \mathbf{w} + \nabla \cdot \mathbf{D} dt \quad (2.11)$$

where the vectors $x_1, x_2, \phi_1,$ and ϕ_2 are the positions and orientations of the molecules, dt is the time step size, T is the temperature, k_b is Boltzmann's constant, the matrix \mathbf{D} is a generalization of the diffusivity which can also depend on positions and orientations. F_1, F_2, T_1, T_2 are the forces and torques respectively. The matrix \mathbf{S} is the matrix square root of \mathbf{D} . The vector \mathbf{w} comprises 12 independent random Gaussian variables, each with zero mean, and unit variance and ∇ is the gradient operator.

The forces in equation 2.11 are grouped in two categories: long-range and short-ranged. For protein-protein association the dominant long-range force is provided by electrostatic interactions. The electrostatic forces are computed from precomputed solutions to the Poisson–Boltzmann equation describing the electric field around each molecule in isolation:

$$\nabla \cdot [\epsilon \nabla \phi] = -\rho - \lambda \sum_i c_i z_i \exp \left[-\frac{z_i \phi}{k_b T} \right] \quad (2.12)$$

where ϕ is the electric potential, ϵ is the dielectric of the material (either solvent or macromolecule), ρ is the permanent charge density in the macromolecule, and c_i and z_i are the bulk concentration and charge of the i th ion species in the solvent. The factor λ depends on position. The electric potential around each molecule is typically computed by using the APBS software package [195].

The short-ranged forces are also of two types. The default short-ranged force is simply a hard-sphere exclusion force. If, two spheres on different molecules overlap after a time step, the step is rejected. BrownDye also includes an option to use Lennard–Jones forces between spheres. This is described by the pairwise potential energy between two spheres:

$$V(r) = \epsilon \left[\frac{r_m^{12}}{r} - 2 \frac{r_m^6}{r} \right] \quad (2.13)$$

where r is the intersphere distance, ϵ is energy well depth, and r_m is the sum of the two radii

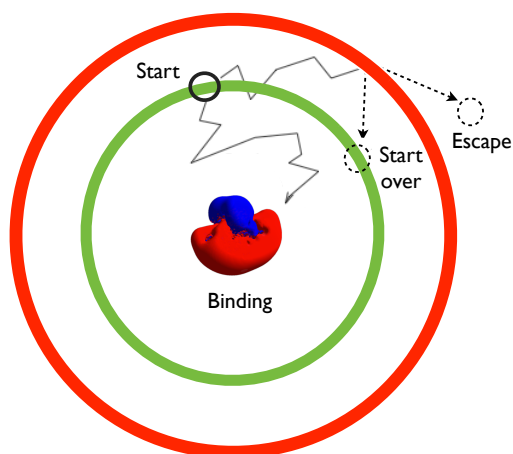


Figure 2.5: Schematic representation of BD simulations. A trajectory is started by positioning the molecule centers a distance b (green spheres) from each other, with randomly chosen orientations. The molecules are stepped forward until they either react, in which case the trajectory ends, or they attain a distance q (red sphere) slightly larger than b . A random number is then used to decide whether the molecules escape from each other, thereby ending the trajectory. If they do not escape, they are placed at a distance b from each other but, with the orientations drawn from a specific probability distribution, and the trajectory continues.

After the desired number of completed trajectories has been generated (usually several thousands), the rate constant is computed from the ratio of reacted versus escaped trajectories [214]. Probabilities and rate constants are reported with 95% of confidence intervals.

One of the limitations of this algorithm is the approximation of flexible molecules as rigid bodies ruling out realistic simulations where fluctuations are important. However, it is very useful for computing relative rates among different experimental conditions. BD is computationally less expensive than other approaches and it is complementary to MD.

2.4 Analysis of trajectories

GROMACS offers plenty of analytical tools for the analysis of MD trajectories. In the current section a summary of the most important tools used in this study is given.

A rate constant can be computed from the trajectories according to equation 2.11, besides, the definition of a reaction criterion to decide whether a reaction takes place or not is required. BrownDye package has implemented a newer version of the Northrup–Allison–McCammon (NAM) algorithm [205] called Luty-McCammon-Zhou (LMZ) algorithm [214]. It allows the use of a smaller q -radius (Figure 2.5), but either places the smaller molecule back on the b -sphere according to an analytically computed probability distribution without yet terminating the trajectory, or assuming an escape and terminating the trajectory. This implementation also takes into account the change in orientation of the smaller molecule during its presumed journey from the q -sphere back to the b -sphere, using an enhanced probability distribution that includes not only the new starting position but also the time taken to get there. In general, BD simulations are performed as is described in Figure 2.5.

2.4.1 Root mean square deviation

The root mean square deviation (RMSD) is a quantity that can be used to estimate the structural stability of a protein. By least-square fitting it allows to evaluate the deviation of a given structure from the original starting structure over the course of the simulation. Each structure in the trajectory is compared to a reference structure.

$$RMSD(t) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|r_i(t) - r_i^{ref}\|^2 \right]^{\frac{1}{2}} \quad (2.14)$$

where $M = \sum_i m_i$, and $r_i(t)$ is the position of atom i at time t . The fitting can use the same group of atoms as the RMSD calculation or a different one, e.g., the RMSD can be computed of the backbone only or the whole protein is fitted on the backbone atoms.

2.4.2 Root mean square fluctuation

The root mean square fluctuation (RMSF) is a measure of the deviation between the position of particle i and some reference position²:

$$RMSF_i = \left[\frac{1}{T} \sum_{t_j=1}^T |r_i(t_j) - r_i^{ref}|^2 \right]^{\frac{1}{2}} \quad (2.15)$$

where T is the time over which one wants to average and r_i^{ref} is the reference position of the particle i . Since it give some insights into the mobility of the protein, is a valuable tool to analyze the flexibility of a protein.

2.4.3 Temperature factors derived from RMSF

Temperature factors or also called B-factors can be computed as:

$$B = \frac{8\pi^2}{3} RMSF^2 \quad (2.16)$$

The B-factors can be interpreted as indication of the relative vibrational motion of different parts of the protein. Atoms with low B-factors belong to a part of the structure that is well-ordered. Atoms with large B-factors generally belong to part of the structure that is very flexible. The computed B-factors can be compared to X-ray crystal structure B-factors.

²Differences between RMSF and RMSD: The RMSF is averaged over time giving a value for each particle i . The RMSD calculates the average over the particles giving time specific values.

2.4.4 Cluster analysis

Clustering of MD trajectories can be performed using different methods. In this study, the single-linkage algorithm was used. It is one of the several methods of agglomerative hierarchical clustering where each element starts in its own cluster and at each step pairs of clusters are merged until certain termination conditions are satisfied. The proximity of two clusters is defined as the minimum distance between any two points in the two clusters. The linkage function is described by the expression:

$$D(X, Y) = \min_{x \in X, y \in Y} d(x, y) \quad (2.17)$$

where X and Y are any two sets of elements considered as clusters, and $d(x, y)$ denotes the distance between the two elements x and y . This algorithm adds a structure to a cluster when the distance to any element of the cluster is less than a certain RMSD cutoff. Representative structures (cluster centers) are then selected for further analysis. The center of a cluster is the structure with the smallest average RMSD from all other structures of the cluster.

2.4.5 Principal Components Analysis

Through Principal Components Analysis (PCA) -also called covariance analysis- one can determine what internal motions contribute the most to the overall dynamics of the protein. In a system of N atoms, there are $3N-6$ modes of possible internal fluctuations (six degrees of freedom are required to describe the external rotation and translation of the system). It uses the covariance matrix C of the atomic coordinates:

$$C_{ij} = \left\langle M_{ii}^{\frac{1}{2}}(x_i - \langle x_i \rangle) M_{jj}^{\frac{1}{2}}(x_j - \langle x_j \rangle) \right\rangle \quad (2.18)$$

where M is a diagonal matrix containing the masses of the atoms (mass-weighted analysis) or the unit matrix (non-mass weighted analysis). C is a symmetric $3N \times 3N$ matrix, which can be diagonalized with an orthonormal transformation matrix R :

$$R^T C R = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N}) \text{ where } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{3N} \quad (2.19)$$

The columns of R are the eigenvectors, also called principal or essential modes. R defines a transformation to a new coordinate system. The trajectory can be projected on the principal modes to give the principal components $p_i(t)$:

$$p(t) = R^T M^{\frac{1}{2}}(x(t) - \langle x \rangle) \quad (2.20)$$

The eigenvalue (λ_i) is the mean square fluctuation of principal component i . The first few principal modes often describe collective, global motions in the system. The trajectory can be filtered along one (or more) principal modes.

2.4.6 Buried surface area calculation

The solvent-accessible surface area (SASA) was measured with the program NACCESS [215] which implements the algorithm of Lee & Richard [216]. In simple words, a protein is described by a set of solvated van der Waals spheres, and a probe of given radius is rolled around the surface of the molecule. The SASA is calculated then as the surface traced out by the center of the probe (see Figure 2.6). In this work, a probe of 0.14 nm radius (which approximates the radius of a water molecule) and default van der Waal radii were used.

Once the SASA is determined, the buried surface area (BSA) is calculated as the sum of the SASA of the two components or proteins (x and y; in equation 2.21) less that of the complex. Measurements of SASA and BSA are usually described in units of \AA^2 or nm^2 .

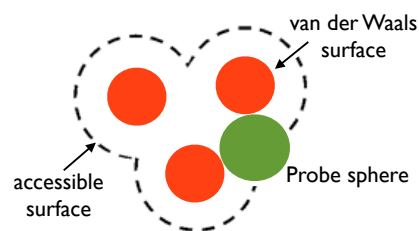


Figure 2.6: Solvent-accessible surface area of a molecule. The accessible surface is drawn with dashed lines and is created by tracing the center of the probe sphere (green) as it rolls along the van der Waals surface (red).

$$BSA = (ASA_x + ASA_y) - ASA_{xy} \quad (2.21)$$

2.4.7 Hydrogen bond calculations

Hydrogen bonds were identified with program HBPLUS [217]. The algorithm for locating hydrogen bonds involves two steps. Firstly, finding the positions of the hydrogens, and secondly, calculating the hydrogen bonds. OH and NH groups are regarded as donors, O is an acceptor always and N is a donor by default. If the donor and acceptor are only one or two covalent bonds apart, the interaction is not counted as a hydrogen bond. Default parameters were used for the calculation and a distance of 0.3 nm was settled as a cut-off in all systems.

“The **task** is, *not*
so much to see
what no one has
yet seen; **but** to
think what *nobody*
has **yet** thought,
about *that* which
everybody **sees.**”

– Erwin Schrödinger –

CHAPTER 3

CALCIUM BINDING REGULATES PUTATIVE SONIC HEDGEHOG PEPTIDASE

3.1 Introduction

HhN proteins are important morphogens, essential for development in bilateral animals [15]. As morphogens, they diffuse through the tissues, presumably in the form of large oligomers [47], and form extracellular concentration gradients. Given the profound impact of this process, it is clear that it has to be tightly controlled. Mathematical modeling [218, 219] has pointed to self-enhanced degradation of the morphogen as one possible mechanism to establish the predicted concentration gradients. Molecular feedback loops that lead to self-enhanced removal of Hh, e.g. by receptor–ligand internalization in the target cells, have indeed been found [220]. However, the most simple molecular implementation of this Hh removal would be self-digestion, requiring, of course, that HhN is a peptidase.

A conspicuous feature of the HhN protein is its triple metal center with 1 Zn^{2+} and 2 Ca^{2+} ions. Despite the similar arrangement of this zinc center to other LAS enzymes, ShhN is the only member of this family without confirmed enzymatic activity. One proven function of the zinc center is the recognition of cell-surface receptors and their antagonists, though without cleaving these binding partners [94, 93, 158]. The double- Ca^{2+} binding site is evolutionary highly conserved and important for the specific recognition of several binding partners of ShhN [87, 94, 93]. Moreover, a number of missense mutations that cause BDA1 and HPE are within this calcium metal center [133, 122].

In this chapter, based on the available ShhN sequences and X-ray structures, the possible functions of the metal ion centers in ShhN has been explored. One of the hypotheses that are suggested by what is known from other calcium dependent zinc proteases is that ShhN with its LAS peptidase-like Zn^{2+} center could also be a Ca^{2+} activated LAS peptidase. To study this hypothesis and to characterize other possible roles of Ca^{2+} in ShhN, the effect of metal ions, in particular the effect of Ca^{2+} on the structure, dynamics,

and electrostatics of ShhN have been investigated by computational means.

3.2 Computational Methods

Molecular structures: All molecular structures were retrieved from the PDB. Three different states (here called Ca0, Ca1, Ca2) of murine ShhN with respect to Ca^{2+} binding were studied: ShhN without calcium (Ca0) based on PDB entry 1vhh [141], ShhN with one Ca^{2+} (Ca1) based on PDB entry 3n1r [88], and ShhN with two Ca^{2+} (Ca2) based on PDB entries 2wfx [94] and 3d1m [87]. In the latter two structures, ShhN is bound to Cdo and Hhip, respectively, and are called here as Ca2_{Cdo} and Ca2_{Hhip} . For consistent comparison, all systems of Ca0-Ca2 were considered between residues L40 and E189, a range common to all structures. As the few clipped residues were disordered and far from the metal binding sites, this manipulation was not considered critical.

ShhN structures were compared to structures of five representative LAS enzymes: Streptomyces Albus G D-Ala-A-Ala Carboxypeptidase (PDB entry 1lbu) [151], L-alanoyl-D-glutamate endopeptidase (PDB entry 2v0g) [153], VanX amino-peptidase (PDB entry 1r44) [148], peptidoglycan amidase MepA (PDB entry 1u10) [149], and glycyl-glycine endopeptidase LytM (2bop) [150]. Seven ShhN variants with mutations at the calcium binding site were generated with the program SCWRL4 [221]: D89V, E91K, E91G, D96N, D96E, T126N and E127K. The X-ray structure of 3d1m (Ca2) was taken as reference. The program uses a new backbone-dependent rotamer library based on kernel density estimates and a new potential function that results in improved accuracy of the predicted side-chain conformations.

MD simulations: The MD of mouse ShhN and LAS enzymes was simulated with the GROMACS 4.5 package [166]. The GROMOS96 43a1 force field was used with optimized Lennard-Jones parameters $C_6 = 2.0 \cdot 10^{-3} \text{ kJ/mol nm}^6$ and $C_{12} = 1.67 \cdot 10^{-6} \text{ kJ/mol nm}^{12}$ for Ca^{2+} interactions taken from Ref. [180].

Each system was simulated in triplicate using the following protocol. Initial protein structures were solvated in a rhombic dodecahedron box of SPC/E water with a minimum of 1.0 nm distance between protein and faces of box. Residues were assumed to be protonated according to their normal states at pH 7, with the exception of histidines. The protons were assigned to histidines after inspection of H-bond patterns in ShhN X-ray structures to the following nitrogens: $\text{N}_{\epsilon 2}$ for H134, H181, H183; $\text{N}_{\delta 1}$ for H141, and both for H135. Na^+ and Cl^- ions were added to neutralize the system at an ionic strength of 0.15 mol/l. The Particle Mesh Ewald method was used to compute electrostatic interactions under periodic boundary conditions. Structures were energy minimized and equilibrated by molecular dynamics for 1 ns. Production simulations were run for 10 ns with a time step of 2 fs. NPT conditions were used with temperature and pressure stabilized at 300 K by Nosé-Hoover thermostat and 1 atm by Parinello-Rahman barostat, respectively. Bonds were constrained using the LINCS algorithm. Snapshots of the trajectories were saved every 100 ps. The last 5 ns of all trajectories were used for the analysis. Representative structures for the different calcium binding states were extracted from trajectories with `g_cluster` based on mutual RMSDs; the structures selected are close to the cluster centers

and represent about 90% of the trajectories. For all statistical analyses R version 2.12 [222] was used.

Matching of molecular structures: Similarity of molecular structures was quantified with EpitopeMatch [223], version 2011.11.22 (<http://www.EpitopeMatch.org>). For two molecular structures or structures of discontinuous fragments given by their geometries (e.g. atomic coordinates) and physico-chemical properties (e.g. atom types), EpitopeMatch aligns the two structures so that the Euclidean distances between physico-chemically similar parts are minimized. The quality of the match between the two structures can be described in terms of RMSD, number of matched components, or a similarity index that combines several quantities. Here, it has been used a matching mode of EpitopeMatch in which the two structures are aligned so that a maximum number of chemically identical atoms from both structures is superimposed, and between these the match with the lowest RMSD is selected.

EpitopeMatch was used for three purposes: (1) to characterize similarity between LAS enzyme zinc centers taken from X-ray structures and ShhN structures taken from MD trajectories; (2) to characterize the similarity between LAS enzyme zinc centers based on their MD trajectories and (3) to identify amongst the complete PBD (about 88.741 structures) for an arrangement similar to the zinc center of ShhN.

Electrostatic calculations: Electrostatic potentials were computed by solving the non-linear Poisson-Boltzmann Equation with APBS [195] using a grid with 0.05 nm spacing, dielectric constants of 79 and 2 outside and inside ShhN, respectively, and a water probe of radius 0.14 nm. Temperature was set to 310 K, ionic strength to 0.1 mol/l NaCl outside ShhN and 0 mol/l inside ShhN. Zinc and calcium pqr files were generated using a charge of +2 and an ionic radius of 0.088 and 0.114 nm respectively. Charges and radii were assigned to the protein with PDB2PQR v.1.8 [224] using AMBER99 [225] force field parameters.

BD simulations: BrownDye program [167] was used first to estimate the calcium association of ShhN, and second, to investigate the influence of the Ca^{2+} ions in the binding of partners such as Hhip, Cdo and the monoclonal antibody 5e1 to ShhN. In general, charges and radii were assigned with PDB2PQR version 1.8 [224] using the AMBER99 [225] force field parameters. The electric field around proteins was calculated using APBS [195], with a salt concentration of 0.1 mol/l. As described above, zinc and calcium pqr files were generated using a charge of (+2) and an ionic radius of 0.088 and 0.114 nm respectively. For both scenarios, 500 000 single trajectory simulations using nam_simulation were performed.

For the first purpose, a phantom atom of zero charge and negative radius of -0.114 nm was placed at the position of each calcium ion independently. In this way, two different systems were created; one contains the phantom atom at the position of Ca1 and a "real" calcium ion at the position of Ca2 (system called, Ca2) while the other system contains the phantom atom at the position of Ca2 and a "real" calcium ion at the position of Ca1 (system called, Ca1). This phantom atom has no influence in the association rate

constant calculation, its function is to define a reaction criterion that is spherically symmetric around the expected binding position of the calcium. The reaction criterion was chosen to be 0.15 nm within the calcium binding site. Simulations were performed using the same conditions but changing the salt concentration to (0.005, 0.01, 0.015, 0.022, 0.075, 0.1 mol/l). The X-ray structure 3d1m (2Ca²⁺) as well as representative structures taken from MD simulations were used to calculate the second-order rate constant. For the second purpose, PDB entries 3d1m, 2wfx and 3mxw [158] in complex with Cdo, Hhip and 5e1 respectively were studied. The three states Ca0, Ca1 and Ca2 were compared.

pKa calculations: PROPKA version 3.1 [226] was used to estimate the pKa values of side chains in ShhN. Default parameters were used to perform the calculations. Coupling effects between protonatable groups were included.

Phylogeny: From UniProtKB (last access Feb, 2013) all 30 reviewed amino acid sequences of full length Hh proteins were retrieved. The set contained sequences of Dhh, Ihh, Shh, and Tiggy-Winkle Hedgehog from *Drosophila* species and several vertebrates. The sequences were aligned with t-coffee [227], version 8.99, with default settings. Based on this alignment a distance based tree with BioNJ [228] was computed. Branching confidence was assessed by 100 bootstraps.

Docking: The cholesterol molecule was taken from ZINC database [229] (ID: 3869467). The mol2 file was translated to PDB format using Chimera v.1.6.2 [230]. The protein structure used was PDB entry 1vhh. The docking was prepared with AutoDockTools / MGLTools v.1.5.6 (The Scripps Research Institute) by adding polar hydrogens and assigning charges to all atoms. AutoDock Vina 1.1.2 [231] was used for all docking runs. General docking parameters for both calculations were kept at their default values, except for the exhaustiveness value, which was increased to 12. The docking grid had a size of 5.8 nm × 5.2 nm × 5.2 nm and covered the entire ShhN protein.

3.3 Results and Discussion

Both metals calcium and zinc stabilize ShhN structure

In order to investigate whether the calcium ions stabilize the structure of ShhN, MD simulations of ShhN structures with 0, 1, or 2 calcium ions were performed. For clarity and better comprehension of the results the following color code is used: red for Ca0 state (based on PDB entry 1vhh), blue for Ca1 state (based on PDB entry 3n1r) and black for Ca2 state (based on PDB entries 3d1m and 2wfx).

Figure 3.1 shows that the RMSF of the protein backbone strongly depends on the number of calcium ions. While the states Ca1 (1 Ca²⁺) and Ca2 (2 Ca²⁺) can barely be distinguished, state Ca0 (no Ca²⁺) has overall a higher RMSF, especially in the two Ca²⁺ binding loops L₂ (residues 88–94) and L₃ (residues 128–139), and in the neighboring L₁ (residues 66–72).

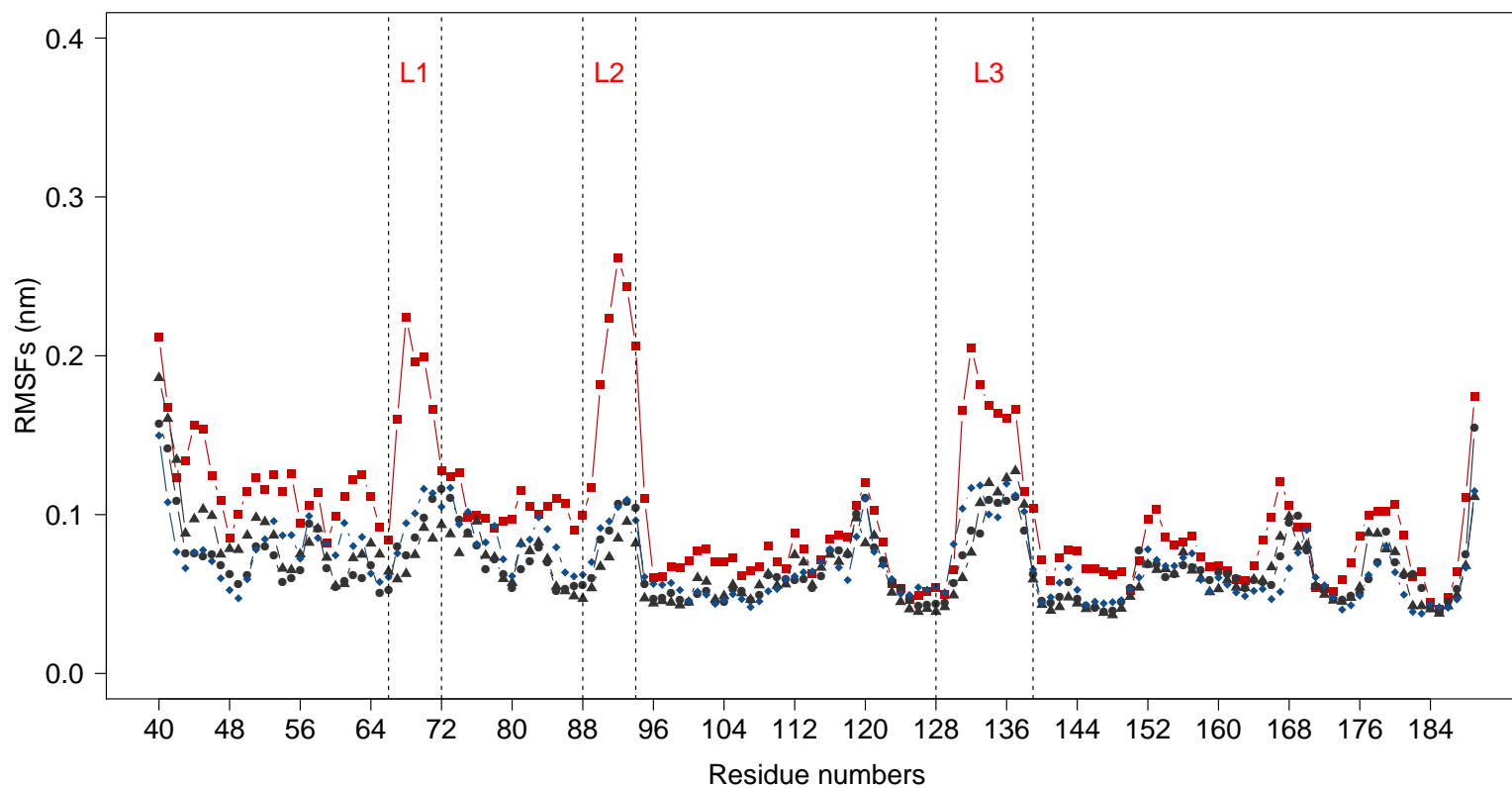


Figure 3.1: Root mean square fluctuation (RMSF) of protein backbone as function of ShhN residue number. The three calcium binding loops L₁, L₂, L₃ are delimited by vertical dotted lines. Color code: state Ca0 in red, state Ca1 in blue, state Ca2 in grey (based on 3d1m, triangles, and 2wfx, circles).

The flexibility pattern observed in the MD simulations is in good agreement with that derived from crystallographic B-factors (Figure C1 in Appendix C) and the experimental observation that L_2 , the loop with the highest RMSF values, is disordered in Ca0 [94]. The residue averaged RMSF values are 0.12 nm for Ca0, 0.09 nm for Ca1, Ca2_{Hhip}, and Ca2_{Cdo} (the latter two based on ShhN complexes with Hhip [94] and Cdo [87]). Wilcoxon tests to all pairs of RMSF distributions, i.e. Ca0 vs. Ca1, Ca0 vs. Ca2_{Hhip}, etc. were applied. It shows that Ca0 was the only different from the other states at a significance level of 0.05, while RMSF distributions Ca1, Ca2_{Hhip}, and Ca2_{Cdo} were not significantly different (see Table C1 in Appendix C). Thus, with respect to flexibility, the binding of the first Ca²⁺ switches ShhN from a more flexible state Ca0 to a more rigid state Ca1, while the binding of the second Ca²⁺ has little impact on flexibility.

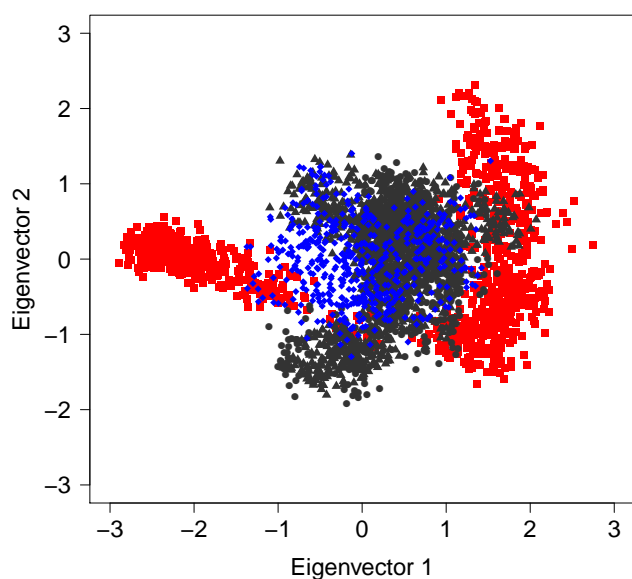


Figure 3.2: PCA plot comparing conformational dynamics sampling of different calcium states. Projections of the most dominant eigenvectors 1 and 2 of all backbone atoms are shown. Ca0 state in red, Ca1 state in blue, Ca2_{Cdo} and Ca2_{Hhip} in grey triangles and circles respectively.

Moreover, PCA analysis was used to compare the different simulated states (Ca0, Ca1, Ca2_{Cdo} and Ca2_{Hhip}) in terms of collective motions of backbone atoms. Given that only very small number of eigenvector contribute significantly to the overall motion of the protein, Figure 3.2 shows the projection of the eigenvectors 1 and 2 (the most dominants) into the principal component space. In Ca1 and Ca2 states several points overlap with each other and the area sampled is smaller than in Ca0 state. Ca0 state (red in Figure 3.2), reveals significant motion differences compared to Ca1 and Ca2 states. It samples a distinct section of the principal component space that correlates with the higher fluctuations discussed above. These motions are indica-

tive of the flexibility exhibit by loops L_1 , L_2 and L_3 in absence of calcium ions.

The transition from Ca2 to Ca0 not only increases the flexibility of ShhN, but induces a different group of structures: the removal of the calcium ions breaks up the binding site and pushes the loops L_1 , L_2 , L_3 close to the calcium ions away from each other (Figure 3.3). This is not surprising as the Ca²⁺ binding site abounds with negatively charged side chains that repel each other.

To further test the stabilizing effect of the calcium ions, the MD of states Ca0, Ca1, Ca2 based on X-ray structures of the corresponding states were simulated. The RMSD between initial structures and structures sampled in simulation was measured, expecting to see a higher RMSD for Ca0 and a lower RMSD for the calcium stabilized Ca1 and Ca2.

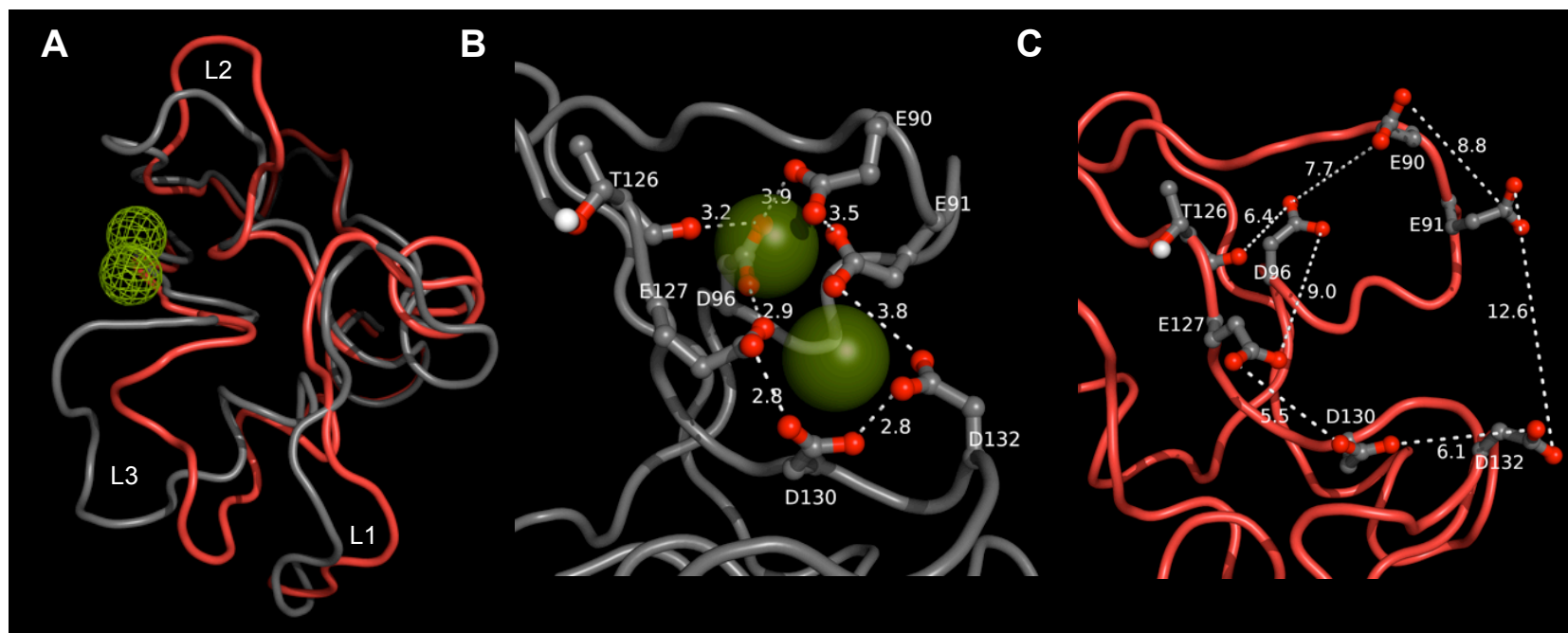


Figure 3.3: Effect of Ca^{2+} binding on ShhN structure. Structures taken from MD trajectories for Ca2 (grey; 3d1m) and Ca0 (red; 1vhh). Shown are sampled structures close to the cluster centers of Ca2 and Ca0 ensembles. (A) Overview and comparison of full ShhN structures in Ca2 and Ca0 states. Position of calcium ions marked by green mesh. Calcium binding pocket formed by loops L₂, L₃ breaks up as Ca^{2+} is removed (transition grey to red). Neighboring loop L₁ is also affected although it does not co-ordinate Ca^{2+} . (B) Close-up of Ca^{2+} (green spheres) binding site in Ca2 state. Calcium ions are surrounded by a cage of anionic side chains. (C) Same region as in panel (B), but now in Ca0 state. Note the large distance differences of anionic groups between (B) and (C). Numbers at the dashed lines are distances in 0.1 nm.

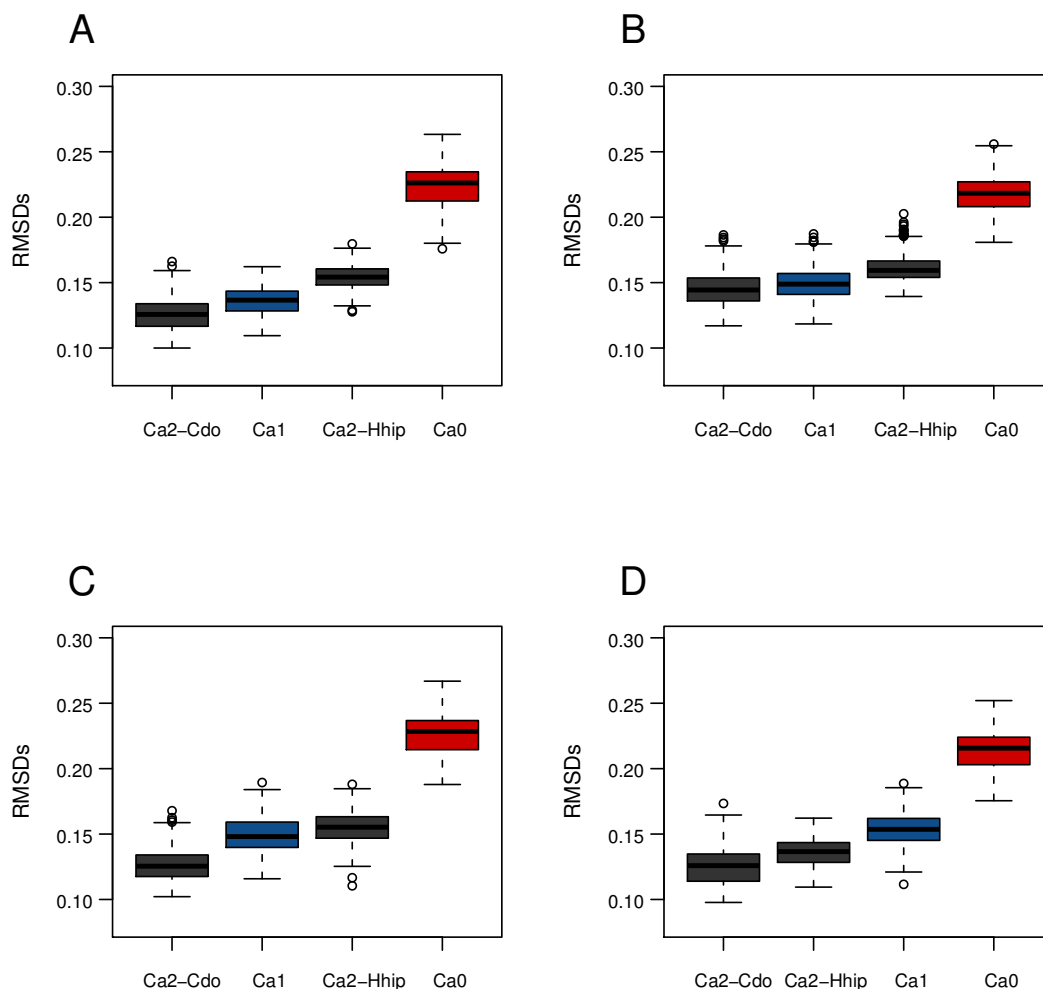


Figure 3.4: Time evolution of RMSDs backbone atoms from murine ShhN. The RMSDs values were measured between, on one hand, the four different X-ray structures representative of Ca2_{Cdo} (A), Ca2_{Hhip} (B), Ca0 (C), Ca1 (D), and, on the other hand, MD trajectories of ShhN in these states; e.g. the first boxplot (grey) in panel A gives the RMSD between 3d1m and an MD trajectory based on 3d1m structure. Boxplots within each of the four panels are sorted according to ascending median of RMSD.

Figure 3.4 shows that RMSDs are consistently higher (usually between 0.15 nm and 0.2 nm) for all Ca0 simulations, and consistently lower (usually 0.15 nm or lower) for Ca2 simulations, irrespective of whether the state was the original one in the X-ray structure, or artificially generated. Additionally, MD of artificial states Ca0, Ca1, Ca2, generated by removing calcium ions from X-ray structures of Ca2 or by introducing calcium ions into the binding pocket in the Ca0 X-ray structure were simulated.

The effect of the zinc ion in the stability of ShhN was also studied and MD simulations of ShhN based on 1vhh structure with and without zinc were performed. It is clearly observed in Figure 3.5 that the RMSD of the structure with Zn^{2+} (red) is lower than the structure without Zn^{2+} (gold) meaning that, besides the calcium ions the presence of zinc is required for the stability of ShhN. These results are in agreement with experiments [146].

Differences between the binding of the two calcium ions

Since ShhN in state Ca0 was found to be more flexible than both Ca1 and Ca2, while the latter two had similar flexibility (Figure 3.1), it seems that the two calcium ions are not equally important for structural stabilization. A closer inspection of the structure shows that, whereas one calcium is more exposed to the solvent, the other is almost completely engulfed by its ligands (see subsection 1.3.2 and in particular Figure 1.7).

Ca1 is the only calcium ion present in the PDB entry 3n1r that was used for the simulation of the Ca1 state (blue trace in Figure 3.1). In the following, this calcium ion is called Ca1, while the more solvent exposed calcium ion that completes the Ca2 state is called Ca2 calcium ion.

BD simulations have been used to estimate the association rate constant for Ca1 and Ca2 to ShhN respectively. The average calcium association rate constants for representative structures taken from MD simulations is shown in Table 3.1. It is shown, as well as, the rate constant estimated for the X-ray structure 3d1m (see Computational Methods for a detailed explanation about how these systems were created). Appreciable variation in the

association rate between the two calcium ions is observed in Table 3.1, strongly marked in MD simulations. This indicates that, once the Ca1 ion is bound to ShhN, the reaction is faster and the probabilities that Ca2 binds are higher than in the opposite scenario.

Another system with zero calcium ions at the binding pocket was also simulated to distinguish which of the two calcium binding sites is the favorite. In other words, the most likely binding pocket occupied by calcium. From 500 00 single trajectories performed, 419 reaction occurred and a rate constant of $3.1 \cdot 10^7 M^{-1} \cdot s^{-1}$ was obtained. In about 95 % of such reactions the calcium ion was bound to the position of Ca1 meaning that it is the preferential site occupied by calcium. If the extracellular concentration of calcium is considered

(about 1 mM) then Ca1 calcium ion should be always present under physiological conditions while the occupancy of the bound Ca2 calcium ion might depend on the specific

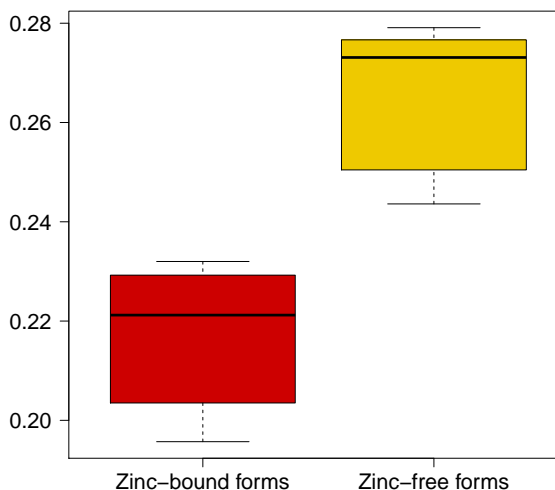


Figure 3.5: Effect of zinc ion on the stability of Shh. Whiskers and edges of boxplots mark quartiles of sampled backbone RMSD values (nm), bold bar in the colored boxes represents the median.

Table 3.1: Estimated second-order rate constants of calcium ions. Rate constants are reported with 95% confidence intervals and expressed in $M^{-1} \cdot s^{-1}$. Ca1 stands in this case for a calcium ion present at this position and a phantom atom at Ca2 position so, the reaction takes place at Ca2 position. The opposite is applied for Ca2.

Calcium states	Second-order rate constant	
	X-ray	MDs
Ca1	$1.8 \cdot 10^7$	$2.4 \cdot 10^7$
Ca2	$1.2 \cdot 10^7$	$0.5 \cdot 10^7$

environment of ShhN. For instance, extremely sensitive to low pH conditions. The latter, potentially provides a kind of Ca^{2+} sensor.

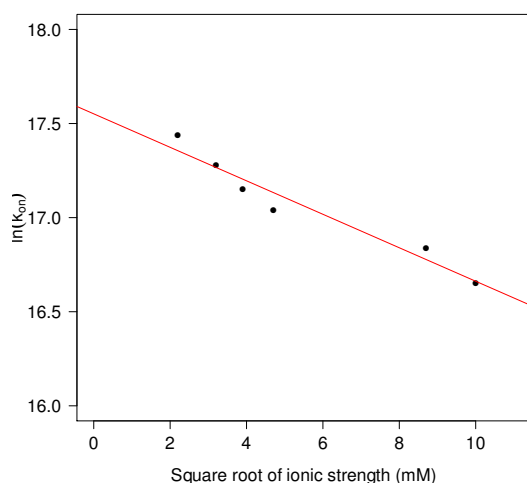


Figure 3.6: Dependence of ionic strength on calcium association rate. The square root of ionic strength is plotted versus the natural logarithm of the calcium association rate. k_{on} decays linearly with the root of ionic strength having a correlation coefficient of -0.95. The calculations were performed for Ca_2 calcium ion.

The role that electrostatic interactions play in the association of calcium can be appreciated in Figure 3.6. There is a linear correlation between the square root of ionic strength and the logarithm of the association rate constant (k_{on}) where as the ionic concentration increases the association rate decreases. Thus, electrostatic interactions are crucial in determining the binding strength of the calcium ions.

Naturally, the question arises whether binding of the Ca_2 calcium ion has specific effects on the structure and function of ShhN. A candidate mechanism could be the activation of the thermolysin-like or LAS zinc peptidase function by binding of the Ca_2 calcium ion.

Ca_2 calcium ion changes conformation of putative catalytic center

Although the zinc center of ShhN is usually referred as a "pseudo-active site" or "putative active site", ShhN is grouped into LAS family due to the similar arrangement of the zinc center to other LAS enzymes (see subsection 1.3.1 for more details). Figure 3.7 shows evidence of the high similarity between zinc centers of several peptidases [141, 147] and the zinc environment of ShhN. In all structures, the zinc ion is co-ordinated by nitrogens of two imidazole rings contributed by histidines, and a carboxylate group from a glutamate or aspartate residue. Both imidazole rings are fixated by hydrogen bonds between their N-H groups and neighboring hydrogen bond acceptors. In all four cases, one of these acceptors is a carbonyl oxygen, the other a carboxylate group. Nevertheless, ShhN is the only protein in the LAS group that exhibits a second metal center (double calcium binding). Therefore, the effect of Ca^{2+} binding on ShhN structure regarding to other LAS enzymes was investigated.

To address this issue, MD of ShhN in states Ca_0 , Ca_1 , and Ca_2 were performed and compared the conformations of the zinc environments sampled in this way with the X-ray structures of LAS peptidases with PDB entries 1lbu, 2v09, 1u10, 1r44, and 2bop. Contrary to the initial hypothesis, the addition of the Ca_2 calcium makes the zinc center of ShhN less similar to the zinc center of other LAS peptidases. Figure 3.8A shows that the median of the RMSD between the sampled zinc center conformations and the LAS peptidase zinc

centers were significantly *higher* in the Ca2 state. In other words, the binding of the second calcium makes the ShhN zinc center *less* similar to a LAS enzyme zinc center. The differences observed in Figure 3.8A between Ca0 and Ca2, and between Ca1 and Ca2 are significant (indicated by two stars on top of boxplots) while the differences between Ca0 and Ca1 are *not* significant (see Tables C2–C6 in Appendix C). This means that, whereas the Ca1 calcium ion is responsible for stabilizing the overall structure as described in the previous sections, the binding of the Ca2 calcium ion switches the zinc center from a LAS enzyme conformation to a significantly different conformation.

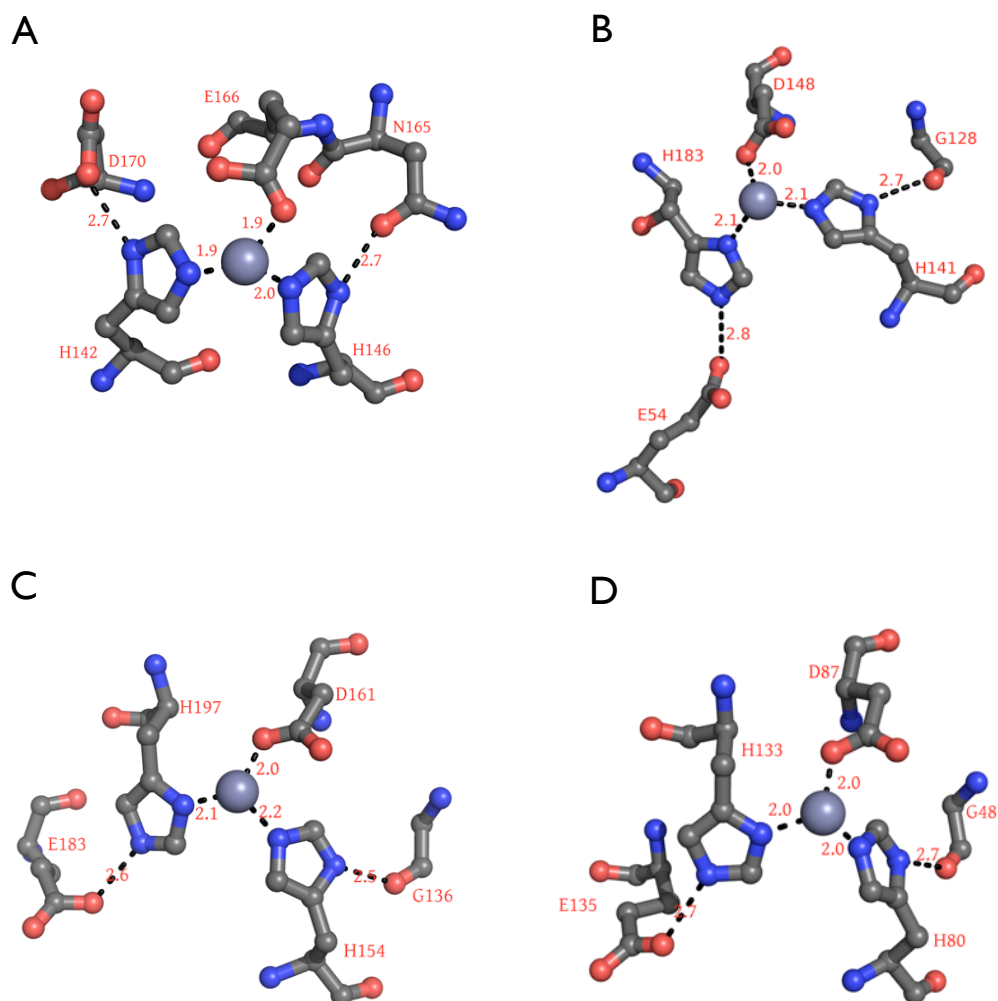


Figure 3.7: Zinc centers of different peptidases and ShhN. (A) Thermolysin, (B) ShhN, and LAS peptidases (C) *Streptomyces Albus G* D-Ala-A-Ala Carboxypeptidase (PDB: 1lbu), and (D) L-alanoyl-D-glutamate endopeptidase of a bacteriophage (PDB: 2v09). Numbers at dashed lines are characteristic distances in 0.1 nm.

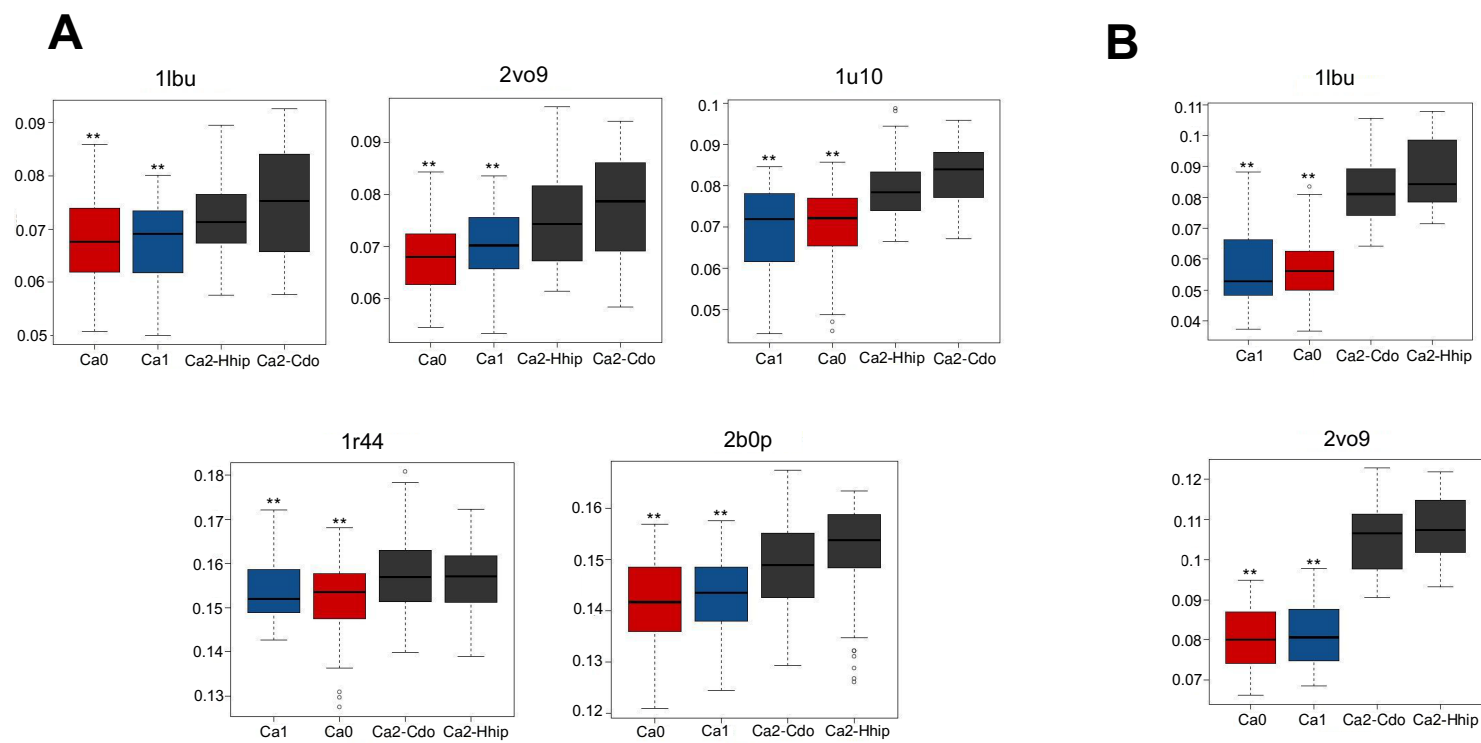


Figure 3.8: RMSDs between ShhN and LAS enzymes. (A) RMSDs between zinc centers from LAS enzyme X-ray structures to ShhN in states Ca0 (red), Ca1 (blue), Ca2 (grey) sampled by MD. (B) RMSDs between zinc centers from MD simulations of LAS enzymes and MD simulations of ShhN in states Ca0, Ca1, Ca2. The two stars on top of boxplots indicate significant differences between states Ca0 and Ca2, and between Ca1 and Ca2 ($\alpha=0.05$).

Although this pattern was consistent across the tested LAS enzyme structures, it could theoretically be an artifact related to the comparison of LAS enzyme X-ray structures and ShhN MD simulations. To exclude this, two LAS enzyme structures (PDB entries 1lbu and 2v09) that according to their X-ray structures had zinc centers geometrically most similar to the ShhN zinc center were simulated. Further, zinc center structures sampled by MD for both the LAS enzymes and the ShhN states were compared (Figure 3.8B). It was found that, ShhN zinc centers in Ca0 and Ca1 were significantly more similar to LAS zinc centers than ShhN zinc centers in Ca2 (see Tables C7 and C8 in Appendix C). Ca0 and Ca1 did not differ significantly in this respect, neither did different versions Ca2_{Cdo}, Ca2_{Hhip}. This result agrees with the previous one, again suggesting that binding of Ca2 calcium switches the zinc center of ShhN from a LAS enzyme conformation to a significantly different conformation.

Switch mechanism

The evidence above suggests the switches of the ShhN zinc center by binding of Ca2. The next question to address will be: By which mechanism does the binding of the Ca2 calcium switch the zinc center, which lies about 1.1 nm away?. The authors of the first X-ray structure [141] had proposed a reaction mechanism, closely related to that of Thermolysin involving seven residues and the zinc ion. They were unaware of the fact that ShhN binds Ca²⁺ in two well-defined binding pockets. Figure 3.9A shows the residues of the putative catalytic center in states Ca0, Ca1, Ca2 as obtained by X-ray crystallography, and, for orientation, the positions of the calcium ions. According to the postulated reaction mechanism E177 abstracts a proton from the catalytic water at the fourth tetrahedral co-ordination site of the zinc ion, followed by a nucleophilic attack of the OH⁻ on the substrate carbon.

Amongst the residues in the putative catalytic center, E127 is directly affected by Ca²⁺ binding as it co-ordinates both the Ca1 and the Ca2 calcium. The X-ray structures (Figure 3.9A) show that as the Ca2 calcium ion binds, the carboxylate of E127 is markedly drawn towards this calcium ion. Further, there is a hydrogen bond between E127 and H135 observed in the X-ray structures and MD simulations of all three calcium binding states. Hence, as E127 is dragged towards the calcium center, it pulls H135 with it. According to Ref. [141], the second N-H of the H135 imidazole could stabilize the peptidase substrate in a conformation amenable to hydrolysis by forming a hydrogen bond with the carbonyl-O of the substrate. In Ca2, with H135 pulled away, the substrate stabilization in this critical conformation will be affected.

Opposite to H135, at the other side of the substrate, lie the zinc bound water molecule and E177, the two actually catalytically active components according to the enzymatic model in Ref. [141]. As the Ca2 calcium ion is introduced, the gap between H135 and E177 widens by about 0.1 nm according to the X-ray structures (Figure 3.9C). This observation is in agreement with the previous statement that the binding of the Ca2 calcium ion is accompanied by a significant perturbation of the putative catalytic center, possibly affecting substrate stabilization.

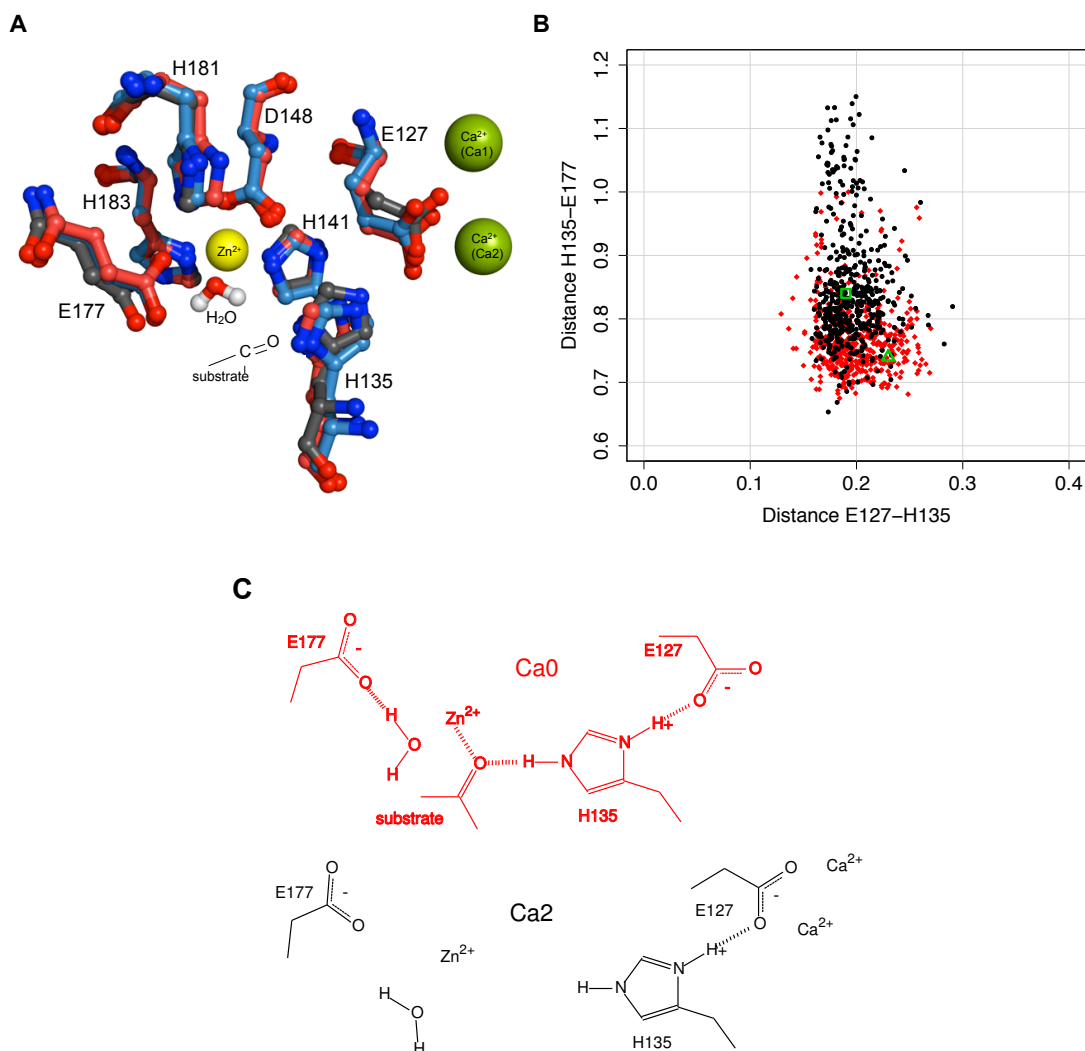


Figure 3.9: Switch mechanism triggered by Ca²⁺ calcium ion. (A) ShhN zinc center in states Ca0 (1vhh, red), Ca1 (3n1r, blue), Ca2 (3d1m, grey). Putative catalytic water from 1vhh is close to the zinc ion. From Ca0 to Ca1 and Ca2, E127 carboxylate is drawn towards Ca²⁺ and drags H-bonded H135 side chain with it, away from substrate and the active E177. While Ca0 and Ca1 superimpose well, Ca2 is clearly different. (B) Distances between H-bonded E127 carboxylate-O and H135 imidazole-proton, and between substrate-clamping side chains of H135 and catalytically active E177. Black (Ca2) and red (Ca0) points are sampled by MD simulations. Green triangle (Ca0) and green square (Ca2) are the corresponding values directly taken from X-ray structures 1vhh and 3d1m, respectively. (C) Central components of the switch mechanisms in states Ca0 and Ca2.

This pattern of conformational change due to the binding of the Ca²⁺ calcium ion is corroborated by the analysis of the MD trajectories of the different states. Figure 3.9B compares MD simulations of Ca0 and Ca2. It shows that the hydrogen bond between E127 and H135 is conserved between the Ca²⁺ binding states, despite the shift of E127 from Ca0 to Ca2 due to its attraction to the calcium center. Conversely, from Ca0 to Ca2 the distance widens between H135 and E177, thus opening the tight clamp previously gripping putative

substrate and catalytic water. Moreover, the variance of this distance increases with the binding of the Ca2 calcium. Numerical details characterizing the distance distributions are reported in Tables C9 and C10 in Appendix C.

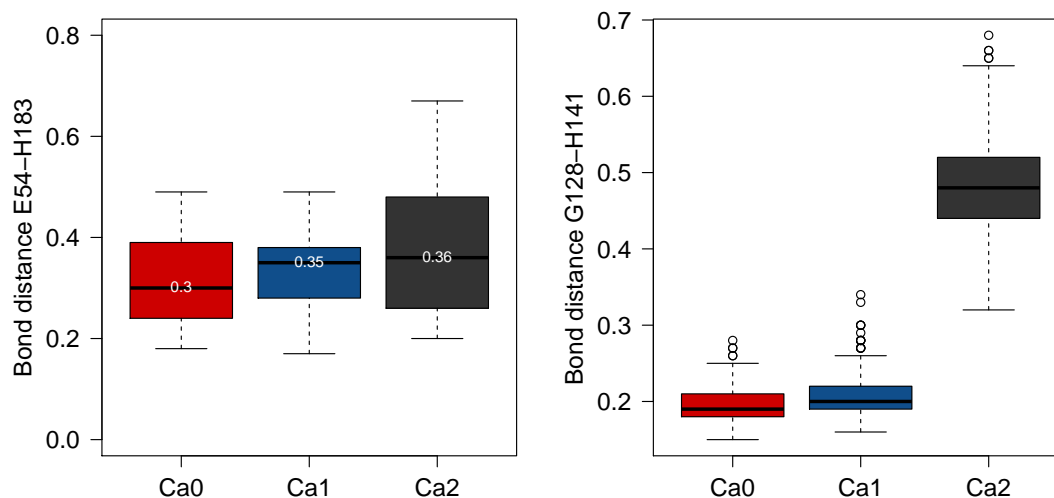


Figure 3.10: Effect of calcium binding on geometry of interactions that stabilize the zinc center. Shown are the distributions of distances (in nm; sampled by MD simulations) between carboxylate or carbonyl groups and histidine imidazole rings that co-ordinate the zinc. Left: the distance between groups from E54 and H183, both distal to the Ca^{2+} binding site, is barely affected by Ca^{2+} binding. Right: Binding of the Ca2 calcium ion breaks the hydrogen bond between G128 carbonyl and H141 imidazole N-H.

The pulling of E127 percolates towards the zinc center also along another route. In the description of the zinc center it was mentioned that the zinc co-ordinating histidines are stabilized by hydrogen bonds with carbonyl and carboxylate groups (Figure 3.7). One of these carbonyl groups comes in ShhN from the backbone of G128, the sequence neighbor of the Ca^{2+} co-ordinating E127. In Ca0 and Ca1, the G128 carbonyl forms a hydrogen bond with the zinc co-ordinating H141. In Ca2, as E127 is pulled towards the Ca2 calcium ion, the neighboring G128 is twisted away from the H141, and the bond between G128 and H141 is broken, the distance and distance variance is increased, and thus the zinc environment destabilized (Figure 3.10 right panel).

H183, the other zinc co-ordinating histidine, lies distal to the Ca^{2+} binding sites and is stabilized by a hydrogen bond with the carboxylate of E54, also distal to the Ca^{2+} binding sites. Consequently, the geometry of this interaction is barely affected by calcium ion binding (Figure 3.10 left panel).

Ca2 calcium ion switches the putative catalytic water

According to the enzymatic mechanism postulated by [141], there is another essential component of the catalytic zinc center: a water molecule occupying the fourth corner of the tetrahedron formed by the zinc ligands, as appears in the first X-ray structure [141].

This is another commonality with active LAS enzymes as noted by [147]. The accessibility of this co-ordination site to water is a necessary condition for an active enzyme with the same reaction mechanism, and it is fulfilled in ShhN.

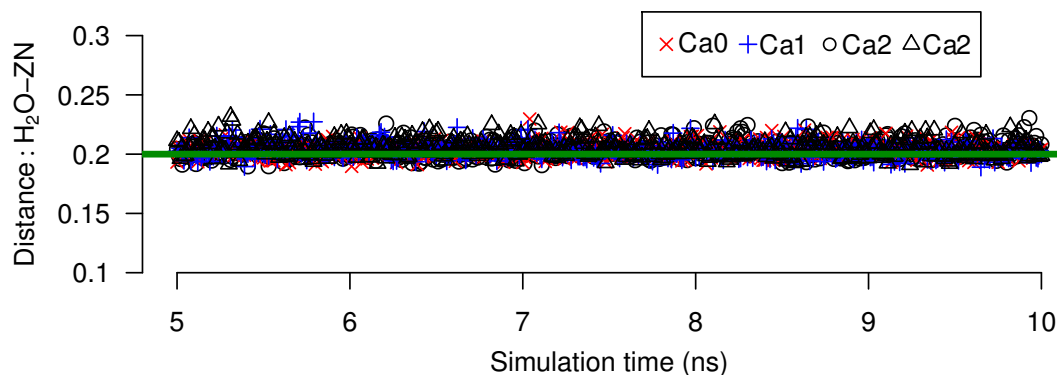


Figure 3.11: Distance between zinc ion and oxygen water close to the position of the putative catalytic water. The line represents the distance measured in the X-ray structure 1vhh [141].

If ShhN is a peptidase that employs the postulated mechanism, including the involvement of the water molecule at the zinc ion, and if further, this peptidase function is switched off by the Ca2 calcium ion, then one could expect that states Ca0 and Ca1 favor a water molecule at that position compared to Ca2. The behavior of water molecules close to the position of the putative catalytic water was analyzed in the MD simulations in terms of distance to the zinc ion and angles, e.g. the angle between the water oxygen, Zn^{2+} and $N_{\epsilon 2}$ of H148. There was in all simulations, irrespective of state (Ca0, Ca1, Ca2) almost always a water molecule at a distance of 0.2 nm to the zinc (Figure 3.11). This is not too surprising as the oxygen atom of water and the zinc have opposite partial charges, and therefore, given the water accessible zinc co-ordination site, there will always be a water molecule realizing this interaction. The angle measurements showed a more interesting pattern. For states Ca0 and Ca1 a water molecule constantly has an angle value very close to the crystallographically determined for all ligands (Figure 3.12). In these states, this water molecule does not exchange with other water molecules during the simulations. The picture looks qualitatively different for state Ca2. There, the water molecule closest to the position of the putative catalytic water assumes clearly different positions, and also flips between distinct angle values (Figure 3.12) exhibiting a dynamic behavior. An exchanges of water molecules at the zinc ion can be also observed in Ca2 state. This loss of a well-defined water co-ordination also affects the flexibility and position of catalytic E177, linked to this water by a H-bond. This contributes to the widening and loosening of the H135-E177 clamp introduced above.

This behavior of the putative catalytic water molecule is consistent with the conformational switching of the zinc environment between more LAS-like in Ca0 and Ca1 to less LAS-like in Ca2. The conservation of a water molecule at this well-defined position in states Ca0 and Ca1 makes also sense in the light of a hydrolase function in these states, and the loss of this water co-ordination is in agreement with a loss of hydrolase function in state Ca2.

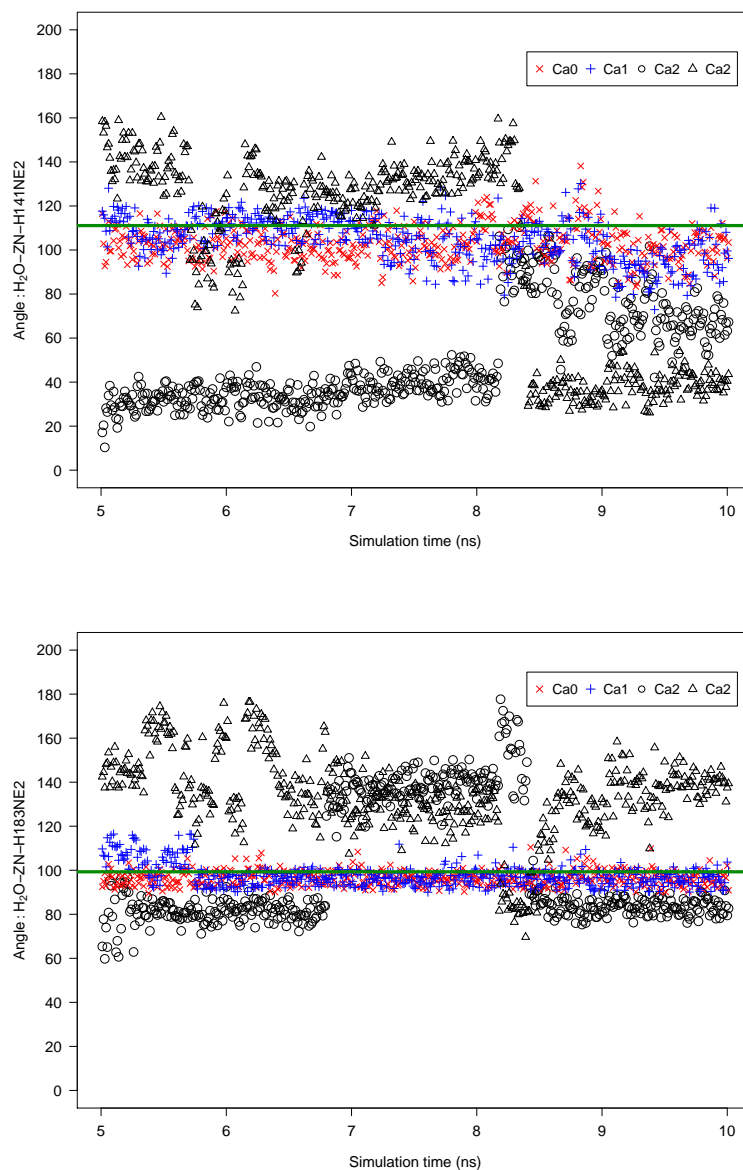


Figure 3.12: Angles (in degrees) as sampled by MD simulations, between oxygen of putative catalytic water, zinc ion and zinc ligands Zinc ligands are: $N_{\epsilon 2}$ of H141, $N_{\delta 1}$ and H183 and $O_{\delta 1}$; from left to right respectively. The green line marks the angle measured in the X-ray structure 1vhh [141].

Do non-enzymes have zinc centers similar to ShhN?

While the evidence presented so far is compatible with an enzymatic function of ShhN, it is unclear whether the structure of the zinc center is specific for enzymes, or whether other non-enzymes show a similar structure. To test this, EpitopeMatch [223] was used to search in the complete PDB for an arrangement similar to the one in the putative active site of ShhN [141], including the following eight groups: zinc ion, E127, H135, H141, D148, E177, H181, H183 (PDB entry 1vhh). Not surprisingly, all 20 Hh structures were

found with complete coverage of all eight groups and low RMSD (0-0.1 nm). Further, two matching proteins covering seven of the groups, a phosphodiesterase (1bf6) and a putative metalloprotease (3iuu) were found. Allowing a conservative exchanges of amino acids yielded in another group of proteins also with full coverage of eight residues, all of them enzymes, including eight LAS enzymes (PDB entries 1lbu, 1qwy, 1r44, 1u10, 2b13, 2b44, 2v09, 4f78), a hydrolase (3csq), and Thermolysin (8tln). No other proteins with a similar center were found. Thus in the set of currently available protein structures, the zinc center in ShhN is characteristic of enzymes.

Evolutionary conservation

Having an enzymatic function in an organism or not having it seems to be a fundamental property of a protein. One therefore would intuitively expect that the putative catalytic center of ShhN is evolutionary conserved. However, it had been noted from early on that in *Drosophila* the zinc binding site is not conserved [141, 86]. To assess the degree of evolutionary conservation a phylogenetic tree of Hh proteins (see Methods) was computed and checked where on the tree the putative catalytic center is present (Figure 3.13). The gross topology of the tree agrees with that of earlier trees [232] based on fewer sequences, and shows a clear branching between *Drosophila* hedgehog proteins and vertebrate hedgehog proteins, including Shh, Ihh, Dhh, and Tiggy-Winkle hedgehog. It was found that 16 out of 17 vertebrate Hh sequences contain the full putative EHHDEHH catalytic motif corresponding to mouse ShhN E127, H135, H141, D148, E177, H181, H183, while all of the *Drosophila* sequences carried instead the motif EHHTVHY, with the D148 and H183 co-ordinating zinc in ShhN replaced by a threonine and tyrosine in *Drosophila* hedgehog. The only sequence in the vertebrate branch that has not the full catalytic motif is ShhN of rat. In this sequence, all zinc co-ordinating residues are present, but H181 is replaced by an arginine. H181 is not co-ordinating the zinc ion but may fixate the catalytically active E177 in a position appropriate for interaction with the catalytic water. To test this, the structural comparison of the ShhN zinc center with the zinc centers of the LAS enzymes was checked again and revealed that there is at least one confirmed LAS enzyme, L-alanoyl-D-glutamate endopeptidase of a bacteriophage (PDB entry 2v09; [153]) that has a proline at the structural position corresponding to H181. This means that a histidine at this position is not strictly required for peptidase function, and hence rat ShhN could still be a peptidase. These results are compatible with a conservation of ShhN enzyme function in vertebrates. In the insect class the picture is less clear: while *Drosophila* hedgehog has no zinc center, Mosquito hedgehog has all seven residues of the putative ShhN active site [141].

The carboxylate carrying residues (in mouse ShhN E90, E91, D96, E127, D130, D132) that co-ordinate the calcium ions are almost completely conserved in all sequences, including *Drosophila* and vertebrates. The only exception is *Drosophila ananassae* Hedgehog protein in which E90 is conservatively replaced by an aspartate.

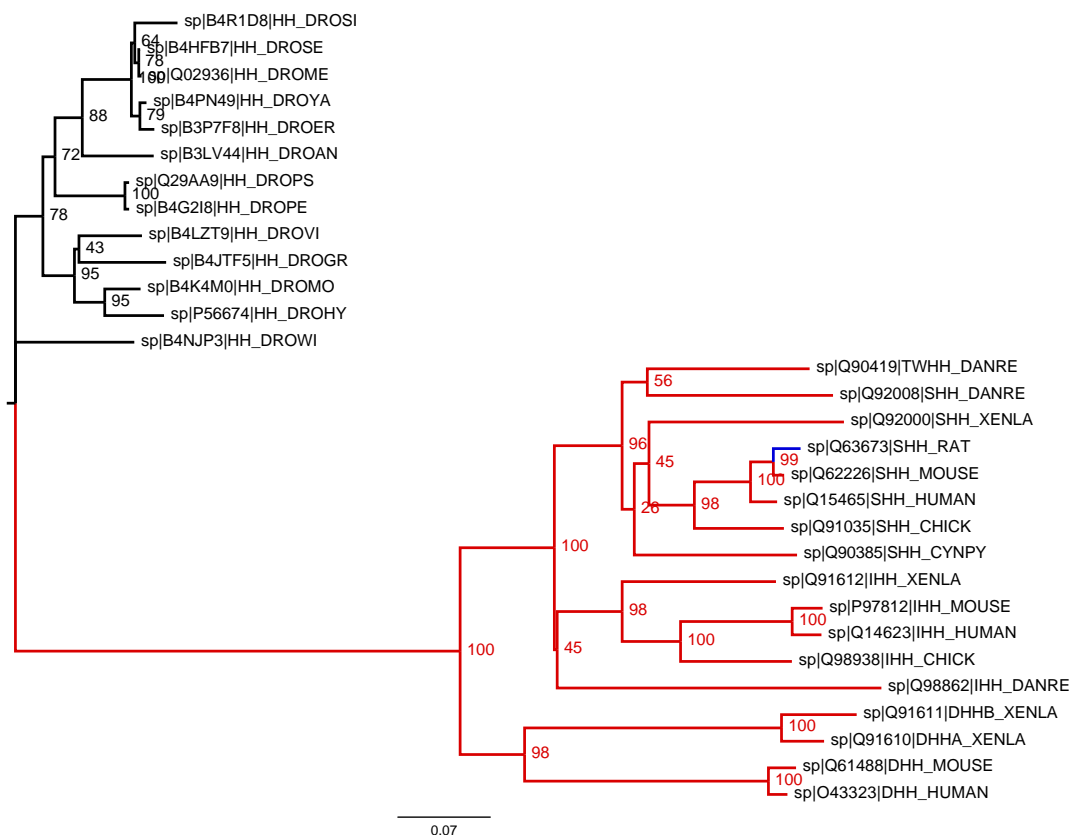


Figure 3.13: Phylogenetic tree of all 30 reviewed full length Hedgehog proteins from UniProtKB. The vertebrate Hedgehogs (bottom subtree) are clearly separated from the *Drosophila* Hedgehogs (top subtree). In all vertebrates the full catalytic motif is absolutely conserved (red), except in rat with one conservative exchange (blue).

Peptidase substrate

If ShhN is an enzyme, there should be a substrate. So far, no substrates of ShhN have been described, despite efforts to identify such substances [141, 145, 146]. By using the available ShhN structures, the set of possible substrates was narrowed. The search was restricted to peptide substrates, as most of the proteins with the highest structural similarity to the putative ShhN catalytic center are peptidases or peptidoglycan amidases [147].

The authors of the $[Ca^{2+}]$ free ShhN X-ray structure (1vhh) [141] noted that in the crystal, one ShhN molecule binds to the C-terminal peptide of its crystal lattice neighbor, so that this peptide comes close to the zinc ion but without replacing the putative catalytic water. The C-terminal carboxylate oxygen forms an isosceles triangle of side length 0.34 nm with the H_{δ} of H135 and the zinc ion, but is only 0.25 nm away from the oxygen of the catalytic water. It is well imaginable that this situation represents the state of the peptidase before the dissociation of the product. Except for the last amino acid, the C-terminal peptide SVAAK in 1vhh is not particularly polar. The amino group of the terminal lysine side chain is not engaged in contacts with the protein, but points away from it. This means that the

putative active site binds non-polar peptides, and thus that the peptide substrate could be non-polar. If this is true, the putative enzymatically active states Ca0 and Ca1 should present a low electrostatic potential close at that binding site, while the potential should flip to higher values in Ca2 to allow for instance the binding of Hhip. To test this the X-ray structure of Ca2 state (PDB: 3d1m) has been taken and computed the electrostatic potential around ShhN in Ca0, Ca1, Ca2 states, with the former two generated by removing calcium ions from the structure. Figure 3.14 shows that the zinc center region indeed changes electrostatic potential from slightly negative (Ca0), over slightly positive (Ca1), to high (Ca2). If one assumes that the strongly bound Ca1 calcium ion is mostly present under physiological conditions, the remaining active candidate state is Ca1. In this state, the zinc center region has an electrostatic potential close to zero, suggesting non-polar substrates, similar to the C-terminal peptide of ShhN discussed above.

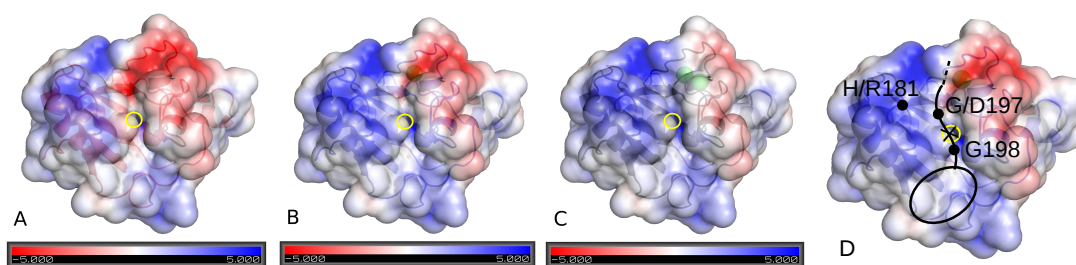


Figure 3.14: Electrostatic potential on solvent accessible surface (SAS) of ShhN in three calcium binding states. Panels show Ca0 (A), Ca1 (B), Ca2 (C), and model of ShhN-substrate complex (D) (PDB entry 3d1m). SAS is transparent to allow view on zinc (highlighted by yellow circle) and calcium ions (larger green spheres in upper right part of protein). Electrostatic potential ϕ coded by color and given on SAS. ϕ is given in units of $k_B T/e$, and a value of $+5$ corresponds to about $+12.5$ kJ/mol for singly charged molecules at room temperature. View is directly on putative catalytic center around zinc ion. Panel (D) shows Ca1 based model of the complex of ShhN peptidase with its substrate (black lines), i.e. the C-terminus of a neighboring ShhN, including the cholesterol modification (ellipse). Cleavage position is indicated by cross at zinc center. Substrate is embedded into hydrophobic pocket (cholesterol) and crevice (peptide G197, G198). Note proximity of H/R181 and G/D197 (compensatory mutations in rat).

In functional vertebrate ShhN, the C-terminal peptide is not SVAAK as in the X-ray structure 1vhh, but SVAAK(S/T)GG, which is still compatible with a rather non-polar peptide substrate. The only known exception of this rule in vertebrates is again rat where the first G (G197) is replaced by D197. Recall that in rat also has the mutation H181R. Hence, two neutral residues, H181 and G197, are replaced by two oppositely charged residues, R and D. This could be a compensatory pair of mutations if both residues were close in space. In fact, 1vhh shows that, if the C-terminus would be extended by (S/T)(G/D)G, the G/D197 in the C-terminal peptide of one ShhN protein could be very close to H/R181 of a neighboring ShhN protein (Figure 3.14D). Such a compensatory mutation makes, of course, only sense if the proximity between zinc center and C-terminal peptide observed in Ca0 (1vhh) is functionally relevant, e.g. if the C-terminal peptide is a substrate of a ShhN peptidase.

As mentioned above, ShhN proteins form large oligomers [47], a situation in which many ShhN molecules are close neighbors, and thus may expose their C-termini to cannibalysis [233] by neighboring ShhN molecules.

Conserved glycines in ShhN as possible cleavage sites

Most Lysostaphin-like enzymes (one of the LAS peptidase groups) are believed to cleave glycyl-glycin or glycyl-alanine peptide bonds [147]. Remarkably, there is a GG motif right at the C-terminus of ShhN. This GG motif is conserved in vertebrates, i.e. in organisms with an intact zinc center, but not in *Drosophila*, where the zinc center is missing. Cleavage of the peptide bond between these glycines would be functionally relevant, as the C-terminal G198 carries an important cholesterol modification [28, 74], and cleavage of the C-terminal glycine would remove this cholesterol. At the molecular level, removal of cholesterol may e.g. dissolve ShhN oligomers [63] and thus expose it to other peptidases, or it could increase the mobility of ShhN [234]. Very recently, it has been reported that a fraction of ShhN may act in an unlipidated form, though it remained unclear how this form is generated [52]. The removal of the cholesterol by ShhN proposed here could contribute to this fraction.

A GG-cholesterol motif is also a noteworthy substrate candidate because of its hydrophobic nature. There is a sizable, shallow hydrophobic pocket close to the zinc (marked with black ellipse in Figure 3.14D), extending from L140 over W173 to F48. This pocket has an electrostatic potential close to zero, independently of the Ca^{2+} binding state (Figure 3.14), which means that the binding of the cholesterol group could be decoupled from the peptidase activity.

To test whether the shallow hydrophobic pocket is suitable for binding cholesterol, molecular docking have been performed to the complete surface of ShhN with cholesterol as ligand. Due to the proposed binding pocket lies in the vicinity of the zinc ion, and the zinc is presumed to be co-ordinated in Ca0 and Ca1 by a well-defined water molecule, two docking runs were carried out, one with this water and one without. In both runs the hydrophobic pocket close to the zinc was identified as the dominant binding site, accommodating the cholesterol molecule in various orientations and poses (Figure C2 in Appendix C). There was only one further binding site in a distance of more than 2 nm to that pocket, but this alternative binding site turned up at rank 14 or worse amongst the 20 best poses (Tables C11 and C12 in Appendix C). The binding of the cholesterol in this hydrophobic pocket is compatible with the observed pair of assumed compensatory mutations between mouse (H181, G197) and rat (R181, D197): if one places G/D197 close to H/R181, the C-terminal G198 will hover just above the zinc ion, pointing with its cholesterylated C-terminus to the hydrophobic pocket (Figure 3.14D).

This extended hydrophobic pocket also indicates a possible mechanism for the extraction of cholesterol from the membrane. The pocket is large, shallow, and well-accessible at the surface. Thus, ShhN may dip with this part into the membrane and recruit a cholesterol C-terminally attached to a neighboring ShhN. Similar mechanisms have been proposed for other lipid binding proteins with extended hydrophobic regions [235, 236].

The other substrate pattern suggested by comparison with lysostaphins is glycyl-alanine. There are two such GA motifs conserved in ShhN. The first lies at G58, and this GA motif is conserved in vertebrates, but not in *Drosophila*. The other lies at G94 and is almost conserved in all Hh proteins (including *Drosophila*), except for Tiggy-Winkle Hh in zebrafish where this G is replaced by N. The GA motif at G58 seems to be a better candidate as the glycine is well-accessible, and it is part of a predominantly hydrophobic peptide TLGASG that is strictly conserved in all vertebrate Hh, but not in *Drosophila*.

The set of potential cleavage sites discussed above could be extended to a total of 13 conserved glycines, several of which are well exposed at the surface and have hydrophobic neighbor residues.

pH dependency of ShhN stability

The pK_a values for all protonatable groups in ShhN based on PDB entry 1vhh [141] were estimated. We found that the protonatable groups in the calcium binding site have pK_a values of 6 and lower (Table C13 in Appendix C). This means that at neutral pH the binding site should be fully deprotonated. However, there are in the calcium binding site three carboxylate groups (E91, D96, E127) with pK_a between 5 and 6, indicating a tendency for deprotonation at pH 5 and thus a weaker affinity to Ca^{2+} . Thus, ShhN peptidase could be switched on by shifting the pH from neutral to pH 5. *In vitro* experiments confirmed these predictions [237].

There is only one further group with a pK_a between 5 and 6, namely H181, with a pK_a of 5.5 in Ca0 (1vhh). In the first X-ray structure [141], H181 is described as forming a charged hydrogen bond with the carboxylate of putative catalytic E177, and thus stabilizing the latter. According to our pK_a estimate, the shifting of pH from 5 to 6 would make protonation of H181 less likely and thus destabilize E177.

Calcium ions enhance interaction between ShhN and binding partners

Table 3.2: Estimated second-order rate constants of the association between ShhN and regulators of the signaling. Rate constants are reported with 95% confidence intervals and expressed in $M^{-1} \cdot s^{-1}$.

Complex	Ca states		
	Ca0	Ca1	Ca2
ShhN-Cdo	-	$290.2 \cdot 10^5$	$464.6 \cdot 10^5$
ShhN-Hhip	$2.2 \cdot 10^5$	$5.9 \cdot 10^5$	$57.6 \cdot 10^5$
ShhN-5e1	$0.1 \cdot 10^5$	$0.4 \cdot 10^5$	$2.2 \cdot 10^5$

The evidence showed so far, supports the hypothesis that ShhN is a peptidase that is switched off by the binding of the Ca2 calcium ion. However, this fact does not imply that this is the only function of this calcium ion. It has been shown experimentally that calcium ions contribute also to the interaction between ShhN and positive or negative regulators of

the signaling such as Cdo, Hhip, Ptc and Gas1 [87, 88, 93, 94, 158]. To investigate the implication of calcium in the association of ShhN with different partners, BD simulations of three complexes were performed. ShhN-Cdo (based on 3d1m), ShhN-Hhip (based on

2wfx) and ShhN-5e1¹ (based on 3mxw) in Ca0, Ca1 and Ca2 states.

Table 3.2 shows the estimated second-order rate constants of the association between ShhN and binding partners. It is remarkable how the presence of Ca2 enhances the interaction in the three complexes while in absence of calcium ions the association between ShhN and Cdo does not occur. These results are in agreement with the data reported by [87] in which, the binding of ShhN to Cdo is abolished in the presence of a calcium chelator like EGTA. The rate value of $2.2 \cdot 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$ obtained in BD calculations for the complex ShhN-5e1 in Ca2 state (see last row and column in Table 3.2), reproduces well the experimentally observed rate of $2.09 \cdot 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$ [158]. Moreover, it has been shown experimentally that the presence of both Zn^{2+} and Ca^{2+} increases the binding affinity of partners for HhN from 10 to 20 fold on average [87, 94].

Despite some differences are appreciated in the association of ShhN with partners, in general, the interface area of Cdo, Hhip and 5e1 overlaps and the calcium ions seem to be critical for this interaction. Interestingly, the electrostatics of the region where the interaction takes place, heavily depends on the presence of calcium.

Disruption of the calcium binding site

Mutations within the Ca^{2+} center (D89V, E91K, E91G, D96N, D96E, T126N and E127K) were introduced in mouse ShhN to investigate whether they affect the stability of the calcium binding site. Four out of five of these residues are coordinating the calcium ions. Therefore, one should expect that somehow these single mutations alter the calcium center. Despite residue D89 does not contact the ion directly, it has been shown that stabilizes the conformation of the loop (L₂) by forming hydrogen bonds to two backbone amides [87].

All mutations disrupted the calcium binding to a greater or lesser extend compared to the wild type (Ca2 state). The introduction of single mutations as D96E and T126N are not so critical for the stability and maintenance of the calcium metal center. This fact is not surprising considering that both are neutral mutations. The analysis of first coordination shell of calcium ions with a cut-off of 0.3 nm reveals that, for D96E and T126N mutants Ca1 and Ca2 remain mainly engulfed by their native ligands. Residue E131 takes the position of D132 and this replacement permits that loops L₂ and L₃ stay close to each other and both, Ca1 and Ca2 keep bound to ShhN. Nevertheless, mutations like D89V, E91K, E91G, E127 and D96N grossly affect the calcium binding. The flexibility of binding loops L₂ and L₃ increases breaking up the binding site.

Figure 3.15 shows the behavior of ligand residues and oxygen waters residing in the first coordination shell of Ca1 (left panel) and Ca2 (right panel) for E91K mutant. While Ca1 is still coordinated by the side chain of carboxylate oxygens (note that D89 forms part of coordinating ligands), no residue ligands are present in the coordination of Ca2.

¹The monoclonal antibody 5e1 is a Hh pathway inhibitor that has been extensively used to elucidate vertebrate Hh biology due to its ability to block binding of Hh to the receptor Ptc.

This means that, the binding of Ca²⁺ is completely abolished and the pocket has been filled with solvating waters. Taking into account that, electrostatics play a crucial role in the association of calcium ions (see above) the change in the electrostatic charge from negative to positive (for instance, mutations E91K and E127K) might be one explanation for these findings.

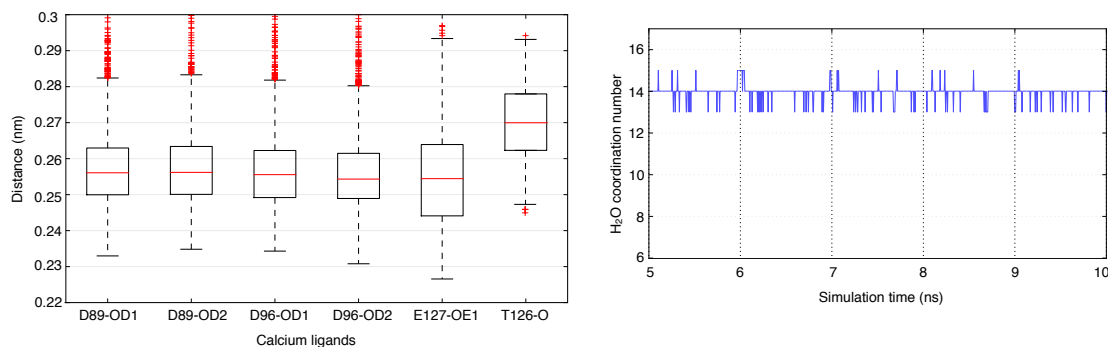


Figure 3.15: Ligand variations of E91K mutant in the first coordination shell of calcium ions. Left panel shows the residues residing in Ca1 during the simulation time analyzed and right panel shows the number of water molecules coordinating Ca²⁺. The analysis was done applying a distance cut-off of 0.3 nm.

These results are in agreement with experiments where the crystal structures of some IhhN mutants (Ihh mouse nomenclature: E95K, D100E and E131K) are reported [238]. Although these experiments were performed using Ihh protein, the high degree of similarity shared by ShhN and IhhN (about 92%) and the level of conservation of these residues in Hh sequences make possible the extrapolation of results. The authors found that, especially E95K (E91K in Shh mouse) alters the range of Ihh signaling *in vivo*. Besides, impaired interactions of Ihh with Ptc, Cdo and Gas1 caused by these mutations have been reported [87, 139]. Based on this, the disruption of Ca²⁺ binding observed here might be a plausible explanation for the impaired interaction observed between Hh and partners that can also alter Hh signaling. An alternative might be that these mutations affect the multimerization and the stability of HhN proteins.

3.4 Conclusions

First, the results described in this chapter are in agreement with the hypothesis of ShhN as a zinc exopeptidase that is stabilized by one calcium ion binding in a specific pocket, and switched off by a second calcium ion binding with a lower affinity in another pocket. The switching mechanism involves dragging of glutamic acid residue 127 towards the calcium center, a movement that is transmitted to the zinc center through several residues coupled to E127, notably histidine 135 which is bound to E127 by a hydrogen bond. The dragging of E127 pulls this H135 away from the putative substrate. Another notable effect of the binding of the second calcium is the destabilization of a catalytic water molecule that co-ordinates the zinc ion.

The C-terminal peptide of ShhN is proposed as peptidase substrate, including the cholest-

terol covalently attached to the C-terminal G198. This terminal cholesterylated glycine could be cleaved off by ShhN at low calcium concentrations, when only the Ca1 calcium binding site is occupied.

From the two mutations in rat ShhN compared to mouse ShhN one can conclude that these are a pair of compensatory mutations: H181_{mouse} → R181_{rat}, and G197_{mouse} → D197_{rat}, with the latter a direct neighbor of the C-terminal G198. Accordingly, it is predicted that chimeric ShhN with combinations H181/D197 or R181/G197 should be partially compromised in the assumed peptidase function. The predictions above are clearly experimentally testable. However, it is less clear where and when *in vivo* the calcium regulation of peptidase function comes into play, though phylogenetic conservation of all parts of its mechanism supports its relevance.

The question arises where and when ShhN encounters such low Ca²⁺ concentrations. [Ca²⁺] in blood is usually tightly controlled at a relatively high level, and hence this does not seem a likely place for enzymatic activation. More promising are interstitial or intracellular compartments. In various interstitial fluids [Ca²⁺] values between 30 μM [239] and 20 mM [240] have been reported, so that there the hypothesized enzymatic function could be switched on and off. In neural tissue, Ca²⁺ concentrations can be transiently depleted [241, 242]. ShhN could also be internalized into cells and exposed to lower [Ca²⁺] there.

The biological context is unknown that combines low [Ca²⁺] conditions with the necessity of removing the cholesterylated C-terminal glycine of ShhN. The fundamental importance of the ShhN cholesterol modification is well-supported by experiment [28, 71, 74]. At the molecular level, removal of cholesterol may e.g. dissolve ShhN oligomers [63] and increase the mobility of ShhN [234]. Very recently, it has been reported that a fraction of ShhN may act in an unlipidated form, though it remained unclear how this form is generated [52]. The removal of the cholesterol by ShhN proposed here could contribute to this fraction.

Second, the fact that ShhN could be a peptidase that is switched off by the binding of the Ca²⁺ calcium ion, does not exclude the possible function of calcium ions as modulators of the HhN signaling. The disruption of the calcium binding site by the introduction of single mutations, and the remarkable contribution of Ca²⁺ calcium to strength the interaction between ShhN and binding partners, support this hypothesis. A significant issue to keep in mind is that, the electrostatics of the region involved in the binding of HhN regulators, heavily depends on the presence of calcium. *Drosophila* is one of the few known cases where HhN has no zinc center and the Cdo analog Ihog binds to a different region independently of calcium. The differences observed among vertebrate and invertebrates, could indicate that at the level of the molecular interaction network, both functions, binding of positive regulators and ShhN peptidase activity, may be linked.

**“Education is
the most
powerful
weapon which
you can use
to change the
world.”**

– Nelson Mandela –

CHAPTER 4

MOLECULAR INSIGHTS INTO SONIC HEDGEHOG DIMERIZATION

4.1 Introduction

The signaling activity of HhN proteins is controlled at different levels such as morphogen multimerization, multimer release, spread through tissues and reception at the target cells. It has been suggested that this process is also regulated by heparan sulfate proteoglycans (HSPGs) expression. In this regard, two highly positive HhN regions have been identified in the interaction with negatively charged heparan sulfate chains [46, 111].

In mammals, HhN proteins are mostly released from producing cells as multimeric forms [36]. However, how HhN proteins recognize each other is still under investigation. Therefore, how the HhN multimers form remains an unsolved question. The identification of the different mechanisms that control the formation and routing of Hh represents one of the major challenges nowadays.

In this chapter, assuming that the simplest form of ShhN multimerization is a homodimer, the role of calcium ions on ShhN-ShhN association was investigated by computational tools. Taking as starting point the hypothesis that the presence of the second calcium contributes to the stability of ShhN_{dimer}, the dynamics of the recognition sites in ShhN complexes that have different interface sizes were compared and described at the molecular level. Two hydrophobic residues buried at the dimeric interface might be critical for the formation of a stable homodimer complex by organizing core residues and providing favorable electrostatic interactions between adjacent molecules. Finally, a plausible biologically relevant multimerization model of HhN proteins is proposed based on crystal symmetry analysis and both, experimental and theoretical evidence.

4.2 Computational Methods

Homodimer model: The molecular structure of ShhN_{dimer} model was taken from Ref. [63]. Based on previous experimental studies [64], simulations were performed with and without the N-terminal extension of ShhN (blue in Figure 4.1 and here called "tail"). The tail comprises residues from G25 to K38 (human ShhN nomenclature). As in Chapter 3, three different states with respect to Ca²⁺ binding were studied for both tail (DT) and no tail (DNT) systems. ShhN complexes without calcium (DT0, DNT0) were based on PDB entry 3m1n [243]. Positions of the calcium ions were taken after structural alignment from: ShhN with one Ca²⁺ (DT1, DNT1) based on PDB entry 3n1r [88], and ShhN with two Ca²⁺ (DT2, DNT2) based on PDB entry 3mxw [158]. For convenience, when referring just to the calcium states independently of the complex studied (DT or DNT) they will be named after the number of calcium ions: D0 (0 calcium), D1 (1 calcium) and D2 (2 calciums).

Two point mutations (I48A, V166A) were introduced at the interface between monomeric subunits with the program SCWRL4 [221]. Default parameters were used. The complex of ShhN-ShhN with tail and zero calcium at the binding site (DT0) was taken as a template.

MD simulations: The MD of human ShhN_{dimer} and mutant_{dimers} (I48A, V166A) was simulated with the GROMACS 4.5 package [166]. The GROMOS96 43a1 force field was used with optimized Lennard-Jones parameters $C_6 = 2.0 \cdot 10^{-3}$ kJ/mol nm⁶ and $C_{12} = 1.67 \cdot 10^{-6}$ kJ/mol nm¹² for Ca²⁺ interactions.

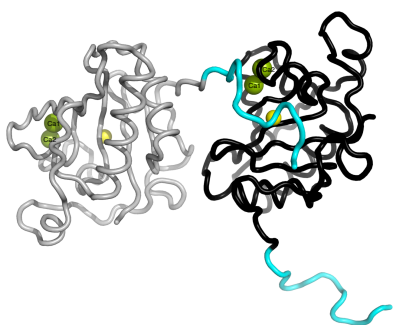


Figure 4.1: ShhN_{dimer} model used. The model is based on PDB entry 3m1n and taken from [64]. Residues G25-K38 composing the N-terminal extension called "tail" are highlighted in cyan. Subunit 1 (light grey) and 2 (black) are depicted as cartoon tubes. Calcium (green) and zinc (yellow) ions are shown as spheres.

Each system was simulated in triplicate using the following protocol. Initial protein structures were solvated in a rhombic dodecahedron box of SPC/E water with a minimum of 1.0 nm distance between protein and faces of box. Residues were assumed to be protonated according to their normal states at pH 7. Na⁺ and Cl⁻ ions were added to neutralize the system at an ionic strength of 0.15 mol/l. The Particle Mesh Ewald method was used to compute electrostatic interactions under periodic boundary conditions. Structures were energy minimized and equilibrated by molecular dynamics for 2 ns. Production simulations were run for 20 ns with a time step of 2 fs. NPT conditions were used with temperature and pressure stabilized at 300 K by V-rescale thermostat and 1 atm by Parinello-Rahman barostat, respectively. Bonds were constrained using the LINCS algorithm. Snapshots of the trajectories

were saved every 100 ps. The last 15 ns of all trajectories were analyzed. RMSD and RMSF were calculated from trajectories for all interface backbone atoms. The interface between subunits was defined as any residue of one ShhN subunit that was in contact with any residue of the adjacent ShhN subunit within a distance of 0.6 nm. For all sta-

tistical analyses R version 2.12 [222] was used.

Buried surface area calculation: Solvent accessible surface areas were calculated from trajectories with the program NACCESS [215]. Default van der Waal radius was used according to [244] and the radius of the water probe was set to 0.14 nm. Subsequently, the buried surface area (the area of the accessible surface on both partners that becomes inaccessible to solvent due to protein-protein contacts) was calculated. This area is the sum of the SASA of the isolated components less that of the complex.

Hydrogen bonds calculation: Polar interactions at dimeric interfaces were determined using the program HBPLUS [217]. Standard geometrical parameters were used and a cut-off of 0.3 nm was applied.

Electrostatic calculations: Electrostatic potentials were computed by solving the non-linear Poisson-Boltzmann equation with APBS [195] using a grid with 0.05 nm spacing, dielectric constants of 79 and 2 outside and inside $ShhN_{dimer}$, respectively, and a water probe of radius 0.14 nm. Temperature was set to 310 K, ionic strength to 0.1 mol/l NaCl outside $ShhN$ and 0 mol/l inside $ShhN$. Zinc and calcium pqr files were generated using a charge of +2 and an ionic radius of 0.088 and 0.114 nm respectively. Charges and radii were assigned to the protein with PDB2PQR v.1.8 [224] using AMBER99 [225] force field parameters.

Binding free energy calculations: The nonlinear Poisson-Boltzmann equation calculations were also used to estimate the electrostatic contributions to the binding (dissociation) free energy between $ShhN$ monomeric subunits as a function of the monovalent salt concentration. The three calcium states were evaluated to estimate the influence of the ion in the stability of multimeric $ShhN$. This binding free energy is calculated in terms of transfer free energies from a homogeneous dielectric environment (where interactions are described by Coulomb's law) to an inhomogeneous dielectric environment with differing internal ($\epsilon = 2$, protein) and external ($\epsilon = 79$, solvent) dielectric constants (Figure 4.2).

The general method divides the binding process into solvation (G_{solv}) and Coulombic (electrostatic) (G_{coul}) components:

$$\Delta G_{bind} = \Delta\Delta G_{solv} + \Delta G_{coul} \quad (4.1)$$

where, the solvation energy contribution to the binding is then given by steps 4 and 2 in Figure 4.2:

$$\Delta\Delta G_{solv} = \Delta G_4 - \Delta G_2 = \Delta G_{solv}(complex) - \Delta G_{solv}(mol1) - \Delta G_{solv}(mol2) \quad (4.2)$$

and the change in Coulombic electrostatic energy upon binding, has the form:

$$\Delta G_{coul} = -\Delta G_1 = G_{coul}(complex) - G_{coul}(mol1) - G_{coul}(mol2) \quad (4.3)$$

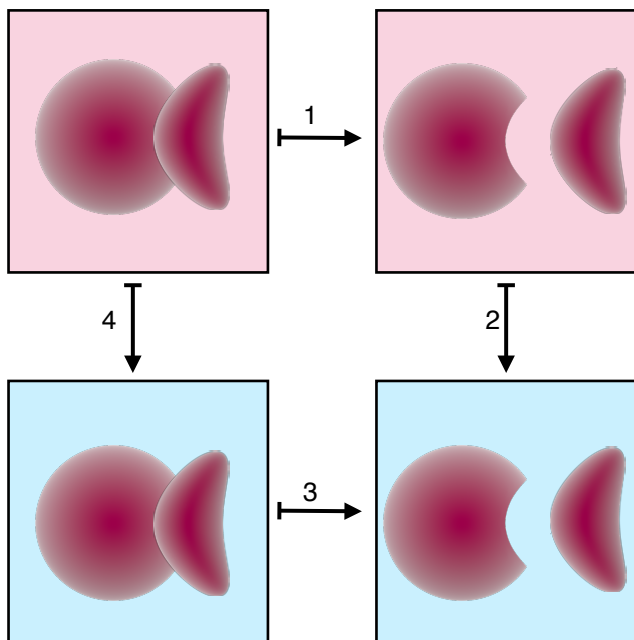


Figure 4.2: Thermodynamic cycle representing the steps of the binding free energy calculation for two solute proteins. Free energies from a homogeneous dielectric environment are transferred to an inhomogeneous dielectric environment with differing internal (red) and external (cyan) dielectric constants. The binding (dissociation) free energy is depicted in step 3.

Coulombic interactions should be calculated using the same uniform dielectric constant as the reference state of the solvation energy above. If the reference dielectric (protein dielectric constant) is different from 1 then, all electrostatic energies need to be divided by the internal dielectric; in this particular case, a dielectric constant of 2.

Matching of molecular structures:

EpitopeMatch [223] was used to identify amongst 5913 calcium binding proteins in the PDB calcium centers that are similar to the calcium center of ShhN. The following nine groups were taken as the query epitope: E90, E91, D96, T126, E127, D130, D132 and two calcium ions. The quality of the matches was evaluated in terms of RMSD and similarity index.

Was used a matching mode in which the two structures are aligned so that a maximum number of chemically identical atoms from both structures was superimposed, and between these the match with the lowest RMSD was chosen.

Inspection of the crystal lattice: PDB entries 1vhh [141], 3m1n [46], 4c4n [111] and 4c4m [111] were inspected with Pymol Molecular Graphics System, version 1.4.1 Schrödinger, LLC. A set of protein coordinates related by crystal symmetry was generated in order to investigate a new biologically relevant mode of multimerization.

4.3 Results and Discussion

Ca²⁺ calcium ion does not enhance ShhN dimerization

The first question addressed in this study was whether the calcium ions stabilize the dimeric ShhN structure. Therefore, the three different calcium states D0, D1 and D2 were examined. Based on this, other aspects related to the recognition sites of ShhN_{dimer}

were considered: the size and chemical character of the protein surface buried at the interfaces; atoms that make contacts across the interface and polar interactions through hydrogen bonds.

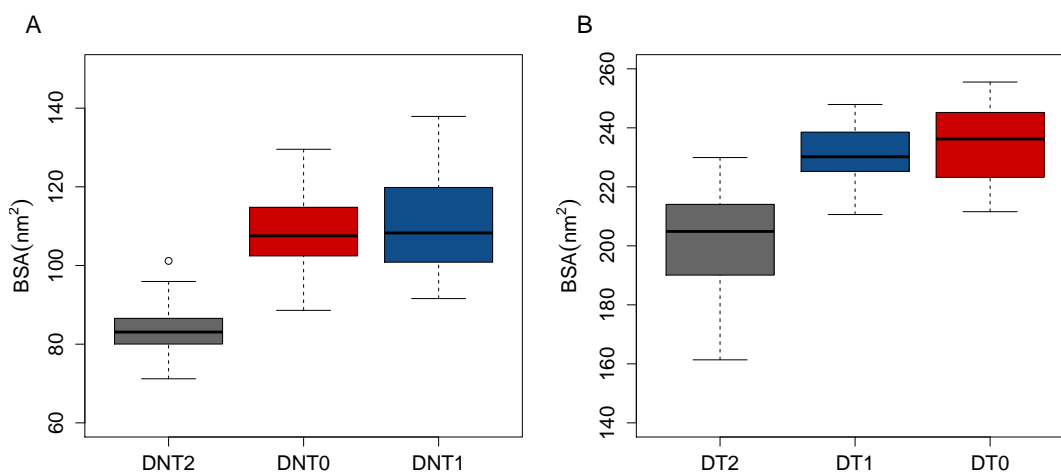


Figure 4.3: Buried surface area of $\text{ShhN}_{\text{dimers}}$. (A) BSA of DNT complexes and (B) DT complexes. A color code according to the number of calcium ions present in the simulated systems is shown: zero calcium (red), one calcium (blue) and two calciums (black). Complexes D2 are significantly different from D0 and D1 respectively at 0.05 significance level (see Tables D1 and D2 in Appendix D). BSA is expressed in nm^2 .

The total buried dimeric interface of DT (tail) and DNT (no-tail) complexes with 0 Ca^{2+} , 1 Ca^{2+} , and 2 Ca^{2+} is shown in Figure 4.3. As was expected, interface sizes of DT and DNT_{dimers} are quite different. DT complexes have a BSA almost 100 nm^2 larger than DNT complexes. Tail residues seem to contribute significantly to the buried area. Surprisingly, the smallest interfaces in both DT and DNT complexes correspond to D2 state. D2 was the only different from other states at a significance level of 0.05 (see Table D1, D2 in Appendix D). The median value of DNT2 complexes is in the range of 81 nm^2 which represents, just 5.6% of the total accessible surface area of the dimer. Although this interface size is small compared to the other states D0 and D1, it still exceeds the minimum size of 50 nm^2 required for physiological protein-protein interfaces [245]. The number of interface residues and the calculated BSAs are linearly correlated with a Pearson's correlation coefficient of $R^2 = 0.93$ (See Figure D3 in Appendix D). As the buried interface area in DNT complexes is significantly smaller than in DT complexes, is not unexpected to find less residues at the interface between the two subunits in DNT_{dimers}.

The chemical composition of ShhN buried protein surfaces is mixed. It comprises non-polar groups that form van der Waals and hydrophobic interactions, and polar groups that form H-bonds and/or salt bridges. While buried atoms are more non-polar for DNT complexes (56%), for DT complexes are more polar (55%). These results are not unexpected if the presence of the highly positive Cardin-Weintraub (CW) sequence is considered in DT complexes.

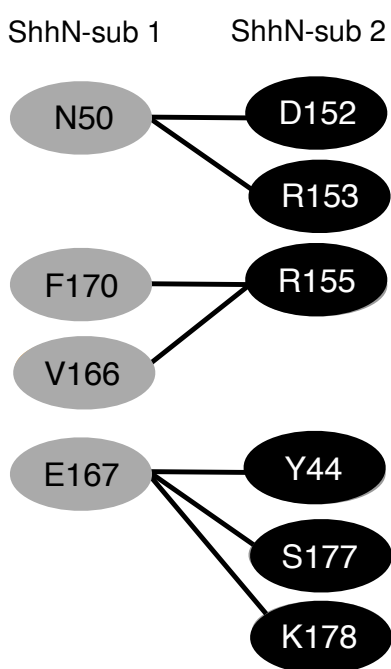


Figure 4.4: Stable hydrogen bonds at ShhN_{dimer} interfaces that are not affected by the presence of calcium. ShhN subunits are colored light grey and black respectively. Residues D152 and R155 have been reported to be involved in Hh multimerization [64]. Hydrogen bonds are shown as black lines.

On average, all interfaces share seven H-bonds (mainly involving side chains atoms) that remain quite stable during dynamics, see Figure 4.4. Two out of ten residues involved in these H-bonds, D152 and R155, have been reported before to contribute to Hh multimerization [64]. For both DT and DNT complexes, the number of H-bonds tends to increase as the number of calcium ions decrease. Interestingly, the formation of complexes that have larger interfaces implicate the interaction of the N-terminal residues including the tail (residues 25-38) and one calcium binding loop defined in Chapter 3 as L_3 (residues 128-139). The more rigid structure of ShhN monomer due to the presence of calcium ions implies less mobility, particularly for calcium binding loops L_2 and L_3 . Therefore, it can be inferred that the less flexible the loops at the interface, the less contacts can be established with the adjacent molecule. Moreover, the contact between the N-terminal and L_3 implicate the electrostatic interaction of a positively charged patch (especially the CW motif) of one subunit and a highly negative area of the other subunit. Thus, it is very likely that this area may be affected by the presence of calcium ions.

Figure 4.5 shows electrostatic potentials calculated for DT complexes in the three states D0, D1 and D2.

The calculation is based on the crystal arrangement proposed by [64]. The area where the interaction between residues in the tail and loop L_3 occurs is enclosed by an ellipse. As it was expected, in D2 state the number of positive charges at the interface strongly increases in comparison with D0 and D1 states.

The removal of calcium ions, however, favors the electrostatic interaction between the positive N-terminal extension (CW motif) and negative residues located in the vicinity of calcium pocket. Although in Figure 4.5 just the DT complexes are shown, the same pattern is observed for DNT complexes. It is important to note that, in DNT complexes the highly positive CW motif is missed but polar contacts are observed in DNT0 and DNT1 states between N-terminal and L_3 calcium loop.

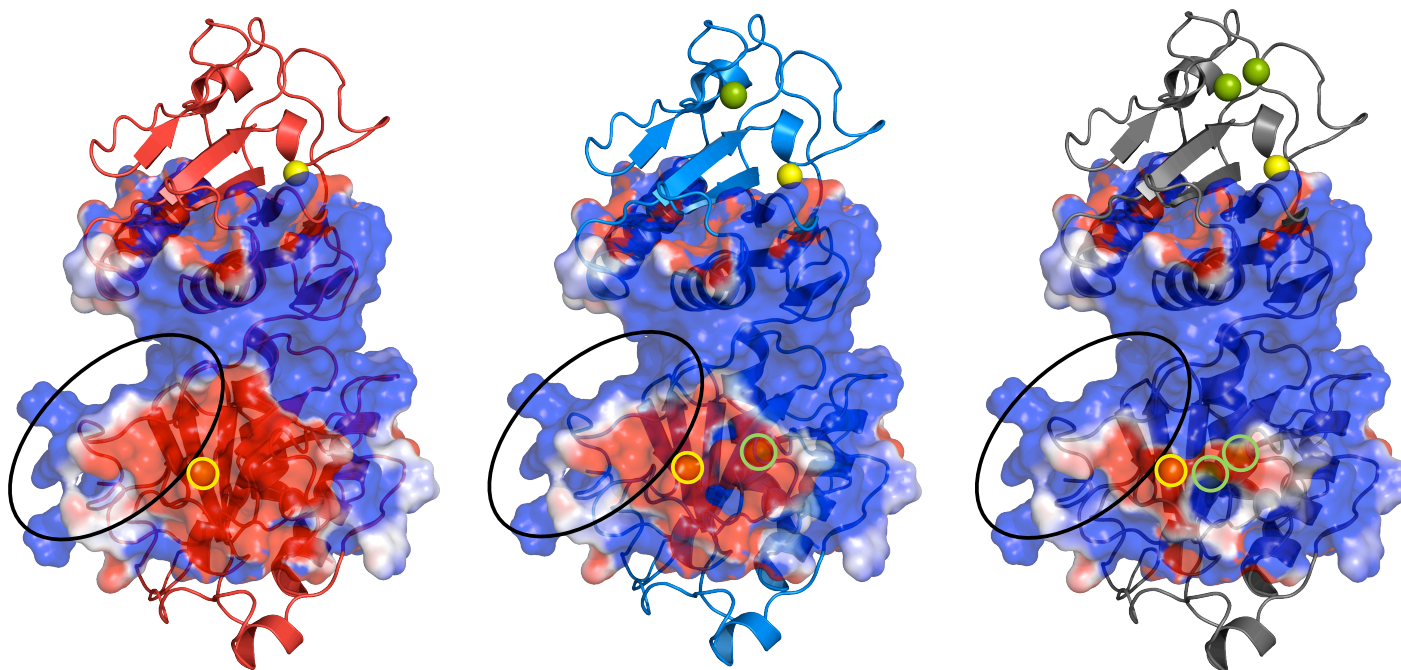


Figure 4.5: Electrostatics of ShhN_{dimers} with zero, one and two calcium ions. From left to right D0, D1, and D2 states. The molecular surface of the proteins is colored according to the electrostatic potential; blue is positive and red is negative. SAS is transparent for a better visibility of the molecular structure. The ellipse encloses the region where the tail (residues 25-38) and loop L₃ interact. Calcium (green) and zinc (yellow) ions are depicted as spheres and highlighted in the lower subunit by green and yellow circles respectively. Electrostatic potentials were scaled to the range of -1 (red) and 1 kT/e (blue).

To further investigate how much the number of calcium ions affect the ShhN dimerization, the electrostatic contribution to the solvation free energies was determined with APBS program [195]. The binding free energies were computed by taking the difference between the free energy of the complex and that of the summation of the ShhN subunits. As shown in Figure 4.6, D0 and D1 states have the most favorable binding free energy. Consistent with the previous findings, D2 state shows the most unfavorable energy values. The presence of the tail strength the binding between ShhN subunits particularly, for DT0 and DT1 complexes where the binding energy increases by a factor of about 3.5 compared to DNT0 and DNT1 complexes. Importantly, all these observations point out that in contrast to the initial hypothesis, the presence of the second calcium (Ca₂; see Figure 4.1) does not contribute to the stability of ShhN_{dimer}, suggesting a weaker complex that seems to have fewer contacts between the ShhN subunits.

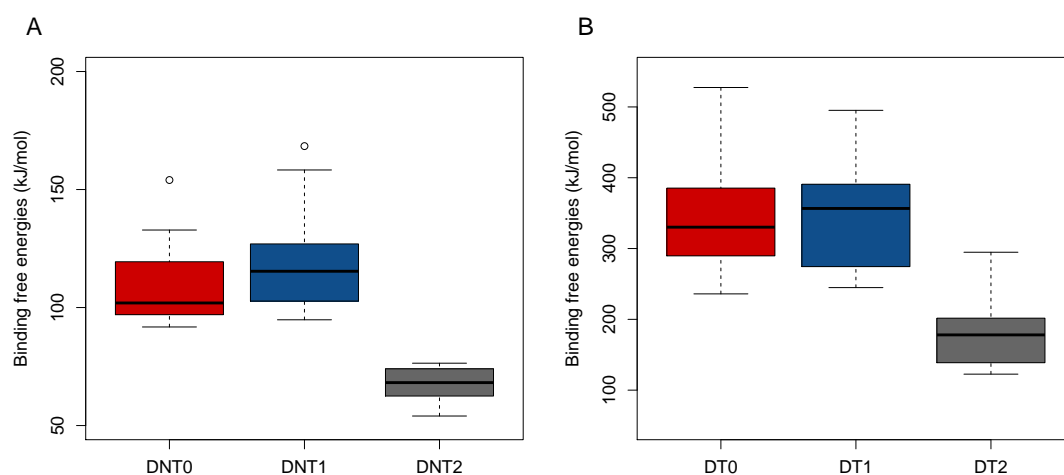


Figure 4.6: Electrostatic contribution to the binding free energies of ShhN_{dimers}. Binding free energies have been calculated for the last 15 ns simulated of DNT complexes (A) and DT complexes (B). In both DNT and DT complexes, two calcium bound dimers exhibited significant differences from one and zero calcium dimers. See Tables D3 and D4 in Appendix D for details.

Diversity of interfaces

To examine in more detail whether the ShhN protein conformation is changed upon association of Ca₂ calcium ion, as well as, the contribution of the CW motif to ShhN multimerization, the recognition sites of all complexes were investigated. Analysis of the trajectories revealed that on average, one feature that all (or almost all) complexes share is that the number of polar and non-bonded interactions decreases as the number of calcium ions bound to ShhN_{dimer} increases.

Unlike D2 state, D0 and D1 states show a similar binding mode in both DT and DNT complexes. In DNT0 and DNT1 complexes, contacts between the two ShhN subunits involve the N-terminal residues (L₃₉-A₄₃, Q₄₆, F₄₇) besides calcium binding loops L₂ and L₃ from the complementarity subunit. DNT2 complexes exhibited a more open conformation with fewer contacts in this area (Figure 4.7). This can be attributed to an electrostatic

repulsion within the whole positive patch observed in Figure 4.5). The same pattern is appreciated in DT complexes but the analysis was not that obvious here because, some residues in the CW motif participate actively in the ShhN dimerization. In DT0 and DT1 complexes, residues R33, R34 and K37 of one subunit were always involved in H-bonds or electrostatic interactions with adjacent residues of L₃ loop of the other subunit, these interactions comprised principally side chains atoms.

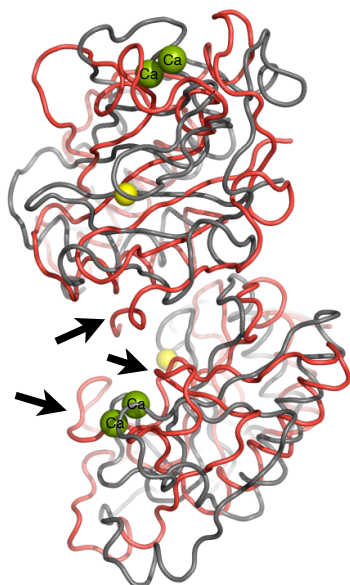


Figure 4.7: Comparison of ShhN_{dimer} in DNT0 and DNT2 complexes. Sample structures close to the cluster centers of DNT0 and DNT2 ensembles were taken from MD simulations. DNT2 dimer (black) shows fewer contacts between ShhN subunits in comparison to DNT0 (red) dimer. Black arrows point out to the N-terminal extension "tail" from one ShhN subunit and L2 and L3 loops of the adjacent ShhN subunit respectively.

Together, these findings suggest that association of Ca²⁺ calcium ion induces a different arrangement of ShhN_{dimer} in comparison to one and zero states. The conformational change that takes place seems to be primarily governed by different small loop movements. Indeed, less contacts were found in DT2 and DNT2 complexes which is consistent with the smaller size of BSA determined (see Figure 4.3). In agreement with experiments reported by Farshi *et al.* [48], from the five positive charged residues that form the CW motif just three, residues R33, R34 and K37 appear to have the major contribution to the ShhN multimerization. Residues K32 and K38 were not involved in any contact between ShhN subunits; which could have two different meanings: one, K32 and K38 may interact with solvent rather than with residues at the dimeric interface and two, they may interact with highly sulfated molecules as heparin or heparan sulfate also implicated in Hh multimerization.

Critical hydrophobic residues at the ShhN dimeric interface

So far, only the contribution of polar residues and calcium ions to ShhN multimerization has been analyzed. However, considering the role that hydrophobic interactions play in protein-protein interactions, the focus lays on the search of non polar residues that

DT0 complex exhibited an extra H-bond between the side chain of H35 and the side chain of E131. Nevertheless, in DT2 complexes only the main chain of residue R34 is engaged in a H-bond with E136; the side chain of R34 was exposed to the solvent most of simulated time. Side chain of R33 showed the same behavior in all three calcium states and was implicated in many bonds. Additional contacts were observed for DT0 and DT1 complexes (in contrast to DT2) between N-terminal residues and the adjacent R123. Furthermore of R123, another residue, K87 have been predicted to contribute to ShhN dimerization [64]. Opposed to experimental reports, the examination of all trajectories showed that residue K87 was not nearly contacting the ShhN neighbor molecule.

Together, these findings suggest that association of Ca²⁺ calcium ion induces a different arrangement of ShhN_{dimer} in comparison to one and zero states. The conformational change that takes place seems to be primarily governed by different small loop movements. Indeed, less contacts were found in DT2 and DNT2 complexes which is consistent with the smaller size of BSA determined (see Figure 4.3). In agreement with experiments reported by Farshi *et al.* [48], from the five positive charged residues

might play a pivotal role in multimerization. To this end, the buried surface area of all hydrophobic residues located at the interface of ShhN_{dimer} was calculated.

Two buried residues, I48 and V166 were identified among all complexes and surprisingly, they are positioned in the center of a hydrogen bond network formed by residues located at the dimeric interface (Figure 4.8). As shown in Table 4.1, I48 and V166 have about 59 and 66%, respectively, of their solvent accessible surface area buried within the dimeric interface. Besides, they have the larger buried surface area found (9.7 nm²) among other hydrophobic residues. While V166 is very well conserved along Hh proteins and was directly involved in a H-bond with R155 (see Figure 4.4; I48 (Shh) is replaced by a serine in Ihh homolog and by a valine in Dhh homolog.

The fact that I48 is replaced by a different residue in each Hh vertebrate homolog could indicate that this residue has little or no importance for Hh biological activity. However, its physical location, hydrophobicity index and buried surface were issues that favored its selection as a potential candidate to affect the ShhN multimerization. To test the contribution of both, I48 and V166 in ShhN dimerization, two single point mutant proteins were generated (I48A and V166A), and further simulated by molecular dynamics.

Comparison of dimeric buried surface areas, as well as, atomic fluctuations and RMSD values at the interfaces revealed that, the introduction of alanine mutations at these specific sites are detrimental for the association of ShhN_{dimers} (see Figure 4.9 and Tables D5–D8 in Appendix D). Mutant_{dimers} exhibit a significantly less buried surface area than the wild type (red in Figure 4.9A). Besides, RMSDs of mutant proteins are about 0.1 nm higher than DTO_{dimer} indicating a less stable structure during dynamics (Figure 4.9B). The flexibility observed for backbone atoms of I48A and V166A mutants, particularly for tail residues (see Figure 4.9C) of one subunit might suggest a rearrangement of this area that is not favorable to yield an stable ShhN_{dimer} . To determine how much each mutation affects ShhN dimerization, binding free energies between the two associating subunits were calculated for the wild type and mutants. The results showed in Figure 4.9D indicate that the binding strength of ShhN_{dimers} was also affected by these mutations yielding a less favorable electrostatic interaction between the ShhN subunits.

Table 4.1: BSA of residues I48 and V166. Residues I48 and V166 are located at the dimeric interface of ShhN. The number in parenthesis is the percentage of buried surface upon dimer formation.

Hydrophobic residues	BSA (nm ²)
I48	9.7 (58.8)
V166	9.67 (65.8)

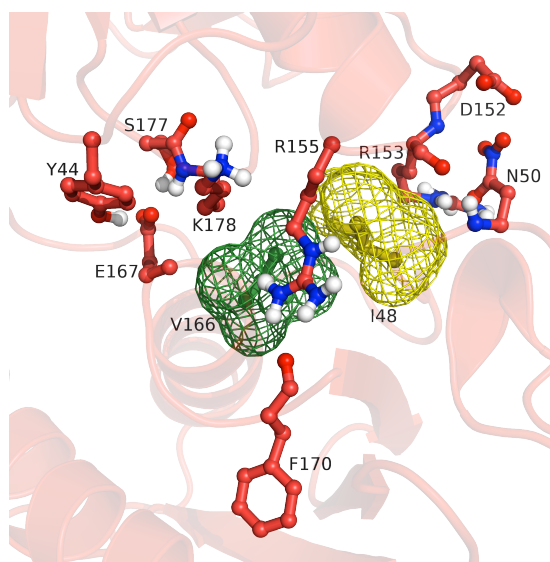


Figure 4.8: Location of hydrophobic residues I48 and V166 at the ShhN dimeric interface. Residues involved in H-bonds (see Figure 4.4) are shown as sticks. I48 (yellow) and V166 (green) are located in the center of the H-bond network that all these residues form.

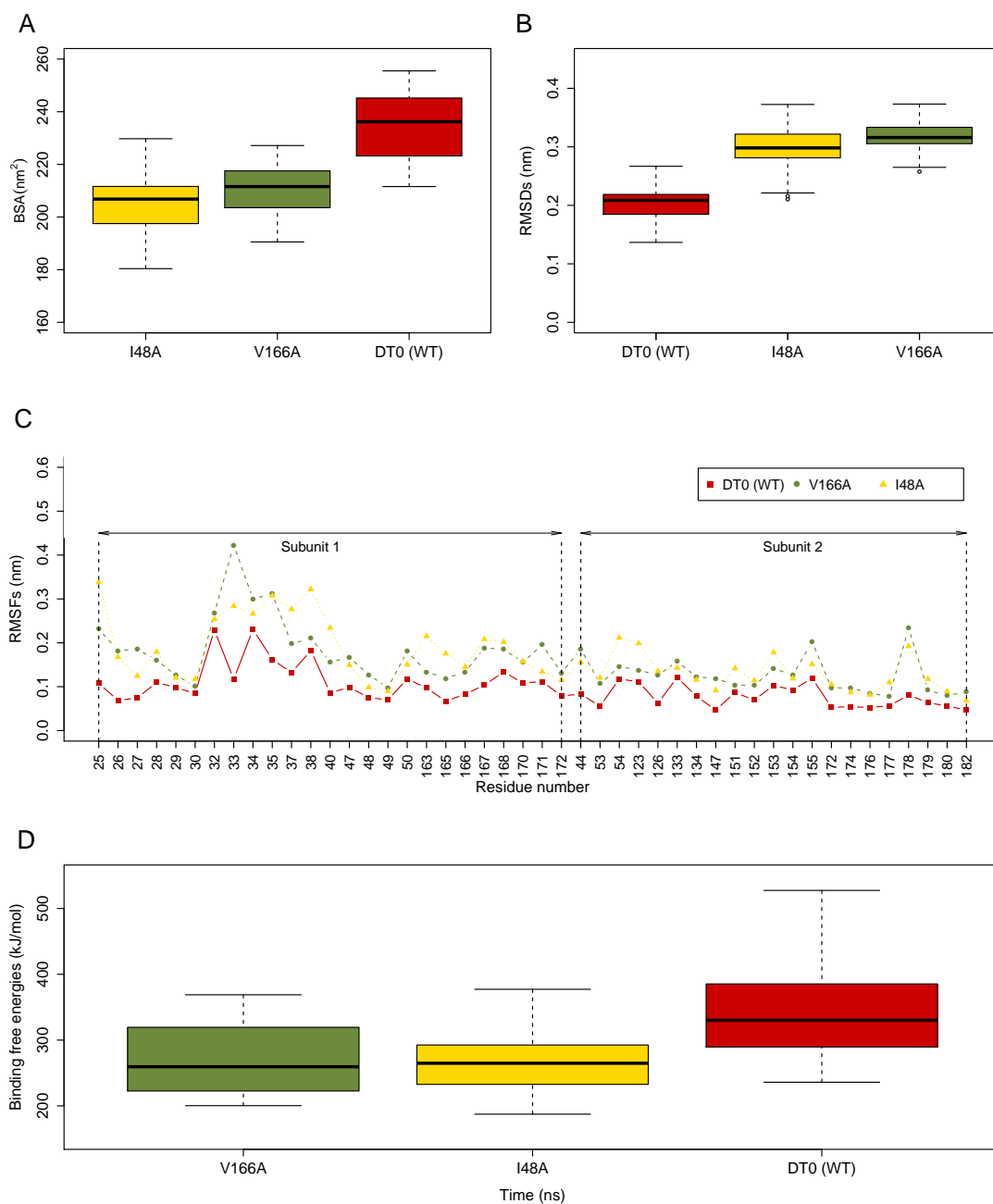


Figure 4.9: Comparison of mutant *dimers* I48A and V166A regarding to the reference DT0 complex. Different computational tools were considered in the analysis (A) Buried surface area between the subunits, (B) and (C) RMSD and RMSF respectively, of backbone atom residues at the interface. (D) electrostatic contribution to the binding free energy of dimers. The interface between the subunits was estimated as any residue of one ShhN subunit in contact with the adjacent ShhN subunit at a distance less than 0.6 nm. See Tables D5–D8 in Appendix D for all statistical details.

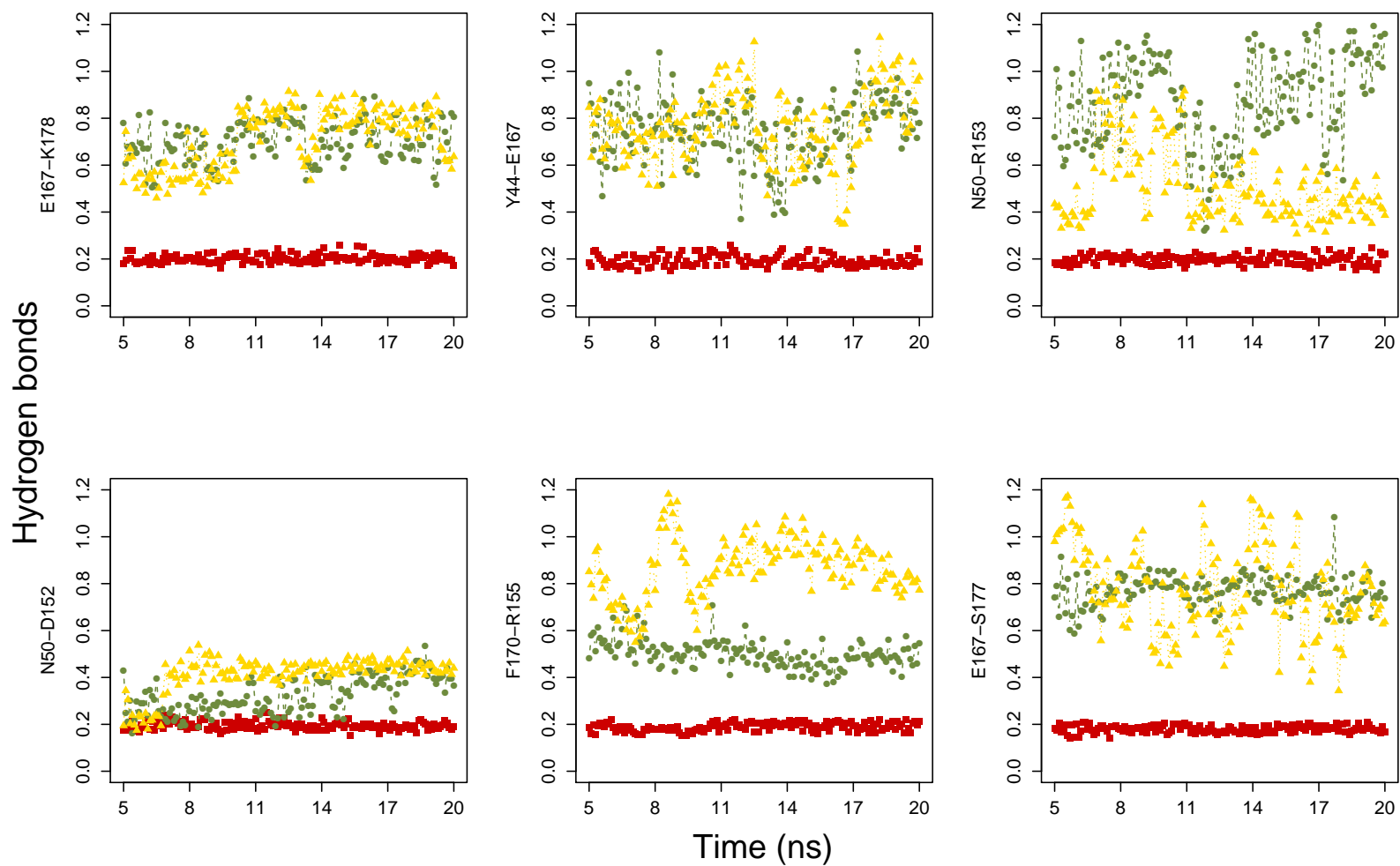


Figure 4.10: Conservation of H-bonds in MD simulations of DT0, I48A and V166A. Hydrogen bonds (measured in nm) of the wild type (red), I48A (yellow) and V166A (green) were measured during the last 15 ns of the simulations. All H-bonds are disrupted in ShhN mutants. The hydrophobic side chains of residues I48 and V166 appear to be required to stabilize this H-bond network between the ShhN subunits. See Tables D9–D14 in Appendix D for all statistical details.

Additionally, the conservation of hydrogen bonds in the mutant dimers was analyzed and compared to the wild type. Figure 4.10 shows how all the six H-bonds studied are significantly different in I48A and V166A mutants with respect to DT0 complex (wild type). An alignment of clustered structures taken from trajectories corroborate these findings. H-bonds of residue N50 with D152 and R153 respectively, and the one between F170 and R155 also show significant differences among the mutants (from left to right in Figure 4.10: third plot in the first row and first and second plots in the second row, respectively). See Tables D9–D14 in Appendix D for all statistical details.

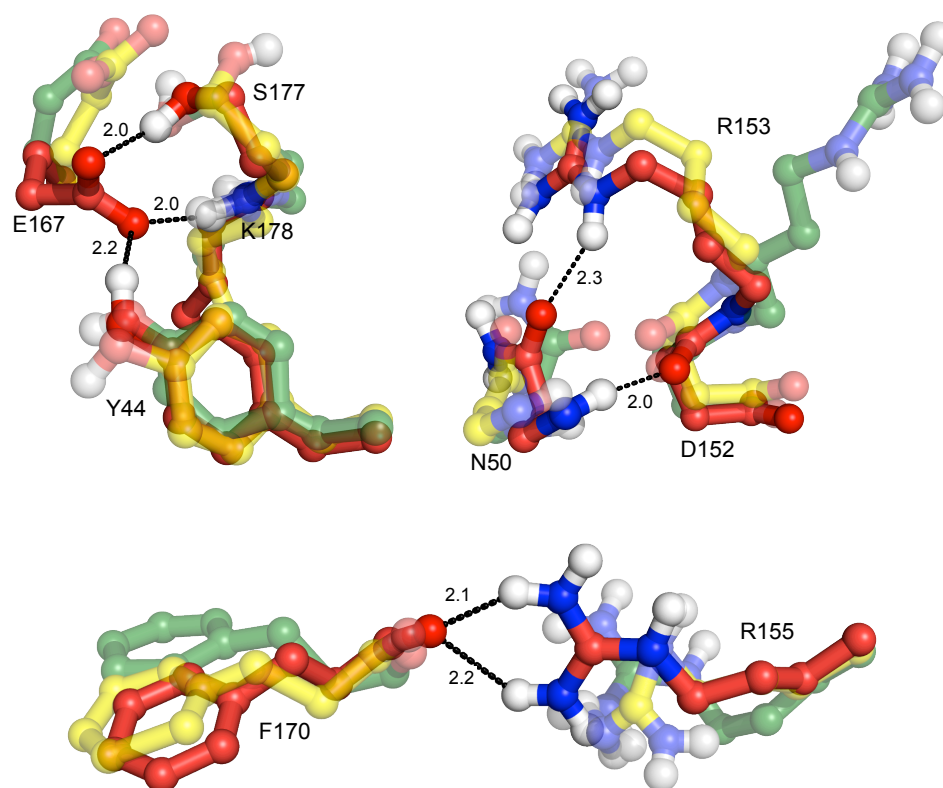


Figure 4.11: Aligned clustered structures taken from MD simulations of DT0 (red), I48A (yellow) and V166A (green) respectively. Displacement of the residues involved in different hydrogen bonds at the dimeric interface of I48A and V166A complexes can be appreciated. Such a change in the spatial conformation leads to the loss of all hydrogen bonds observed in the wild type. Residues are shown in ball and stick representation. Numbers at the dashed lines are distances in 0.1 nm. To ensure a better comprehension the rest of the protein was removed.

As can be noticed in Figure 4.11 the side chain of residue E167 adopts in ShhN mutants a spatial conformation completely different from DT0 complex which lead to a total loss of two hydrogen bonds (E167-Y44 and E167-K178). The same is applied to residues N50, R153 and R155 as is also shown in Figure 4.11. These observations point out that hydrophobic residues may promote ShhN recognition indirectly, by stabilizing the hydrogen bond network.

Overall, the evidence suggest that mutation of key hydrophobic residues buried in the dimeric interface of ShhN greatly compromises the affinity between the two subunits. In absence of either I48 or V166, hydrogen bonds and electrostatic interactions become disorganized and disrupted in the interface. I48 and V166 appear to promote the interaction of adjacent critical residues, favoring in this way the formation of ShhN_{dimers}. I48 is an interesting residue due to the lower degree of conservation in Hh sequences. In principle, it may not classify as a hot spot residue. Yet, the opposite is shown. In general, no significant differences were found between I48A and V166A in the computational estimations, except for the RMSD calculations and three H-bonds mentioned above. Nevertheless, the equal or unequal contribution of each residue to ShhN dimerization can be easily determined with experiments.

Ca²⁺ may promote heparan sulfate binding

In contrast to our initial hypothesis that calcium might stabilize the ShhN_{dimer} and enhance multimerization, so far, all evidence suggest that the binding of Ca²⁺ has a negative effect on ShhN dimerization. Nevertheless, if the role that calcium ions play in the formation of large aggregates in other biological systems is considered [246, 247] the question then arises of how similar is the calcium binding site of ShhN to other calcium binding proteins involved in analogous processes. To answer this question, a match between all available calcium binding proteins in the PDB (5.913 structures, last access 2012) and nine ShhN groups (E90, E91, D96, T126, E127, D130, D132 and two calcium ions) was carried out.

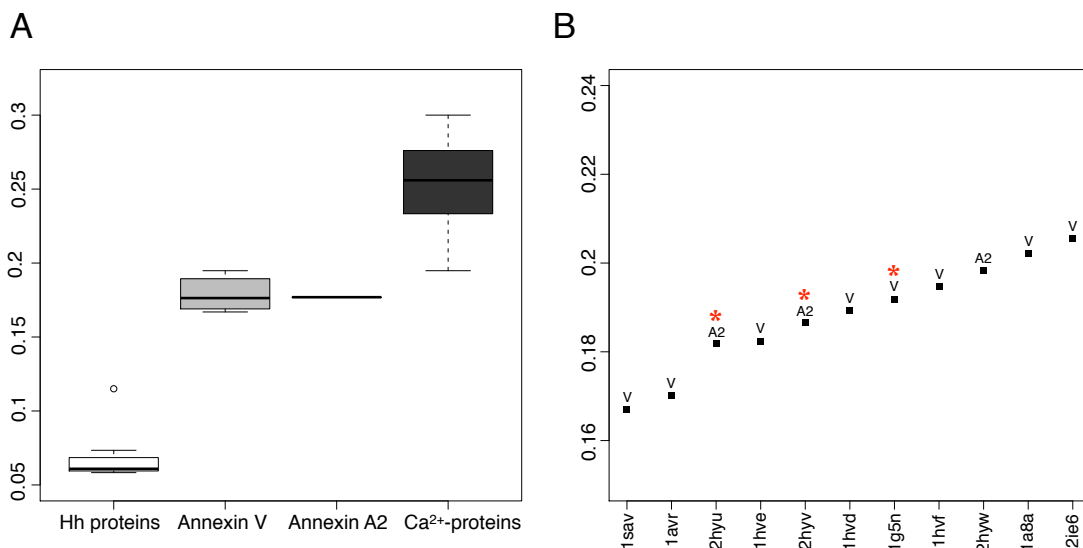


Figure 4.12: RMSDs between ShhN and calcium binding proteins. (A) RMSDs between calcium centers from all available calcium binding proteins in the PDB and ShhN double calcium center. (B) RMSDs between annexin proteins family and ShhN double calcium center. The red stars highlighted X-ray structures that has been crystallized bound to heparin. Calculations were carried out using EpitopeMatch [223].

Three clusters can be clearly distinguished in Figure 4.12A. As expected, the lowest RMSDs values (0.05-0.13 nm) correspond to all Hh family proteins (white boxplot). Surprisingly, the second cluster (grey boxplot) is only composed of annexin proteins V and A2 while the third cluster represents the rest of calcium binding proteins tested. Annexins are calcium and membrane binding proteins that have been implicated in a wide range of functions in eukaryotes including signaling [248]. Besides, they are known to bind glycosaminoglycans (GAGs) in a calcium dependent fashion [248, 249]. While in annexin A2 the calcium ions contact directly the heparin, annexin V does not show the same pattern. The calcium ions induce in annexin V the required loop conformation for heparin binding.

To corroborate the findings described above, a second match was carried out between the nine groups of ShhN and only annexin proteins (73 structures, last access in 2012). Consistent with the previous results, Figure 4.12B shows that among the whole family of Annexins the more similar to Shh are annexins V and A2. The red stars in Figure 4.12B stand for annexin proteins that have been crystallized bound to heparin [248, 249]. It is worth mentioning that outside the calcium binding site, ShhN and annexin proteins are quite different thus the similarities found between the calcium binding site of annexin proteins and ShhN were unexpected.

It is well known that Hh proteins bind GAG molecules, particularly heparan sulfate (HS) or heparin. This interaction appears to facilitate Hh signaling cascades at various levels, including secretion, retention and stabilization of binding partners [105, 106]. A recent study [111] has shown that interaction between ShhN and heparin takes place through two highly positive sites on ShhN. These two sites are: the already known Cardin-Weintraub sequence (residues K32-K38) and another positive patch of residues K87, R123, R153, R155 and K178 (not previously identified). The last one is in very close proximity to the calcium binding site and overlap with many of the residues identified in the Shh_{dimer} model used here.

Based on these new evidence, the arrangement of the ShhN_{dimer} model was analyzed again but this time the two binding sites for GAG molecules were considered. As shown in Figure 4.13A the ShhN_{dimer} model used in this study provides a larger positive patch recreating the ideal scenario for the GAG binding. These two sites are in close proximity in the ShhN homodimer and the positive charge of this area is increased by the presence of both calcium ions which are located at the dimeric interface. Both sites contain five residues but whereas the CW motif has three lysines and two arginines, the recently discovered GAG binding is composed of three arginines and two lysines. Arginine has been shown to bind 2.5 times more tightly than lysine [250]. The guanidino group in arginine forms more stable H-bonds as well as stronger electrostatic interactions with sulfate groups [100]. Although the lack of experimental or theoretical reports about Hh-GAG interactions can not support this argument, it will be very interesting to determine whether the binding strength of GAG molecules is similar or not in these two binding sites.

The sulfate ions shown in Figure 4.13A can be used to predict binding sites for GAGs [251, 252]. These sulfate ions are usually originated from the crystallization buffer found in the crystal structures of proteins. To corroborate the possible binding of GAGs to this large positive patch, the X-ray structure of PDB entry 4c4m [111] was aligned to one of

the subunits of ShhN-ShhN homodimer with an RMSD of 0.025 nm. Figure 4.13B shows how the chondroitin sulfate (CS) coming from 4c4m structure perfectly match this area. Besides, the open conformation observed for D2 complexes (see Figure 4.7) is consistent with these findings indicating that a GAG molecule can be accommodated in this space and contact some residues of the large positive patch between subunits (as shown in Figure 4.13B).

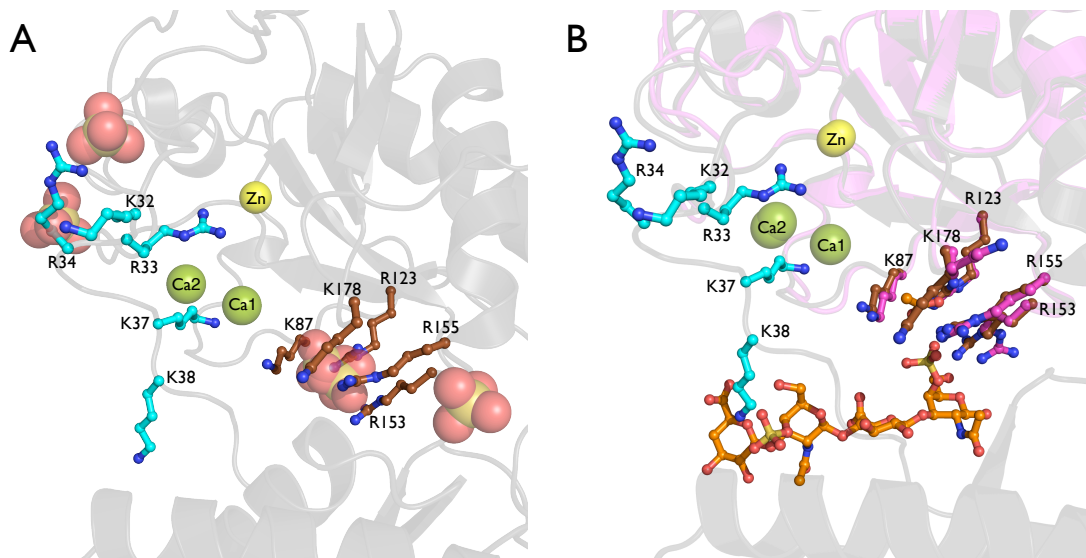


Figure 4.13: Structural analysis of ShhN-GAG interactions. (A) The two binding sites proposed for GAG molecules are in close proximity in the ShhN homodimer arrangement investigated in this study. Sulfate ions shown as spheres indicate possible GAG binding. (B) The same view as in (A) including the X-ray structure of 4c4m [111] depicted in magenta and aligned to one of ShhN subunits (RMSD of 0.025 nm). A chondroitin sulfate molecule from 4c4m (in orange sticks) matches the position of some sulfate ions shown in (A). Residues within the CW motif are colored in brown while residues forming the new GAG binding site are colored in cyan (ShhN homodimer) and in magenta (PDB entry 4c4m).

In addition to this structural analysis, the effect of calcium ions in the binding of heparin molecules to ShhN has been tested by computational tools [253]. The author has investigated the three calcium states of ShhN (0, 1 and 2 bound calciums) and has found that in the presence of Ca2 calcium ion, heparin shows a higher affinity for ShhN. These results are in agreement with all evidence previously discussed and all together suggest that Ca2 calcium ion may influence the specificity and/or the affinity of binding GAG molecules for Hh proteins. In this way, a more stable and compact ShhN multimer can be formed generating a concentration gradient required for the long-range spread.

A biologically relevant mode of ShhN multimerization

Until now, the interactions between ShhN-ShhN have been studied using the model proposed by [63] based on PDB entry 3m1n [243]. This arrangement was suggested as a possible mode of multimerization for Hh proteins where the N-terminal extension has a pivotal role. At first, the authors dismissed another contact observed in this X-ray structure within one unit cell due to the small buried surface area between ShhN molecules (less than 50 nm²). Besides, the lack of the N-terminal region in other Hh crystal structures made impossible to find any contact areas exceeding 50 nm². In this section, a similar crystal symmetry analysis was performed considering three new X-ray structures and the calcium free molecule from murine ShhN .

A set of Hh protein structures (PDB entries 1vhh, 3n1r, 4c4m and 4c4n) in the presence or absence of calcium was analyzed to investigate whether such a ShhN-ShhN arrangement (see Figure 4.1) is conserved in other crystal lattices and the possible existence of new contacts between ShhN molecules.

Analysis of the symmetry contacts in the PDB entry 1vhh revealed a ShhN-ShhN interface which appeared multiple times in the crystal lattice. The C-terminus of one Shh subunit inserts into the binding cleft of an adjacent subunit, and has a buried surface area of about 140 nm². An interesting point is that this arrangement is completely independent from the one proposed by [63]. Both arrangements are physiologically relevant exceeding a buried surface area of 100 nm². But nevertheless, on the one hand the N-terminus extension (residues G25-K38) of one ShhN molecule is extended in an area nearby the zinc center of the partner molecule (3m1n) and, on the other hand, is the C-terminus which contacts the zinc (1vhh). While the crystal contacts of 3n1r did not show any relevant interaction, 4c4n exhibited a ShhN dimeric structure that appears to be held together by the heparin chain.

Unexpectedly, in the crystal lattice of PDB entry 4c4m similar contacts to the ones observed in 1vhh and 3m1n respectively were found multiple times. Figure 4.14 summarizes these findings. The ShhN monomer crystallized in the 4c4m structure is depicted in magenta and was situated at the center of the coordinates axis (Figure 4.14A) where each axis represents a different interaction mode. In Figure 4.14B, an arrangement similar (97.5%) to the displayed in 1vhh is shown together with its surface electrostatic potential. The chondroitin sulfate was extended using all molecules obtained by the same crystal symmetry operation. Note that CS lies in the dimeric interface of an arrangement similar (90.4%) to the observed in 3m1n (Figure 4.14C). All residues involved in ShhN multimerization [46, 47, 63] have been identified in the crystal lattice arrangement shown in Figure 4.14A. Residues E72 and R73 (red square) [47] are located in the opposite face of the dimeric interfaces identified in 1vhh and 3m1n structures respectively. Protein-protein contacts between these three adjacent ShhN molecules are shown in Figure 4.14D.

All in all, a new model of ShhN multimerization has been proposed in this study (Figure 4.15) not only based on crystal structures but considering the dynamic properties of ShhN proteins investigated by MD simulations. The basic idea behind the model is that dually lipidated ShhN monomers produced in the endoplasmic reticulum are self associated at the outer plasma membrane in such a way that, the palmitic acid and cholesterol

modifications from adjacent molecules are embedded into the lipid bilayer. In agreement with experimental data [68, 71], cholesterol modification might be sufficient to allow Hh dimerization yielding soluble cell-surface clusters. To complete the picture, these clusters can interact with heparin/heparan sulfate chains of glypican which enables them to form compact larger clusters or visible nanoscale clusters [47] that can be further released from the plasma membrane. The size of these big clusters can be determined by the length of the GAG molecules. Once in the extracellular space, according to what is described in the literature [36, 46], this large multimer of ShhN may diffuse far from its site of synthesis to engage directly in long range signaling. The change in the pH levels and calcium concentrations can regulate the peptidase activity of ShhN demonstrated in Chapter 3 leading to the formation of different concentration gradients.

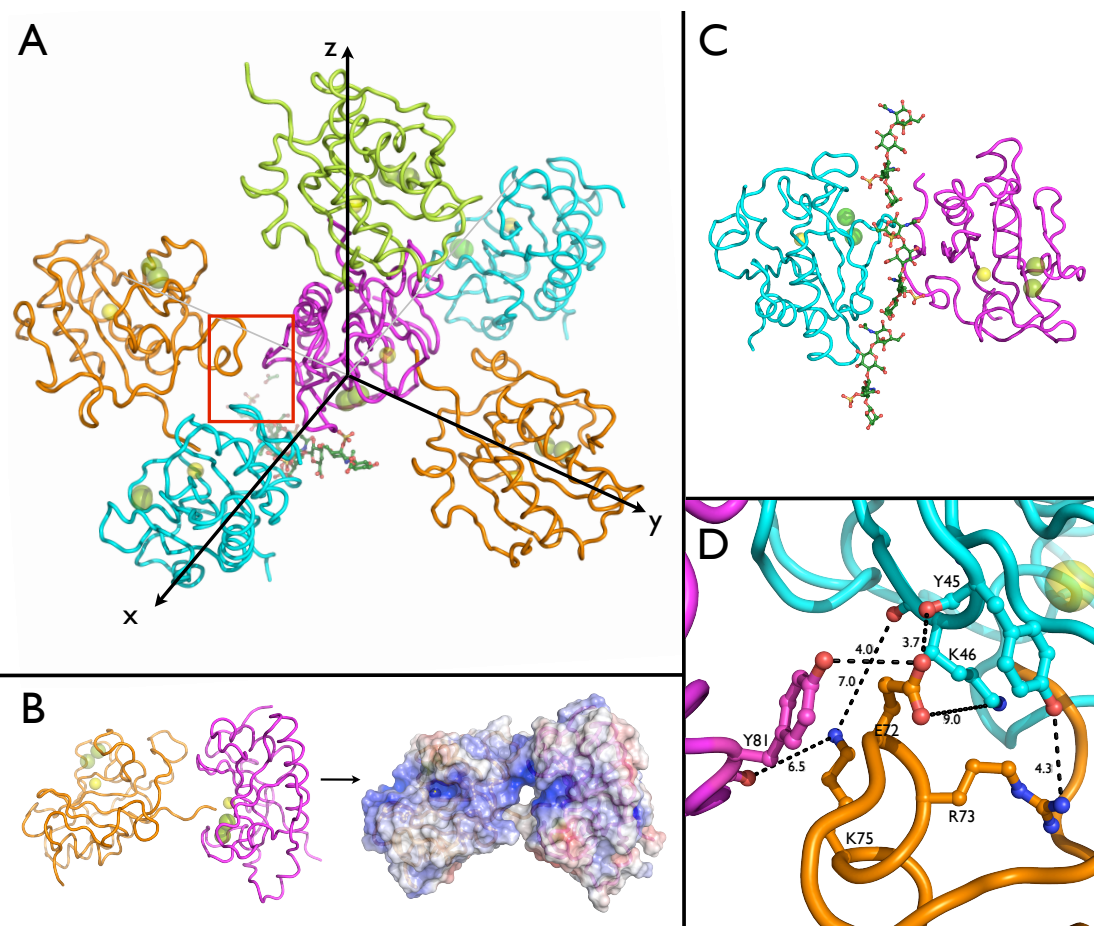


Figure 4.14: Analysis of the crystal lattice contacts of X-ray structure 4c4m. (A) Each axis represents a different interaction mode of ShhN monomers. The red rectangle encloses the area where E72, R73 and K75 residues from one ShhN molecule (orange) interact with Y81 (magenta), Y45 and K76 (cyan) residues of adjacent ShhN molecule. (B) and (C) represent two different arrangements of ShhN molecules. The former is determined by the contact of the C-terminal extension of one ShhN molecule (orange) which lies nearby the zinc cleft from the adjacent ShhN molecule (magenta). The molecular surface was colored according to its electrostatic potential, blue is positive and red is negative, and scaled to the range of -5 (red) and 5 kT/e (blue). In contrast to (B), the arrangement observed in (C) is promoted by GAG molecules. In this particular case, chondroitin sulfate is shown as sticks. (D) Zoom of the red rectangle showed in (A). Numbers at the dashed lines are distances in 0.1 nm.

An element to keep in mind regarding this model is that the way in which ShhN and HS interact may change. It is known that the variation in the degree of sulfation can result in different binding modes of HS or heparin with proteins that affect their activity [100]. For instance, the arrangement of the ShhN-heparin structure (4c4n) differs from the one observed in ShhN-CS.

It is important to note that lack of X-ray structure available where either both lipid modifications (a cholesterol attached to the C-terminus and a palmitate added to the N-terminus) or the complete C and N-terminus sequences appear in the crystal derogates any attempt to test the proposed model. However, all evidence collected in this study (with the accessible data) point to this direction. Further experimental tests are needed.

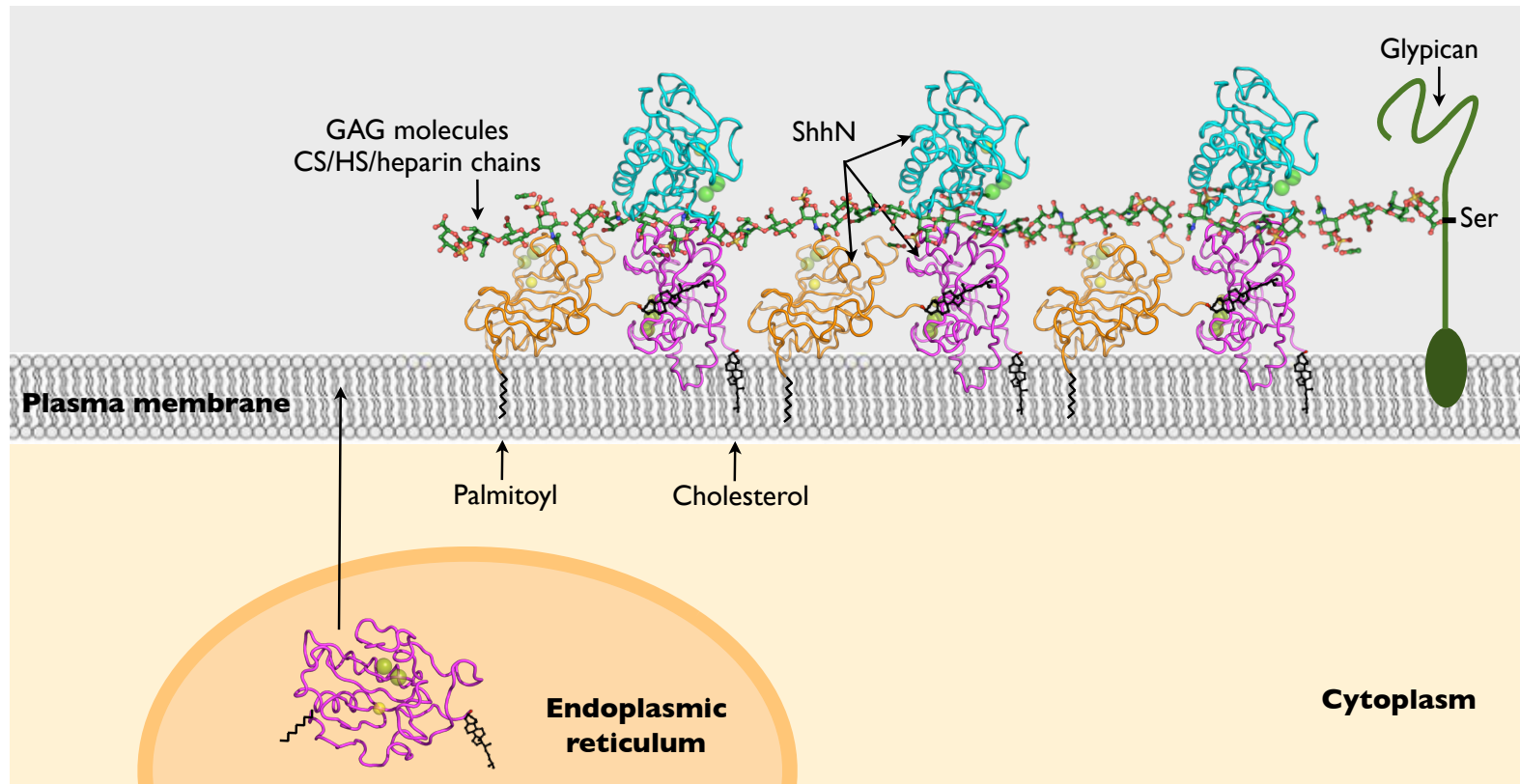


Figure 4.15: Schematic representation of the proposed model for ShhN multimerization. Dually lipidated ShhN monomer produced in the endoplasmic reticulum is self associated at the outer plasma membrane in such a way that the palmitic acid and cholesterol modifications from adjacent molecules are embedded into the lipid bilayer (orange and magenta ShhN monomers). The interaction with GAG chains enables ShhN to establish another way of multimerization (magenta and cyan ShhN monomers) forming compact larger clusters that can be further released from the plasma membrane. For simplicity the model has been represented in a two-dimensional space (a plane) but most likely all monomers are anchored in the membrane. Cholesterol, palmitic acid and GAG molecules are shown as sticks. Calcium (green) and zinc (yellow) ions are displayed as spheres. This scale model has been proposed based on the crystal symmetry analysis showed in Figure 4.14 and evidence from MD simulations.

4.4 Conclusions

Overall, the results described in this chapter are in agreement with experiments [46, 63] where highly conserved residues implicated in the multimerization process of ShhN proteins have been identified. Two not previously identified hydrophobic residues (I48 and V166) which are buried at the interface of ShhN_{dimers} are critical for the maintenance of the complex.

The simple model used to explore the intermolecular interactions of ShhN_{dimers} revealed that the Ca²⁺ calcium ion does not strengthen the affinity of adjacent molecules, contrary to what was initially thought. Although the nature of the interfaces between the ShhN monomers is diverse, non-covalent interactions play a dominant role. Either hydrogen bonds or hydrophobic contacts within adjacent molecules stabilize the ShhN-ShhN interfaces. The contribution of electrostatic interactions is also important, but this is not surprising because both metal centers zinc and calcium are located in the vicinity of the dimeric interfaces.

The open conformation observed in the complexes with two calcium ions (which increases the positive charge at the interfaces) might be required for the interaction of ShhN with GAG molecules as heparan sulfate. Taking into account the nature of the double-calcium center, most likely the Ca²⁺ ions will not participate directly in the interaction with GAG molecules, but the role of calcium ions in the binding of GAG molecules to ShhN monomers or multimers needs to be experimentally tested. The question arises then whether different interaction modes of ShhN-ShhN are given by the variation in the degree of sulfation of heparan sulfate chains (or heparin). This fact can obviously influence the way that ShhN multimerizes and forms compact and very stable multimers. Given the similarity between Hh proteins and other signaling molecules as Wnt which also harbor two lipid moieties, oligomerization might be a common mechanism for long-range signaling among morphogens.

The analysis of crystal lattice contacts of various ShhN proteins led us to propose a multimerization model where different levels of interaction can control the way that multimers form. Despite the lack of experimental evidence that corroborate our model, it can be used to guide further experiments to uncover the details of Hh oligomerization.

**“I am a traveller
to all **parts**, and a
newcomer to **none**:
I am **art** among
the arts, **with** the
mountains, I am
one.”**

– José Martí –

CHAPTER 5

OUTLOOK AND REMAINING CHALLENGES

Understanding the role that metal ions play in proteins is a key issue in biological processes. The activity of a wide range of biomolecules is mediated and regulated by metal ion binding. Despite that, the implication that zinc and calcium ions have in the Hedgehog signaling pathway is not well established and did not get much attention.

Based on our results, Sonic hedgehog proteins could be zinc exopeptidases regulated by calcium. The peptidase function should be switched off if both calcium binding pockets are occupied. A strong pH dependence of protease activity was confirmed by *in vitro* experiments. Although, the peptidase substrate has still to be determined the autodegradation could be an elegant mechanism to tune morphogen gradients. While the N-terminal of ShhN proteins (residues G25-K38) enhances ShhN dimerization, the binding of calcium ions disfavors the electrostatic interaction between ShhN subunits as a result of an increment of positive charges in this region. Nevertheless, this fact can promote the interaction of ShhN dimers with heparan sulfate chains (a key player for HhN function) and enhances the multimerization process. Results built on mutations of buried residues at the ShhN dimeric interface demonstrate the importance of hydrophobic residues for the maintenance of a stable complex. Mutations within the calcium binding site can affect the ShhN multimerization at the molecular level and thus, alter the ShhN signaling.

All the collected evidence over the last decades have enabled to broaden our knowledge about Hedgehog signaling pathway. But, *what are the future challenges?*

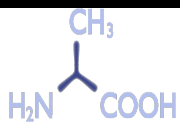
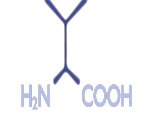

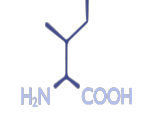
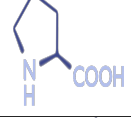
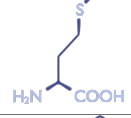
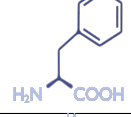
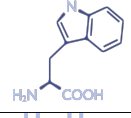

In general, further research has to include some aspects, such as: (1) the regulation of the HhN self-assembly and whether different mechanisms of release HhN from the plasma membrane are tissue dependent or independent, (2) the role of GAG molecules in the HhN oligomerization process and their influence in the formation of a concentration gradient, (3) the way by which Ptc regulates Smo, and (4) the study of the multivalent character of Hh ligands *in vivo*. The emergence of new experimental techniques together with robust and faster computational methods will help to accomplish these challenges.

Finally, the work described in this thesis opens a new direction of research that was

overlooked in the past. The identification of ShhN as a peptidase makes possible to test different hypotheses about this morphogen which might explain some of the still unsolved questions. Similar and/or additional studies need to be performed to characterize potential substrates of this metalloprotease. Despite the similarity between vertebrate homologs, *lhh* and *Dhh* hedgehogs should be included in further investigations.

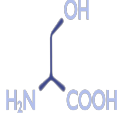
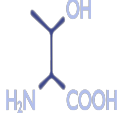
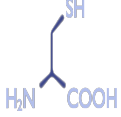
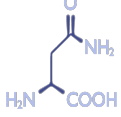
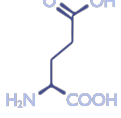
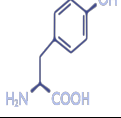
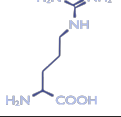
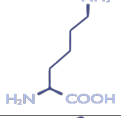
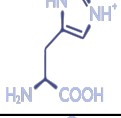
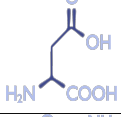
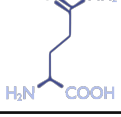
Appendix A

Table A1: Chemical structure of amino acids. One- and three-Letter codes. All twenty amino acids are shown. Amino acids are colored according to their chemical properties: polar-uncharged (G, S, T, Y, C, Q, N) green, basic (K, R, H) blue, acidic (D, E) red and nonpolar (A, V, L, I, P, W, F, M) darkgray.

Chemical name	3-Letter amino acid code	1-Letter amino acid code	Chemical structure
Alanine	Ala	A	
Valine	Val	V	
Leucine	Leu	L	
Isoleucine	Ile	I	
Proline	Pro	P	
Methionine	Met	M	
Phenylalanine	Phe	F	
Tryptophan	Trp	W	
Glycine	Gly	G	

Continued on next page

Table A1 – Continued from previous page

Chemical name	3-Letter amino acid code	1-Letter amino acid code	Chemical structure
Serine	Ser	S	
Threonine	Thr	T	
Cysteine	Cys	C	
Asparagine	Asn	N	
Glutamine	Gln	Q	
Tyrosine	Tyr	Y	
Arginine	Arg	R	
Lysine	Lys	K	
Histidine	His	H	
Aspartate	Asp	D	
Glutamate	Glu	E	

Appendix B

Table B1: Crystal structures of Hh proteins available in the Protein Data Bank. In all X-ray structures except one (1ato; apricot) the signaling domain of Hh (HhN) was crystalized. Invertebrate Hh is colored in gold and does not have Zn^{2+} . Dhh is shown in blue and Shh in purple/pink regarding the organism mouse/human respectively. Ihh is colored in yellow and Ihh mutants are highlighted in green. Complexes from vertebrate homologs (Shh, Ihh and Dhh) contain Zn^{2+} and 2 Ca^{2+} otherwise is specified.

PDB code	Complex	Organism
1ato	HhC	Fruit Fly
2ibg	HhN-IhogFn3	Fruit Fly
1vhh	ShhN (no Ca^{2+})	Mouse
3n1r	ShhN (1 Ca^{2+})	Mouse
3d1m	ShhN-Cdo	Mouse
2wfx	ShhN-Hhip	Mouse
2gw4	ShhN-Hhip	Mouse
4c4n	ShhN-Heparin	Mouse
4c4m	ShhN-CS4	Mouse
3m1n	ShhN (no Ca^{2+})	Human
3mwx	ShhN-5e1	Human
3ho5	ShhN-Hhip	Human
3n1f	IhhN-CdoFn3	Human
3n1m	IhhN-BocFn3	Human
3n1o	IhhN	Human
3n1p	IhhN-BocFn3	Human
3k7g	IhhN (no Ca^{2+})	Human
3k7h	IhhN (E95K) (no Ca^{2+})	Human
3k7i	IhhN (E131K) (no Ca^{2+})	Human
3k7j	IhhN (D100E) (no Ca^{2+})	Human
2wg3	DhhN-Hhip (no Ca^{2+})	Human
2wfr	DhhN	Human
2wfq	DhhN (no Ca^{2+})	Human
3n1g	DhhN-BocFn3	Human
3n1q	DhhN-CdoFn3	Human

Appendix C

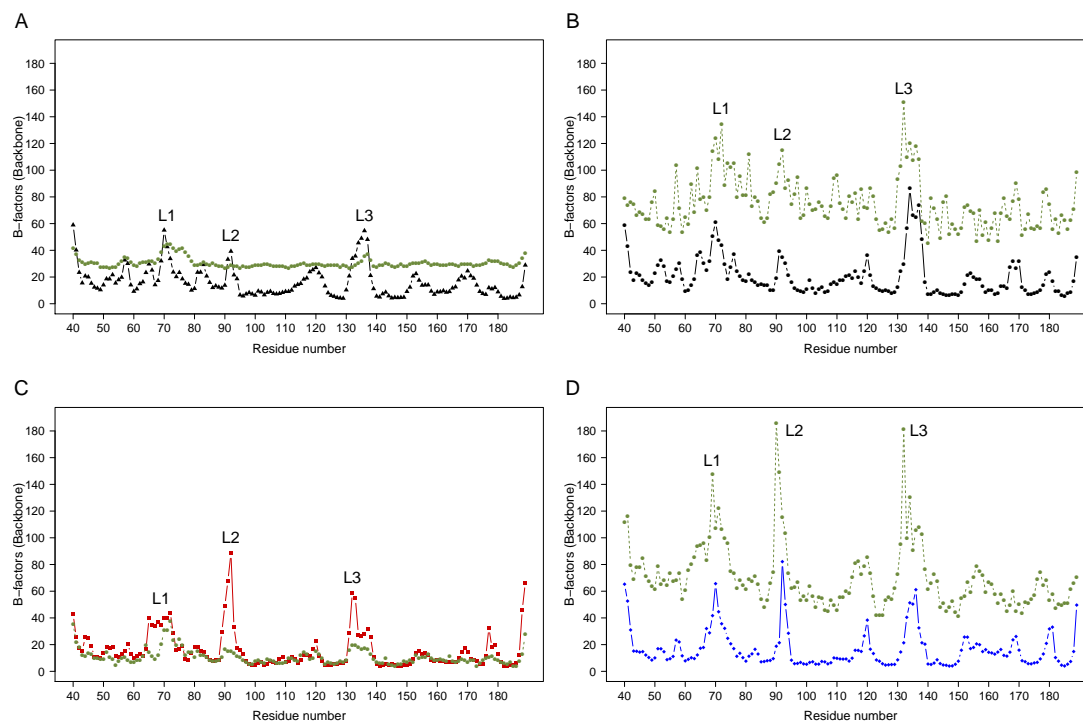


Figure C1: B-factors from molecular dynamics simulations of ShhN in states Ca2_{Ihog} (A), Ca2_{Hhip} (B), Ca0 (C), and Ca1 (D). The green curves show the corresponding experimental B-factors.

Table C1: Mean values of RMSFs from Figure 3.1 of Chapter 3, and p-values from Wilcoxon rank sum tests with null hypotheses that there is no shift between RMSF distributions in the compared pairs of states.

	Mean RMSF (nm)	p-values		
		Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.10	$1.1 \cdot 10^{-12}$	$1.5 \cdot 10^{-15}$	$2.9 \cdot 10^{-14}$
Ca1	0.07	-	0.32	0.33
Ca2 _{Hhip}	0.07	-	-	0.95
Ca2 _{Ihog}	0.07	-	-	-

Table C2: P-values of RMSD comparisons (Wilcoxon tests) between zinc center of LAS enzyme X-ray structure 1lbu and zinc centers of ShhN. All simulated ShhN states Ca0, Ca1, Ca2_{Hhip}, and Ca2_{Ihog} were included (see Figure 3.8A of Chapter 3).

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.4	$5.5 \cdot 10^{-8}$	$4.4 \cdot 10^{-5}$
Ca1	-	$2.0 \cdot 10^{-8}$	$9.4 \cdot 10^{-5}$
Ca2 _{Hhip}	-	-	0.06

Table C3: As Table C2, but referring to X-ray structure of LAS enzyme 2v0g.

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.6	$3.4 \cdot 10^{-7}$	$4.1 \cdot 10^{-10}$
Ca1	-	$5.4 \cdot 10^{-3}$	$4.2 \cdot 10^{-6}$
Ca2 _{Hhip}	-	-	0.09

Table C5: As Table C2, but referring to X-ray structure of LAS enzyme 1r44.

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.7	$4.3 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$
Ca1	-	$1.7 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$
Ca2 _{Hhip}	-	-	0.6

Table C7: P-values of RMSD comparisons (Wilcoxon tests) between zinc center of LAS enzyme structure 1lbu and zinc centers of ShhN. All Shh states Ca0, Ca1, Ca2_{Hhip}, and Ca2_{Ihog} were included (see Figure 3.8B of Chapter 3). In contrast to Table C2, the structures of the zinc centers of both the LAS enzyme and ShhN are taken from MD simulations of the respective proteins.

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.9	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
Ca1	-	$2.4 \cdot 10^{-14}$	$2.9 \cdot 10^{-12}$
Ca2 _{Hhip}	-	-	0.5

Table C4: As Table C2, but referring to X-ray structure of LAS enzyme 1u10.

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.8	$1.6 \cdot 10^{-9}$	$5.0 \cdot 10^{-12}$
Ca1	-	$1.4 \cdot 10^{-7}$	$4.9 \cdot 10^{-11}$
Ca2 _{Hhip}	-	-	0.06

Table C6: As Table C2, but referring to X-ray structure of LAS enzyme 2bop.

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.6	$2.6 \cdot 10^{-5}$	$1.6 \cdot 10^{-7}$
Ca1	-	$7.4 \cdot 10^{-4}$	$8.2 \cdot 10^{-5}$
Ca2 _{Hhip}	-	-	0.07

Table C8: As Table C7, but for comparison of LAS enzyme 2v0g and ShhN.

	p-values		
	Ca1	Ca2 _{Hhip}	Ca2 _{Ihog}
Ca0	0.4	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
Ca1	-	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
Ca2 _{Hhip}	-	-	0.08

Table C9: Distribution of distances (in nm) between carboxylate-O of E127 and imidazole proton of H135, along the charged hydrogen bond between these groups.

Given are the mean distance and the distances demarcating the quartiles, both for Ca0 and Ca2, corresponding to horizontal axis of Figure 3.9B of Chapter 3. There is little change between the states, i.e. the H-bond is conserved.

Ca ²⁺ state	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Ca0	0.129	0.181	0.194	0.199	0.216	0.270
Ca2	0.157	0.179	0.192	0.195	0.207	0.290

Table C10: Distribution of distances (in nm) between side chains of H135 and E177, the “catalytic clamp”.

Given are the mean distance and the distances demarcating the quartiles, both for Ca0 and Ca2, corresponding to vertical axis of Figure 3.9B of Chapter 3. The comparison shows that the clamp opens from Ca0 to Ca2, and that it becomes more flexible.

Ca ²⁺ state	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Ca0	0.675	0.733	0.758	0.768	0.788	1.000
Ca2	0.653	0.788	0.827	0.848	0.878	1.150

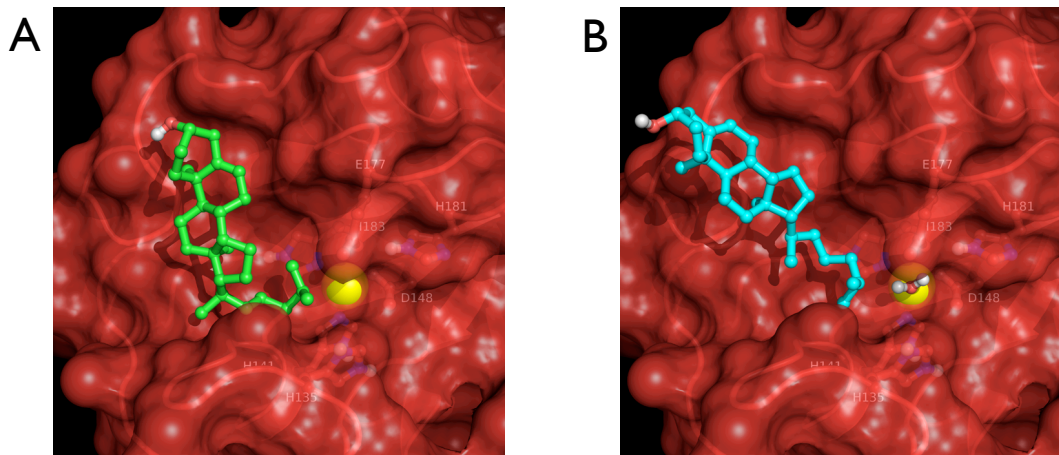


Figure C2: Docking of cholesterol to ShhN (1vhh), without zinc co-ordinating water (A), and with zinc co-ordinating water (B). The best (green) and second best (blue) poses are shown.

Table C11: Summary of docking results of cholesterol to the surface of ShhN, without zinc co-ordinating water molecule. Affinity (kcal/mol): the affinity estimate by AutoDock vina. The two RMSD columns give the root mean square deviation between the respective docking pose and the docking pose with best affinity (top ranking pose); “l.b” is a lower bound of RMSD, considering matches between atoms of same type, while “u.b” is an upper bound, computed for matches between exactly equivalent atoms according to the chosen numbering of atoms in the ligand. The affinity is expressed in kcal/mol.

Modes	Affinity	RMSD l.b	RMSD u.b
1	-6.8	0.000	0.000
2	-6.7	2.452	3.960
3	-6.5	1.887	3.633
4	-6.2	3.433	5.034
5	-6.1	2.102	3.352
6	-6.0	3.974	6.015
7	-6.0	2.316	3.153
8	-5.9	3.329	7.969
9	-5.9	2.378	3.750
10	-5.9	2.786	5.661
11	-5.8	2.734	7.353
12	-5.8	2.944	7.585
13	-5.6	2.186	3.240
14	-5.6	20.878	24.203
15	-5.5	20.929	24.506
16	-5.5	5.170	7.745
17	-5.5	4.188	6.568
18	-5.5	4.828	7.686
19	-5.4	3.379	5.252
20	-5.3	1.940	3.100

Table C12: Summary of docking results of cholesterol to the surface of ShhN including a water molecule. The water molecule of X-ray structure 1vhh was used. The affinity is expressed in kcal/mol.

Modes	Affinity	RMSD l.b	RMSD u.b
1	-7.2	0.000	0.000
2	-7.1	2.400	3.725
3	-7.0	1.461	2.358
4	-6.8	5.002	6.808
5	-6.7	1.545	3.144
6	-6.5	2.334	4.185
7	-6.4	1.579	3.111
8	-6.4	2.344	3.223
9	-6.3	1.941	2.343
10	-6.2	2.393	4.236
11	-6.2	1.668	2.152
12	-6.1	2.319	3.742
13	-6.1	3.150	5.073
14	-6.0	21.800	26.178
15	-5.9	3.228	4.485
16	-5.9	21.796	25.745
17	-5.9	22.408	26.172
18	-5.8	22.385	24.987
19	-5.7	3.241	8.804
20	-5.7	1.651	2.500

Table C13: Predicted pK_a values for calcium ligands. Ca0 (1vhh) and Ca2 states (3d1m) are shown.

Residues	Ca0 state	Ca2 state
E90	3.55	2.16
E91	5.40	5.96
D96	6.00	5.44
E127	5.02	5.97
D130	3.46	2.17
D132	4.37	2.37

Appendix D

Table D1: P-values of BSA comparisons (Wilcoxon tests) between calcium states of ShhN_{dimers} without tail. See Figure 4.3 of Chapter 4.

	p-values	
	DNT1	DNT2
DNT0	0.4	$8.2 \cdot 10^{-12}$
DNT1	-	$7.7 \cdot 10^{-12}$

Table D2: As Table D1, but referring to ShhN_{dimers} with tail.

	p-values	
	DT1	DT2
DT0	0.07	$9.3 \cdot 10^{-10}$
DT1	-	$2.3 \cdot 10^{-9}$

Table D3: P-values of binding free energy comparisons (Wilcoxon tests) between calcium states of ShhN_{dimers} without tail. See Figure 4.5A of Chapter 4.

	p-values	
	DNT1	DNT2
DNT0	0.1	$3.3 \cdot 10^{-9}$
DNT1	-	$3.3 \cdot 10^{-9}$

Table D4: As Table D3, but referring to ShhN_{dimers} with tail.

	p-values	
	DT1	DT2
DT0	0.9	$9.9 \cdot 10^{-8}$
DT1	-	$9.9 \cdot 10^{-8}$

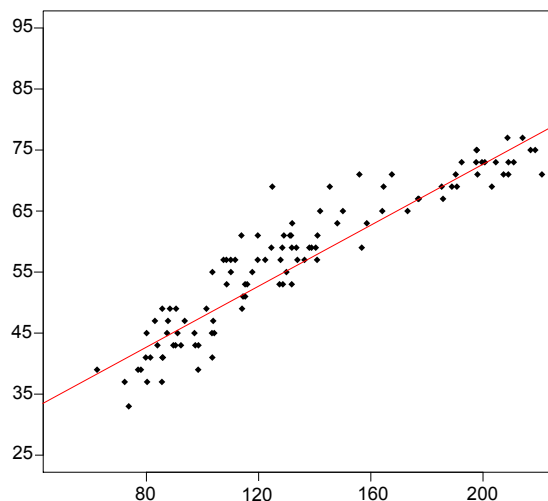


Figure D3: Correlation between the number of residues at the interfaces and the BSA. The number of interface residues (y-axis) is plotted against the ASA lost by dimerization (x-axis). Pearson's correlation between the two groups was applied showing a significant positive correlation with $R^2 = 0.93$.

Table D5: P-values of BSA comparisons (Wilcoxon tests) between DT0_{dimer} and I48A, V166A mutants. See Figure 4.9A of Chapter 4.

	p-values	
	I48A	V166A
DT0	$3.1 \cdot 10^{-15}$	$1.2 \cdot 10^{-15}$
I48A	-	0.09

Table D6: P-values of RMSD comparisons between DT0_{dimer} and I48A, V166A mutants. See Figure 4.9B of Chapter 4.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	$1.3 \cdot 10^{-14}$

Table D7: P-values of RMSF comparisons (Wilcoxon tests) between DT0_{dimer} and I48A, V166A mutants. See Figure 4.9C of Chapter 4.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-8}$	$7.4 \cdot 10^{-8}$
I48A	-	0.7

Table D8: P-values of binding free energy comparisons between DT0_{dimer} and I48A, V166A mutants. See Figure 4.9D of Chapter 4.

	p-values	
	I48A	V166A
DT0	$1.2 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$
I48A	-	0.9

Table D9: P-values of a H-bond (K178 NH-E167 OE2) comparison (Wilcoxon tests) between DT0_{dimer} and I48A, V166A mutants. See Figure 4.10 of Chapter 4.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	0.2

Table D10: As Table D9, but referring to a H-bond between Y44 H-E167 OE1.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	0.4

Table D11: As Table D9, but referring to a H-bond between N50 OD1-R153 NH2.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	$2.2 \cdot 10^{-16}$

Table D12: As Table D9, but referring to a H-bond between N50 NH-D152 O.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	$2.2 \cdot 10^{-16}$

Table D13: As Table D9, but referring to a H-bond between F170 O-R155 NH.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	$2.2 \cdot 10^{-16}$

Table D14: As Table D9, but referring to a H-bond between E167 OE1-S177 OH.

	p-values	
	I48A	V166A
DT0	$2.2 \cdot 10^{-16}$	$2.2 \cdot 10^{-16}$
I48A	-	0.4

References

- [1] Nüsslein-Volhard C and Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785):795–801, 1980. ISSN 0028-0836. doi: [10.1038/287795a0](https://doi.org/10.1038/287795a0). 1
- [2] Tabata T, Eaton S, and Kornberg T B. The drosophila hedgehog gene is expressed specifically in posterior compartment cells and is a target of engrailed regulation. *Genes Dev*, 6(12B):2635–45, Dec 1992. doi: [10.1101/gad.6.12b.2635](https://doi.org/10.1101/gad.6.12b.2635). 1
- [3] Mohler J and Vani K. Molecular organization and embryonic expression of the hedgehog gene involved in cell-cell communication in segmental patterning of drosophila. *Development*, 115(4):957–71, Aug 1992.
- [4] Lee J J, Kessler von D P, Parks S, and Beachy P A. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell*, 71(1):33–50, Oct 1992. doi: [10.1016/0092-8674\(92\)90264-D](https://doi.org/10.1016/0092-8674(92)90264-D). 1
- [5] Beachy P A, Cooper M K, Young K E, Kessler von D P, et al. Multiple roles of cholesterol in hedgehog protein biogenesis and signaling. *Cold Spring Harb Symp Quant Biol*, 62:191–204, 1997. doi: [10.1101/SQB.1997.062.01.025](https://doi.org/10.1101/SQB.1997.062.01.025). 1
- [6] Hall T M, Porter J A, Young K E, Koonin E V, et al. Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell*, 91(1):85–97, Oct 1997. doi: [10.1016/S0092-8674\(01\)80011-8](https://doi.org/10.1016/S0092-8674(01)80011-8). 1
- [7] Ingham P W and McMahon A P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev*, 15(23):3059–87, Dec 2001. doi: [10.1101/gad.938601](https://doi.org/10.1101/gad.938601). 2
- [8] Ingham P W, Nakano Y, and Seger C. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat Rev Genet*, 12(6):393–406, Jun 2011. doi: [10.1038/nrg2984](https://doi.org/10.1038/nrg2984). 1
- [9] Bürglin T R. Evolution of hedgehog and hedgehog-related genes, their origin from Hog proteins in ancestral eukaryotes and discovery of a novel Hint motif. *BMC Genomics*, 9:127, 2008. doi: [10.1186/1471-2164-9-127](https://doi.org/10.1186/1471-2164-9-127). 1
- [10] Bürglin T R. The hedgehog protein family. *Genome Biol*, 9(11):241, 2008. doi: [10.1186/gb-2008-9-11-241](https://doi.org/10.1186/gb-2008-9-11-241). 2
- [11] Krauss S, Concordet J P, and Ingham P W. A functionally conserved homolog of the *Drosophila* segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell*, 75(7):1431–1444, Dec 1993. doi: [10.1016/0092-](https://doi.org/10.1016/0092-)

- 8674(93)90628-4. 2
- [12] Echelard Y, Epstein D J, St-Jacques B, Shen L, et al. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell*, 75(7):1417–1430, 1993. ISSN 0092-8674. doi: [10.1016/0092-8674\(93\)90627-3](https://doi.org/10.1016/0092-8674(93)90627-3).
- [13] Riddle R D, Johnson R L, Laufer E, and Tabin C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, 75(7):1401–1416, Dec 1993. doi: [10.1016/0092-8674\(93\)90626-2](https://doi.org/10.1016/0092-8674(93)90626-2).
- [14] Chang D T, Lopez A, von Kessler D P, Chiang C, et al. Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development*, 120(11): 3339–3353, 1994. 2
- [15] McMahon A P, Ingham P W, and Tabin C J. Developmental roles and clinical significance of hedgehog signaling. *Curr Top Dev Biol*, 53:1–114, 2003. doi: [10.1016/S0070-2153\(03\)53002-2](https://doi.org/10.1016/S0070-2153(03)53002-2). 2, 35
- [16] Wilson C W and Chuang P T. Mechanism and evolution of cytosolic hedgehog signal transduction. *Development*, 137(13):2079–94, Jul 2010. doi: [10.1242/dev.045021](https://doi.org/10.1242/dev.045021). 2
- [17] Pathi S, Pagan-Westphal S, Baker D P, Garber E A, et al. Comparative biological responses to human Sonic, Indian, and Desert hedgehog. *Mech. Dev*, 106(1-2): 107–117, 2001. ISSN 0925-4773. doi: [10.1016/S0925-4773\(01\)00427-0](https://doi.org/10.1016/S0925-4773(01)00427-0). 2
- [18] Dessaud E, McMahon A P, and Briscoe J. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development*, 135(15):2489–503, Aug 2008. doi: [10.1242/dev.009324](https://doi.org/10.1242/dev.009324). 2
- [19] Vortkamp A, Lee K, Lanske B, Segre G V, et al. Regulation of rate of cartilage differentiation by indian hedgehog and pth-related protein. *Science*, 273(5275): 613–22, Aug 1996. doi: [10.1126/science.273.5275.613](https://doi.org/10.1126/science.273.5275.613). 3
- [20] Bijlsma M F, Peppelenbosch M P, and Spek C A. Hedgehog morphogen in cardiovascular disease. *Circulation*, 114(18):1985–91, Oct 2006. doi: [10.1161/CIRCULATIONAHA.106.619213](https://doi.org/10.1161/CIRCULATIONAHA.106.619213). 3
- [21] Varjosalo M and Taipale J. Hedgehog: functions and mechanisms. *Genes Dev*, 22(18):2454–72, Sep 2008. doi: [10.1101/gad.1693608](https://doi.org/10.1101/gad.1693608). 3
- [22] Chiang K E, Ryanand C. Hedgehog secretion and signal transduction in vertebrates. *J Biol Chem*, 287(22):17905–13, May 2012. doi: [10.1074/jbc.R112.356006](https://doi.org/10.1074/jbc.R112.356006). 3
- [23] Kicheva A, Pantazis P, Bollenbach T, Kalaidzidis Y, et al. Kinetics of morphogen gradient formation. *Science*, 315(5811):521–5, Jan 2007. doi: [10.1126/science.1135774](https://doi.org/10.1126/science.1135774). 3
- [24] Müller P and Schier A F. Extracellular movement of signaling molecules. *Dev Cell*, 21(1):145–58, Jul 2011. doi: [10.1016/j.devcel.2011.06.001](https://doi.org/10.1016/j.devcel.2011.06.001). 3
- [25] Wolpert L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol*, 25(1):1–47, 1969. ISSN 0022-5193. doi: [10.1016/S0022-5193\(69\)80016-0](https://doi.org/10.1016/S0022-5193(69)80016-0). 3

- [26] Briscoe J and Théron P P. The mechanisms of hedgehog signalling and its roles in development and disease. *Nat Rev Mol Cell Biol*, 14(7):416–29, Jul 2013. doi: [10.1038/nrm3598](https://doi.org/10.1038/nrm3598). 3, 5
- [27] Ingham P W and Placzek M. Orchestrating ontogenesis: variations on a theme by sonic hedgehog. *Nat Rev Genet*, 7(11):841–50, Nov 2006. doi: [10.1038/nrg1969](https://doi.org/10.1038/nrg1969). 3
- [28] Porter J A, Young K E, and Beachy P A. Cholesterol modification of hedgehog signaling proteins in animal development. *Science*, 274(5285):255–9, Oct 1996. doi: [10.1126/science.274.5285.255](https://doi.org/10.1126/science.274.5285.255). 3, 55, 59
- [29] Chen X, Tukachinsky H, Huang C-H, Jao C, et al. Processing and turnover of the Hedgehog protein in the endoplasmic reticulum. *J Cell Biol*, 192(5):825–38, Mar 2011. doi: [10.1083/jcb.201008090](https://doi.org/10.1083/jcb.201008090). 4
- [30] Ma G, Xiao Y, and He L. Recent progress in the study of Hedgehog signaling. *J. Genet. Genomics*, 35(3):129–137, 2008. ISSN 1673-8527. doi: [10.1016/S1673-8527\(08\)60019-3](https://doi.org/10.1016/S1673-8527(08)60019-3). 4
- [31] Hofmann K. A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling. *Trends Biochem. Sci*, 25(3):111–112, 2000. ISSN 0968-0004. doi: [10.1016/S0968-0004\(99\)01539-X](https://doi.org/10.1016/S0968-0004(99)01539-X). 4
- [32] Chamoun Z, Mann R K, Nellen D, von Kessler D P, et al. Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science*, 293(5537):2080–4, Sep 2001. doi: [10.1126/science.1064437](https://doi.org/10.1126/science.1064437).
- [33] Lee J D and Treisman J E. Sightless has homology to transmembrane acyltransferases and is required to generate active Hedgehog protein. *Curr Biol*, 11(14):1147–1152, 2001. ISSN 0960-9822. doi: [10.1016/S0960-9822\(01\)00323-2](https://doi.org/10.1016/S0960-9822(01)00323-2). 4
- [34] Buglino J A and Resh M D. Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J Biol Chem*, 283(32):22076–88, Aug 2008. doi: [10.1074/jbc.M803901200](https://doi.org/10.1074/jbc.M803901200). 4
- [35] Pepinsky R B, Zeng C, Wen D, Rayhorn P, et al. Identification of a Palmitic Acid-modified Form of Human Sonic hedgehog. *Journal of Biological Chemistry*, 273(22):14037–14045, 1998. doi: [10.1074/jbc.273.22.14037](https://doi.org/10.1074/jbc.273.22.14037). 4
- [36] Zeng X, Goetz J A, Suber L M, Scott Jr W J, et al. A freely diffusible form of Sonic hedgehog mediates long-range signalling. *Nature*, 411(6838):716–20, Jun 2001. doi: [10.1038/35079648](https://doi.org/10.1038/35079648). 4, 5, 6, 7, 61, 78
- [37] Eugster C, Panáková D, Mahmoud A, and Eaton S. Lipoprotein-Heparan Sulfate Interactions in the Hh Pathway. *Dev Cell*, 13(1):57 – 71, 2007. ISSN 1534-5807. doi: [10.1016/j.devcel.2007.04.019](https://doi.org/10.1016/j.devcel.2007.04.019). 6
- [38] Tukachinsky H, Kuzmickas R P, Jao C Y, Liu J, and Salic A. Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. *Cell Rep*, 2(2):308–20, Aug 2012. doi: [10.1016/j.celrep.2012.07.010](https://doi.org/10.1016/j.celrep.2012.07.010). 4, 6
- [39] Nozawa Y I, Lin C, and Chuang P T. Hedgehog signaling from the primary cilium to the nucleus: an emerging picture of ciliary localization, trafficking and transduction. *Curr Opin Genet Dev*, 23(4):429–37, Aug 2013. doi: [10.1016/j.gde.2013.04.008](https://doi.org/10.1016/j.gde.2013.04.008). 4

- [40] Rajat Rohatgi, Ljiljana Milenkovic, and Matthew P Scott. Patched1 regulates hedgehog signaling at the primary cilium. *Science*, 317(5836):372–6, Jul 2007. doi: [10.1126/science.1139740](https://doi.org/10.1126/science.1139740). 4
- [41] Taipale J, Cooper M K, Maiti T, and Beachy P A. Patched acts catalytically to suppress the activity of Smoothed. *Nature*, 418(6900):892–7, Aug 2002. doi: [10.1038/nature00989](https://doi.org/10.1038/nature00989). 5
- [42] Hui C-C and Angers S. Gli proteins in development and disease. *Annu Rev Cell Dev Biol*, 27:513–37, 2011. doi: [10.1146/annurev-cellbio-092910-154048](https://doi.org/10.1146/annurev-cellbio-092910-154048). 5, 10
- [43] Peters C, Wolf A, Wagner M, Kuhlmann J, and Waldmann H. The cholesterol membrane anchor of the Hedgehog protein confers stable membrane association to lipid-modified proteins. *Proc Natl Acad Sci U S A*, 101(23):8531–6, Jun 2004. doi: [10.1073/pnas.0308449101](https://doi.org/10.1073/pnas.0308449101). 5
- [44] Chen M-H, Li Y-J, Kawakami T, Xu S-M, et al. Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates. *Genes & Development*, 18(6):641–659, March 2004. doi: [10.1101/gad.1185804](https://doi.org/10.1101/gad.1185804). 5, 6, 7
- [45] Gallet A. Hedgehog morphogen: from secretion to reception. *Trends Cell Biol*, 21(4):238–46, Apr 2011. doi: [10.1016/j.tcb.2010.12.005](https://doi.org/10.1016/j.tcb.2010.12.005). 5, 7
- [46] Goetz J A, Singh S, Suber L M, F Jon Kull, and Robbins D J. A highly conserved amino-terminal region of sonic hedgehog is required for the formation of its freely diffusible multimeric form. *J. Biol. Chem*, 281(7):4087–93, Feb 2006. doi: [10.1074/jbc.M511427200](https://doi.org/10.1074/jbc.M511427200). 5, 7, 10, 61, 64, 77, 78, 81
- [47] Vyas N, Goswami D, Manonmani A, Sharma P, et al. Nanoscale organization of hedgehog is essential for long-range signaling. *Cell*, 133(7):1214–27, Jun 2008. doi: [10.1016/j.cell.2008.05.026](https://doi.org/10.1016/j.cell.2008.05.026). 5, 35, 55, 77, 78
- [48] Farshi P, Ohlig S, Pickhinke U, Höing S, et al. Dual roles of the cardin-weintraub motif in multimeric sonic hedgehog. *J Biol Chem*, 286(26):23608–19, Jul 2011. doi: [10.1074/jbc.M110.206474](https://doi.org/10.1074/jbc.M110.206474). 10, 69
- [49] Chang S-C, Mulloy B, Magee A I, and Couchman J R. Two distinct sites in sonic Hedgehog combine for heparan sulfate interactions and cell signaling functions. *J Biol Chem*, 286(52):44391–402, Dec 2011. doi: [10.1074/jbc.M111.285361](https://doi.org/10.1074/jbc.M111.285361). 5
- [50] Vincent J-P. Hedgehog nanopackages ready for dispatch. *Cell*, 133(7):1139–41, Jun 2008. doi: [10.1016/j.cell.2008.06.011](https://doi.org/10.1016/j.cell.2008.06.011). 5
- [51] Queiroz K C S, Tio R A, Zeebregts C J, Bijlsma M F, et al. Human plasma very low density lipoprotein carries Indian hedgehog. *J Proteome Res*, 9(11):6052–9, Nov 2010. doi: [10.1021/pr100403q](https://doi.org/10.1021/pr100403q). 6
- [52] Palm W, Swierczynska M M, Kumari V, Ehrhart-Bornstein M, et al. Secretion and signaling activities of lipoprotein-associated hedgehog and non-sterol-modified hedgehog in flies and mammals. *PLoS Biol*, 11(3):e1001505, 2013. doi: [10.1371/journal.pbio.1001505](https://doi.org/10.1371/journal.pbio.1001505). 6, 55, 59
- [53] Ayers K L, Gallet A, Staccini-Lavenant L, and Thérond P P. The long-range ac-

- tivity of Hedgehog is regulated in the apical extracellular space by the glypican Dally and the hydrolase Notum. *Dev Cell*, 18(4):605–20, Apr 2010. doi: [10.1016/j.devcel.2010.02.015](https://doi.org/10.1016/j.devcel.2010.02.015). 6, 10
- [54] Panáková D, Sprong H, Marois E, Thiele C, and Eaton S. Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature*, 435(7038):58–65, May 2005. doi: [10.1038/nature03504](https://doi.org/10.1038/nature03504). 6
- [55] Burke R, Nellen D, Bellotto M, Hafen E, et al. Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell*, 99(7):803–15, Dec 1999. doi: [10.1016/S0092-8674\(00\)81677-3](https://doi.org/10.1016/S0092-8674(00)81677-3). 6
- [56] Ma Y, Erkner A, Gong R, Yao S, et al. Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of dispatched. *Cell*, 111(1):63–75, 2002. ISSN 0092-8674. doi: [10.1016/S0092-8674\(02\)00977-7](https://doi.org/10.1016/S0092-8674(02)00977-7). 6
- [57] Tseng T T, Gratwick K S, Kollman J, Park D, et al. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol*, 1(1):107–25, Aug 1999. 6
- [58] Grimmond S, Larder R, Van Hateren N, Siggers P, et al. Cloning, mapping, and expression analysis of a gene encoding a novel mammalian EGF-related protein (SCUBE1). *Genomics*, 70(1):74–81, Nov 2000. doi: [10.1006/geno.2000.6370](https://doi.org/10.1006/geno.2000.6370). 6
- [59] Kawakami A, Nojima Y, Toyoda A, Takahoko M, et al. The zebrafish-secreted matrix protein you/scube2 is implicated in long-range regulation of hedgehog signaling. *Curr Biol*, 15(5):480–8, Mar 2005. doi: [10.1016/j.cub.2005.02.018](https://doi.org/10.1016/j.cub.2005.02.018). 6
- [60] Hollway G E, Maule J, Gautier P, Evans T M, et al. Scube2 mediates Hedgehog signalling in the zebrafish embryo. *Dev Biol*, 294(1):104–18, Jun 2006. doi: [10.1016/j.ydbio.2006.02.032](https://doi.org/10.1016/j.ydbio.2006.02.032). 6
- [61] Creanga A, Glenn T D, Mann R K, Saunders A M, et al. Scube/You activity mediates release of dually lipid-modified Hedgehog signal in soluble form. *Genes Dev*, 26(12):1312–25, Jun 2012. doi: [10.1101/gad.191866.112](https://doi.org/10.1101/gad.191866.112). 6
- [62] Dierker T, Dreier R, Petersen A, Bordych C, and Grobe K. Heparan sulfate-modulated, metalloprotease-mediated sonic hedgehog release from producing cells. *J. Biol. Chem*, 284(12):8013–22, Mar 2009. doi: [10.1074/jbc.M806838200](https://doi.org/10.1074/jbc.M806838200). 7
- [63] Ohlig S, Farshi P, Pickhinke U, van den Boom J, et al. Sonic hedgehog shedding results in functional activation of the solubilized protein. *Dev Cell*, 20(6):764–74, Jun 2011. doi: [10.1016/j.devcel.2011.05.010](https://doi.org/10.1016/j.devcel.2011.05.010). 55, 59, 62, 77, 81
- [64] Ohlig S, Pickhinke U, Sirko S, Bandari S, et al. An emerging role of sonic hedgehog shedding as a modulator of heparan sulfate interactions. *J Biol Chem*, 287(52):43708–19, Dec 2012. doi: [10.1074/jbc.M112.356667](https://doi.org/10.1074/jbc.M112.356667). 7, 62, 66, 69
- [65] Dierker T, Dreier R, Migone M, Hamer S, and Grobe K. Heparan sulfate and transglutaminase activity are required for the formation of covalently cross-linked hedgehog oligomers. *J. Biol. Chem*, 284(47):32562–71, Nov 2009. doi: [10.1074/jbc.M109.044867](https://doi.org/10.1074/jbc.M109.044867). 7

- [66] Tanaka Y, Okada Y, and Hirokawa N. FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. *Nature*, 435(7039):172–7, May 2005. doi: [10.1038/nature03494](https://doi.org/10.1038/nature03494). 7
- [67] Thérond P P. Release and transportation of Hedgehog molecules. *Curr. Opin. Cell Biol*, 24(2):173–80, Apr 2012. doi: [10.1016/j.ceb.2012.02.001](https://doi.org/10.1016/j.ceb.2012.02.001). 7
- [68] Jianchi Feng, Bryan White, Oksana V Tyurina, Burcu Guner, Theresa Larson, Hae Young Lee, Rolf O Karlstrom, and Jhumku D Kohtz. Synergistic and antagonistic roles of the Sonic hedgehog N- and C-terminal lipids. *Development*, 131(17):4357–70, Sep 2004. doi: [10.1242/dev.01301](https://doi.org/10.1242/dev.01301). 7, 78
- [69] Eaton S. Multiple roles for lipids in the hedgehog signalling pathway. *Nat Rev Mol Cell Biol*, 9(6):437–445, 2008. doi: [10.1038/nrm2414](https://doi.org/10.1038/nrm2414). 7
- [70] Lewis P M, Dunn M P, McMahon J A, Logan M, et al. Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by ptc1. *Cell*, 105(5):599–612, Jun 2001. doi: [10.1016/S0092-8674\(01\)00369-5](https://doi.org/10.1016/S0092-8674(01)00369-5). 7
- [71] Gallet A, Ruel L, Staccini-Lavenant L, and Thérond P P. Cholesterol modification is necessary for controlled planar long-range activity of Hedgehog in Drosophila epithelia. *Development*, 133(3):407–18, Feb 2006. doi: [10.1242/dev.02212](https://doi.org/10.1242/dev.02212). 59, 78
- [72] Guerrero I and Chiang C. A conserved mechanism of Hedgehog gradient formation by lipid modifications. *Trends Cell Biol*, 17(1):1–5, Jan 2007. doi: [10.1016/j.tcb.2006.11.002](https://doi.org/10.1016/j.tcb.2006.11.002).
- [73] Chamberlain C E, Jeong J, Guo C, Allen B L, and McMahon A P. Notochord-derived Shh concentrates in close association with the apically positioned basal body in neural target cells and forms a dynamic gradient during neural patterning. *Development*, 135(6):1097–106, Mar 2008. doi: [10.1242/dev.013086](https://doi.org/10.1242/dev.013086). 7
- [74] Riobo N A. Cholesterol and its derivatives in Sonic Hedgehog signaling and cancer. *Curr Opin Pharmacol*, 12(6):736–41, Dec 2012. doi: [10.1016/j.coph.2012.07.002](https://doi.org/10.1016/j.coph.2012.07.002). 7, 55, 59
- [75] Tokhunts R, Singh S, Chu T, D’Angelo G, et al. The full-length unprocessed hedgehog protein is an active signaling molecule. *J Biol Chem*, 285(4):2562–8, Jan 2010. doi: [10.1074/jbc.M109.078626](https://doi.org/10.1074/jbc.M109.078626). 7
- [76] Kohtz J D, Lee H Y, Gaiano N, Segal J, et al. N-terminal fatty-acylation of sonic hedgehog enhances the induction of rodent ventral forebrain neurons. *Development*, 128(12):2351–63, Jun 2001. 7
- [77] Dawber R J, Hebbes S, Herpers B, Docquier F, and van den Heuvel M. Differential range and activity of various forms of the Hedgehog protein. *BMC Dev Biol*, 5:21, 2005. doi: [10.1186/1471-213X-5-21](https://doi.org/10.1186/1471-213X-5-21). 7
- [78] Wilson C W and Chuang P-T. New “hogs” in Hedgehog transport and signal reception. *Cell*, 125(3):435–8, May 2006. doi: [10.1016/j.cell.2006.04.016](https://doi.org/10.1016/j.cell.2006.04.016). 8
- [79] Beckett K, Franch-Marro X, and Vincent J-P. Glypican-mediated endocytosis of Hedgehog has opposite effects in flies and mice. *Trends Cell Biol*, 18(8):360–3,

- Aug 2008. doi: [10.1016/j.tcb.2008.06.001](https://doi.org/10.1016/j.tcb.2008.06.001). 8
- [80] Kang J S, Zhang W, and Krauss R S. Hedgehog signaling: cooking with Gas1. *Sci STKE*, 2007(403):pe50, Sep 2007. doi: [10.1126/stke.4032007pe50](https://doi.org/10.1126/stke.4032007pe50). 8
- [81] Allen B L, Tenzen T, and McMahon A P. The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. *Genes Dev*, 21(10):1244–57, May 2007. doi: [10.1101/gad.1543607](https://doi.org/10.1101/gad.1543607). 8, 9
- [82] Chuang P T and McMahon A P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature*, 397(6720):617–21, Feb 1999. doi: [10.1038/17611](https://doi.org/10.1038/17611). 8
- [83] Tenzen T, Allen B L, Cole F, Kang J-S, et al. The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev Cell*, 10(5):647–56, May 2006. doi: [10.1016/j.devcel.2006.04.004](https://doi.org/10.1016/j.devcel.2006.04.004). 8
- [84] Yao S, Lum L, and Beachy P A. The Ihog cell-surface proteins bind Hedgehog and mediate pathway activation. *Cell*, 125(2):343–57, Apr 2006. doi: [10.1016/j.cell.2006.02.040](https://doi.org/10.1016/j.cell.2006.02.040). 8
- [85] Zheng X, Mann R K, Sever N, and Beachy P A. Genetic and biochemical definition of the hedgehog receptor. *Genes Dev*, 24(1):57–71, Jan 2010. doi: [10.1101/gad.1870310](https://doi.org/10.1101/gad.1870310). 8
- [86] McLellan J S, Yao S, Zheng X, Geisbrecht B V, et al. Structure of a heparin-dependent complex of Hedgehog and Ihog. *Proc Natl Acad Sci U S A*, 103(46):17208–13, Nov 2006. doi: [10.1073/pnas.0606738103](https://doi.org/10.1073/pnas.0606738103). 8, 52
- [87] McLellan J S, Zheng X, Hauk G, Ghirlando R, et al. The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. *Nature*, 455(7215):979–83, Oct 2008. doi: [10.1038/nature07358](https://doi.org/10.1038/nature07358). 8, 9, 12, 16, 35, 36, 40, 56, 57, 58
- [88] Kavran J M, Ward M D, Olatosu O O, Mulepati S, and Leahy D J. All mammalian Hedgehog proteins interact with cell adhesion molecule, down-regulated by oncogenes (CDO) and brother of CDO (BOC) in a conserved manner. *J Biol Chem*, 285(32):24584–90, Aug 2010. doi: [10.1074/jbc.M110.131680](https://doi.org/10.1074/jbc.M110.131680). 8, 16, 36, 56, 62
- [89] Beachy P A, Hymowitz S G, Lazaru R A, Leahy D J, and Siebold C. Interactions between hedgehog proteins and their binding partners come into view. *Genes Dev*, 24(18):2001–12, Sep 2010. doi: [10.1101/gad.1951710](https://doi.org/10.1101/gad.1951710). 8, 9
- [90] Kang J S, Gao M, Feinleib J L, Cotter P D, et al. CDO: an oncogene-, serum-, and anchorage-regulated member of the Ig/fibronectin type III repeat family. *J Cell Biol*, 138(1):203–13, Jul 1997. 8
- [91] J S Kang, Mulieri P J, Hu Y, Taliana L, and Krauss R S. Boc, an ig superfamily member, associates with cdo to positively regulate myogenic differentiation. *EMBO J*, 21(1):114–124, 2002. doi: [10.1093/emboj/21.1.114](https://doi.org/10.1093/emboj/21.1.114). 8
- [92] Chuang P T, Kawcak T, and McMahon A P. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes Dev*, 17(3):342–7, Feb 2003. doi: [10.1101/gad.108103](https://doi.org/10.1101/gad.108103). 8

- [10.1101/gad.1026303](https://doi.org/10.1101/gad.1026303). 8
- [93] Bosanac I, Maun H R, Scales S J, Wen X, et al. The structure of SHH in complex with HHIP reveals a recognition role for the Shh pseudo active site in signaling. *Nat Struct Mol Biol*, 16(7):691–7, Jul 2009. doi: [10.1038/nsmb.1632](https://doi.org/10.1038/nsmb.1632). 8, 16, 35, 56
- [94] Bishop B, Aricescu A R, Harlos K, O’Callaghan C A, et al. Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. *Nat Struct Mol Biol*, 16(7):698–703, Jul 2009. doi: [10.1038/nsmb.1607](https://doi.org/10.1038/nsmb.1607). 8, 15, 16, 35, 36, 40, 56, 57
- [95] Lee C S, Buttitta L, and Fan C M. Evidence that the WNT-inducible growth arrest-specific gene 1 encodes an antagonist of sonic hedgehog signaling in the somite. *Proc Natl Acad Sci U S A*, 98(20):11347–52, Sep 2001. doi: [10.1073/pnas.201418298](https://doi.org/10.1073/pnas.201418298). 9
- [96] Martinelli D C and Fan C M. Gas1 extends the range of Hedgehog action by facilitating its signaling. *Genes Dev*, 21(10):1231–43, May 2007. doi: [10.1101/gad.1546307](https://doi.org/10.1101/gad.1546307). 9
- [97] Anne Imberty, Hugues Lortat-Jacob, and Serge Pérez. Structural view of glycosaminoglycan-protein interactions. *Carbohydr Res*, 342(3-4):430–9, Feb 2007. doi: [10.1016/j.carres.2006.12.019](https://doi.org/10.1016/j.carres.2006.12.019). 9
- [98] Sasisekharan R and Venkataraman G. Heparin and heparan sulfate: biosynthesis, structure and function. *Curr Opin Chem Biol*, 4(6):626–31, Dec 2000. 9
- [99] Esko J D and Lindahl U. Molecular diversity of heparan sulfate. *J Clin Invest*, 108(2): 169–73, Jul 2001. doi: [10.1172/JCI13530](https://doi.org/10.1172/JCI13530). 9
- [100] Gandhi N S and Mancera R L. The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des*, 72(6):455–82, Dec 2008. doi: [10.1111/j.1747-0285.2008.00741.x](https://doi.org/10.1111/j.1747-0285.2008.00741.x). 9, 75, 79
- [101] Lin X. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development*, 131(24):6009–21, Dec 2004. doi: [10.1242/dev.01522](https://doi.org/10.1242/dev.01522). 10
- [102] The I, Bellaiche Y, and Perrimon N. Hedgehog movement is regulated through tout velu-dependent synthesis of a heparan sulfate proteoglycan. *Mol Cell*, 4(4):633–9, Oct 1999. doi: [10.1016/S1097-2765\(00\)80214-2](https://doi.org/10.1016/S1097-2765(00)80214-2). 10
- [103] Bishop J R, Schuksz M, and Esko J D. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature*, 446(7139):1030–7, Apr 2007. doi: [10.1038/nature05817](https://doi.org/10.1038/nature05817). 10
- [104] Capila I and Linhardt R J. Heparin-protein interactions. *Angew Chem Int Ed Engl*, 41 (3):391–412, Feb 2002. 10
- [105] Kramer K L and Yost H J. Heparan sulfate core proteins in cell-cell signaling. *Annu Rev Genet*, 37:461–84, 2003. doi: [10.1146/annurev.genet.37.061103.090226](https://doi.org/10.1146/annurev.genet.37.061103.090226). 10, 75
- [106] Koziel L, Kunath M, Kelly O G, and Vortkamp A. Ext1-dependent heparan sulfate regulates the range of ihh signaling during endochondral ossification. *Dev Cell*, 6 (6):801–13, Jun 2004. doi: [10.1016/j.devcel.2004.05.009](https://doi.org/10.1016/j.devcel.2004.05.009). 10, 75
- [107] Yan D and Lin X. Shaping morphogen gradients by proteoglycans. *Cold Spring Harb*

- Perspect Biol*, 1(3):a002493, Sep 2009. doi: [10.1101/cshperspect.a002493](https://doi.org/10.1101/cshperspect.a002493). 10
- [108] Carrasco H, Olivares G H, Faunes F, Oliva C, and Larraín J. Heparan sulfate proteoglycans exert positive and negative effects in Shh activity. *J Cell Biochem*, 96(4): 831–8, Nov 2005. doi: [10.1002/jcb.20586](https://doi.org/10.1002/jcb.20586). 10
- [109] Rubin J B, Choi Y, and Segal R A. Cerebellar proteoglycans regulate sonic hedgehog responses during development. *Development*, 129(9):2223–32, May 2002. 10
- [110] Cardin A D and Weintraub H J. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis*, 9(1):21–32, 1989. doi: [10.1161/01.ATV.9.1.21](https://doi.org/10.1161/01.ATV.9.1.21). 10
- [111] Whalen D M, Malinauskas T, Gilbert R J C, and Siebold C. Structural insights into proteoglycan-shaped Hedgehog signaling. *Proc Natl Acad Sci USA*, 110(41):16420–5, Oct 2013. doi: [10.1073/pnas.1310097110](https://doi.org/10.1073/pnas.1310097110). 10, 61, 64, 75, 76
- [112] Zhang F, McLellan J S, Ayala A M, Leahy D J, and Linhardt R J. Kinetic and structural studies on interactions between heparin or heparan sulfate and proteins of the hedgehog signaling pathway. *Biochemistry*, 46(13):3933–41, Apr 2007. doi: [10.1021/bi6025424](https://doi.org/10.1021/bi6025424). 10
- [113] Beachy P A, Karhadkar S S, and Berman D M. Tissue repair and stem cell renewal in carcinogenesis. *Nature*, 432(7015):324–31, Nov 2004. doi: [10.1038/nature03100](https://doi.org/10.1038/nature03100). 10
- [114] Evangelista M, Tian H, and de Sauvage F J. The hedgehog signaling pathway in cancer. *Clin Cancer Res*, 12(20 Pt 1):5924–8, Oct 2006. doi: [10.1158/1078-0432.CCR-06-1736](https://doi.org/10.1158/1078-0432.CCR-06-1736). 10
- [115] Briscoe J and Théron P. Hedgehog signaling: from the Drosophila cuticle to anti-cancer drugs. *Dev Cell*, 8(2):143–51, Feb 2005. doi: [10.1016/j.devcel.2005.01.008](https://doi.org/10.1016/j.devcel.2005.01.008). 10
- [116] Kuo J S, Casey S O, Thompson L, and Truwit C L. Pallister-Hall Syndrome: Clinical and MR Features. *American Journal of Neuroradiology*, 20(10):1839–1841, 1999. URL <http://www.ajnr.org/content/20/10/1839.abstract>. 10
- [117] Naruse I, Ueta E, Sumino Y, Ogawa M, and Ishikiriya S. Birth defects caused by mutations in human gli3 and mouse gli3 genes. *Congenit Anom (Kyoto)*, 50(1):1–7, Mar 2010. doi: [10.1111/j.1741-4520.2009.00266.x](https://doi.org/10.1111/j.1741-4520.2009.00266.x). 10
- [118] Vaz S S, Chodirker B, Prasad C, Seabrook J A, et al. Risk factors for nonsyndromic holoprosencephaly: a Manitoba case-control study. *Am J Med Genet A*, 158A(4): 751–8, Apr 2012. doi: [10.1002/ajmg.a.35240](https://doi.org/10.1002/ajmg.a.35240). 11
- [119] Klein O P, Oberoi S, Huysseune A, Hovorakova M, et al. Developmental disorders of the dentition: An update. *Am J Med Genet C Semin Med Genet*, Oct 2013. doi: [10.1002/ajmg.c.31382](https://doi.org/10.1002/ajmg.c.31382). 11
- [120] Wallis D and Muenke M. Mutations in holoprosencephaly. *Hum Mutat*, 16(2):99–108, 2000. doi: [10.1002/1098-1004\(200008\)16:2<99::AID-HUMU2>3.0.CO;2-0](https://doi.org/10.1002/1098-1004(200008)16:2<99::AID-HUMU2>3.0.CO;2-0). 11
- [121] Pan A, Chang Le, Nguyen A, and James A W. A review of hedgehog signaling in cranial bone development. *Front Physiol*, 4:61, 2013. doi: [10.3389/fphys.2013.00061](https://doi.org/10.3389/fphys.2013.00061). 11

- [122] Maity T, Fuse N, and Beachy P A. Molecular mechanisms of Sonic hedgehog mutant effects in holoprosencephaly. *Proc Natl Acad Sci U S A*, 102(47):17026–31, Nov 2005. doi: [10.1073/pnas.0507848102](https://doi.org/10.1073/pnas.0507848102). 11, 16, 35
- [123] Odent S, Attié-Bitach T, Blayau M, Mathieu M, et al. Expression of the Sonic hedgehog (SHH) Gene during Early Human Development and Phenotypic Expression of New Mutations Causing Holoprosencephaly. *Human Molecular Genetics*, 8(9):1683–1689, 1999. doi: [10.1093/hmg/8.9.1683](https://doi.org/10.1093/hmg/8.9.1683). 11
- [124] Lami F, Carli D, Ferrari P, Marini M, et al. Holoprosencephaly: report of four cases and genotype-phenotype correlations. *J Genet*, 92(1):97–101, Apr 2013. doi: [10.1007/s12041-013-0215-5](https://doi.org/10.1007/s12041-013-0215-5). 11
- [125] Heussler H S, Suri M, Young I D, and Muenke M. Extreme variability of expression of a Sonic Hedgehog mutation: attention difficulties and holoprosencephaly. *Arch Dis Child*, 86(4):293–296, 2002. doi: [10.1136/adc.86.4.293](https://doi.org/10.1136/adc.86.4.293). 11
- [126] Belloni E, Muenke M, Roessler E, Traverso G, et al. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet*, 14(3):353–6, Nov 1996. doi: [10.1038/ng1196-353](https://doi.org/10.1038/ng1196-353). 11
- [127] Roessler E, Belloni E, Gaudenz K, Jay P, et al. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet*, 14(3):357–60, Nov 1996. doi: [10.1038/ng1196-357](https://doi.org/10.1038/ng1196-357). 11
- [128] Lodder E M, Hoogeboom A J M, Coert J H, and de Graaff E. Deletion of 1 amino acid in Indian hedgehog leads to brachydactyly A1. *Am J Med Genet A*, 146A(16):2152–4, Aug 2008. doi: [10.1002/ajmg.a.32441](https://doi.org/10.1002/ajmg.a.32441). 11
- [129] Temtamy S A and Aglan M S. Brachydactyly. *Orphanet J Rare Dis*, 3:15, 2008. doi: [10.1186/1750-1172-3-15](https://doi.org/10.1186/1750-1172-3-15). 11
- [130] Farabee W C. *Hereditary and sexual influence in meristic variation: A study of digital malformations in man*. PhD thesis, Harvard University, 1903. 11
- [131] Bell J. On brachydactyly and symphalangism. *Treasury of Human Inheritance*. Cambridge University Press, pages 1–31, 1951. 11
- [132] Gao B, Guo J, She C, Shu A, et al. Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. *Nat Genet*, 28(4):386–8, Aug 2001. doi: [10.1038/ng577](https://doi.org/10.1038/ng577). 11
- [133] Byrnes A M, Racacho L, Grimsey A, Hudgins L, et al. Brachydactyly A-1 mutations restricted to the central region of the N-terminal active fragment of Indian Hedgehog. *Eur J Hum Genet*, 17(9):1112–20, Sep 2009. doi: [10.1038/ejhg.2009.18](https://doi.org/10.1038/ejhg.2009.18). 11, 35
- [134] Bo Gao and Lin He. Answering a century old riddle: brachydactyly type a1. *Cell Res*, 14(3):179–87, Jun 2004. doi: [10.1038/sj.cr.7290218](https://doi.org/10.1038/sj.cr.7290218). 11
- [135] McCready M E, Sweeney E, Fryer A E, Donnai D, et al. A novel mutation in the IHH gene causes brachydactyly type A1: a 95-year-old mystery resolved. *Hum Genet*, 111(4-5):368–75, Oct 2002. doi: [10.1007/s00439-002-0815-2](https://doi.org/10.1007/s00439-002-0815-2). 11
- [136] Kirkpatrick T J, Au K-S, Mastrobattista J M, McCready M E, et al. Identification of a mutation in the Indian Hedgehog (IHH) gene causing brachydactyly type A1 and evidence for a third locus. *J Med Genet*, 40(1):42–4, Jan 2003.

- [137] Liu M, Wang X, Cai Z, Tang Z, et al. A novel heterozygous mutation in the Indian hedgehog gene (IHH) is associated with brachydactyly type A1 in a Chinese family. *J Hum Genet*, 51(8):727–31, 2006. doi: [10.1007/s10038-006-0012-6](https://doi.org/10.1007/s10038-006-0012-6).
- [138] Zhu G, Ke X, Liu Q, Li J, et al. Recurrence of the D100N mutation in a Chinese family with brachydactyly type A1: Evidence for a mutational hot spot in the Indian hedgehog gene. *Am J Med Genet Part A*, 143A(11):1246–1248, 2007. ISSN 1552-4833. doi: [10.1002/ajmg.a.31728](https://doi.org/10.1002/ajmg.a.31728).
- [139] Gao B, , Hu J, Stricker S, Cheung M, et al. A mutation in *Ihh* that causes digit abnormalities alters its signalling capacity and range. *Nature*, 458(7242):1196–200, Apr 2009. doi: [10.1038/nature07862](https://doi.org/10.1038/nature07862). 58
- [140] Stattin E-L, Lindén B, Lönnerholm T, Schuster J, and Dahl N. Brachydactyly type A1 associated with unusual radiological findings and a novel Arg158Cys mutation in the Indian hedgehog (IHH) gene. *Eur J Med Genet*, 52(5):297–302, 2009. doi: [10.1016/j.ejmg.2009.05.008](https://doi.org/10.1016/j.ejmg.2009.05.008). 11
- [141] Hall T M, Porter J A, Beachy P A, and Leahy D J. A potential catalytic site revealed by the 1.7-Å crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature*, 378(6553):212–6, Nov 1995. doi: [10.1038/378212a0](https://doi.org/10.1038/378212a0). 12, 36, 44, 47, 49, 50, 51, 52, 53, 56, 64
- [142] Christianson D W. Structural biology of zinc. *Adv Protein Chem*, 42:281–355, 1991. ISSN 0065-3233. doi: [10.1016/S0065-3233\(08\)60538-0](https://doi.org/10.1016/S0065-3233(08)60538-0). 12
- [143] Berman H, Henrick K, Nakamura H, and Markley J L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res*, 35(Database issue):D301–3, Jan 2007. doi: [10.1093/nar/gkl971](https://doi.org/10.1093/nar/gkl971). 12
- [144] Lee Y-M and Lim C. Physical basis of structural and catalytic Zn-binding sites in proteins. *J Mol Biol*, 379(3):545–53, Jun 2008. doi: [10.1016/j.jmb.2008.04.004](https://doi.org/10.1016/j.jmb.2008.04.004). 12
- [145] Fuse N, Maiti T, Wang B, Porter J A, et al. Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc Natl Acad Sci U S A*, 96(20):10992–9, Sep 1999. doi: [10.1073/pnas.96.20.10992](https://doi.org/10.1073/pnas.96.20.10992). 12, 53
- [146] Day E S, Wen D, Garber E A, Hong J, et al. Zinc-dependent structural stability of human Sonic hedgehog. *Biochemistry*, 38(45):14868–80, Nov 1999. doi: [10.1021/bi9910068](https://doi.org/10.1021/bi9910068). 12, 42, 53
- [147] Bochtler M, Odintsov S G, Marcyjaniak M, and Sabala I. Similar active sites in lysostaphins and D-Ala-D-Ala metallopeptidases. *Protein Sci*, 13(4):854–61, Apr 2004. doi: [10.1110/ps.03515704](https://doi.org/10.1110/ps.03515704). 12, 44, 50, 53, 55
- [148] Bussiere D E, Pratt S D, Katz L, Severin J M, et al. The Structure of VanX Reveals a Novel Amino-Dipeptidase Involved in Mediating Transposon-Based Vancomycin Resistance. *Mol Cell*, 2(1):75–84, 1998. ISSN 1097-2765. doi: [10.1016/S1097-2765\(00\)80115-X](https://doi.org/10.1016/S1097-2765(00)80115-X). 12, 36
- [149] Marcyjaniak M, Odintsov S G, Sabala I, and Bochtler M. Peptidoglycan amidase MepA is a LAS metallopeptidase. *J Biol Chem*, 279(42):43982–9, Oct 2004. doi: [10.1074/jbc.M406735200](https://doi.org/10.1074/jbc.M406735200). 36

- [150] Firczuk M, Mucha A, and Bochtler M. Crystal structures of active LytM. *J Mol Biol*, 354(3):578–90, Dec 2005. doi: [10.1016/j.jmb.2005.09.082](https://doi.org/10.1016/j.jmb.2005.09.082). 36
- [151] Charlier P, Wery J P, Dideberg O, and Frère J M. Streptomyces albus g d-ala-d-ala carboxypeptidase, 2006. ISBN 9780470028636. doi: [10.1002/0470028637.met032](https://doi.org/10.1002/0470028637.met032). 36
- [152] Firczuk M and Bochtler M. Folds and activities of peptidoglycan amidases. *FEMS Microbiol Rev*, 31(6):676–91, Nov 2007. doi: [10.1111/j.1574-6976.2007.00084.x](https://doi.org/10.1111/j.1574-6976.2007.00084.x).
- [153] Korndörfer I P, Kanitz A, Danzer J, Zimmer M, et al. Structural analysis of the L-alanoyl-D-glutamate endopeptidase domain of Listeria bacteriophage endolysin Ply500 reveals a new member of the LAS peptidase family. *Acta Crystallogr D Biol Crystallogr*, 64(Pt 6):644–50, Jun 2008. doi: [10.1107/S09074444908007890](https://doi.org/10.1107/S09074444908007890). 12, 36, 52
- [154] Crooks G E, Hon G, Chandonia J M, and Brenner S E. Weblogo: a sequence logo generator. *Genome Res*, 14(6):1188–90, Jun 2004. doi: [10.1101/gr.849004](https://doi.org/10.1101/gr.849004). 13
- [155] Wallace A C, Laskowski R A, and Thornton J M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, 8(2):127–34, Feb 1995. doi: [10.1093/protein/8.2.127](https://doi.org/10.1093/protein/8.2.127). 14
- [156] Laskowski R A and Swindells M B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*, 51(10):2778–86, Oct 2011. doi: [10.1021/ci200227u](https://doi.org/10.1021/ci200227u). 14
- [157] Kirberger M, Wang X, Deng H, Yang W, et al. Statistical analysis of structural characteristics of protein Ca²⁺-binding sites. *J Biol Inorg Chem*, 13(7):1169–81, Sep 2008. doi: [10.1007/s00775-008-0402-7](https://doi.org/10.1007/s00775-008-0402-7). 15
- [158] Maun H R, Wen X, Lingel A, de Sauvage F J, et al. Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *J Biol Chem*, 285(34):26570–80, Aug 2010. doi: [10.1074/jbc.M110.112284](https://doi.org/10.1074/jbc.M110.112284). 16, 35, 38, 56, 57, 62
- [159] Karplus M and McCammon J A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 9(9):646–52, Sep 2002. doi: [10.1038/nsb0902-646](https://doi.org/10.1038/nsb0902-646). 19
- [160] Hehre W J, Radom L, Schleyer P, and Pople J A. Ab initio molecular orbital theory. New York, 1986. doi: [10.1002/jcc.540070314](https://doi.org/10.1002/jcc.540070314). 19
- [161] Burkert U and Allinger N L. Molecular mechanics. Washington, D.C., 1982. doi: [10.1002/jcc.540040420](https://doi.org/10.1002/jcc.540040420). 19
- [162] Warshel A and Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*, 103(2):227–249, 1976. ISSN 0022-2836. doi: [10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9). 20
- [163] Gogonea V, Suárez D, van der Vaart A, and Merz K M Jr. New developments in applying quantum mechanics to proteins. *Curr Opin Struct Biol*, 11(2):217–23, Apr 2001. doi: [10.1016/S0959-440X\(00\)00193-7](https://doi.org/10.1016/S0959-440X(00)00193-7).
- [164] Senn H M and Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed Engl*, 48(7):1198–229, 2009. doi: [10.1002/anie.200802019](https://doi.org/10.1002/anie.200802019). 20

- [165] McCammon J A, Gelin B R, and Karplus M. Dynamics of folded proteins. *Nature*, 267(5612):585–90, Jun 1977. doi: [10.1038/267585a0](https://doi.org/10.1038/267585a0). 20
- [166] Pronk S, Páll S, Schulz R, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–54, Apr 2013. doi: [10.1093/bioinformatics/btt055](https://doi.org/10.1093/bioinformatics/btt055). 20, 36, 62
- [167] Huber G A and McCammon J A. Browndye: A software package for brownian dynamics. *Comput Phys Commun*, 181(11):1896–1905, Nov 2010. doi: [10.1016/j.cpc.2010.07.022](https://doi.org/10.1016/j.cpc.2010.07.022). 20, 37
- [168] Hockney R W, Goel S P, and Eastwood J W. Quiet high-resolution computer models of a plasma. *J. Comput. Phys*, 14(2):148 – 158, 1974. ISSN 0021-9991. doi: [10.1016/0021-9991\(74\)90010-2](https://doi.org/10.1016/0021-9991(74)90010-2). 21
- [169] Adcock S A and McCammon J A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*, 106(5):1589–615, May 2006. doi: [10.1021/cr040426m](https://doi.org/10.1021/cr040426m). 21
- [170] Leach A. *Molecular Modelling: Principles and Applications*. second edition. Prentice Hall, ISBN-10: 0582382106, March 2001. 22
- [171] Ponder J W and Case D A. Force fields for protein simulations. *Adv Protein Chem*, 66:27–85, 2003. doi: [10.1016/S0065-3233\(03\)66002-X](https://doi.org/10.1016/S0065-3233(03)66002-X). 22
- [172] MacKerell Jr. A D, Bashford D, Bellot M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, 102(18): 3586–3616, 1998. doi: [10.1021/jp973084f](https://doi.org/10.1021/jp973084f). 22
- [173] Jorgensen W L and Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc*, 110(6):1657–1666, 1988. doi: [10.1021/ja00214a001](https://doi.org/10.1021/ja00214a001). 22
- [174] Jorgensen W L, Maxwell D S, and Tirado-Rives J. Development and testing of the OPLS All-Atom Force Field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc*, 118(45):11225–11236, 1996. doi: [10.1021/ja9621760](https://doi.org/10.1021/ja9621760). 22
- [175] Scott W R P, Hünenberger P H, Tironi I G, et al. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, 103(19):3596–3607, 1999. doi: [10.1021/jp984217f](https://doi.org/10.1021/jp984217f). 22
- [176] van Gunsteren W F, Billeter S R, Eising A A, et al. *Biomolecular simulation: The GROMOS96 manual and user guide*. Zürich, Switzerland: Hochschulverlag AG an der ETH Zürich., 1996. 22
- [177] Daura X, Mark A E, and van Gunsteren W F. Parametrization of aliphatic CH_n united atoms of GROMOS96 force field. *J. Comput. Chem*, 19(5):535–547, 1998. ISSN 1096-987X. doi: [10.1002/\(SICI\)1096-987X\(19980415\)19:5<535::AID-JCC6>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1096-987X(19980415)19:5<535::AID-JCC6>3.0.CO;2-N). 22, 23
- [178] van Gunsteren W F and Berendsen H J. *Gromos-87 manual*. Biomos BV Nijenborgh 4, 9747 AG Groningen, 1987. 22
- [179] Halgren T A. Representation of van der Waals (vdW) interactions in molecular

- mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.*, 114:7827–7843, 1992. doi: [10.1021/ja00046a032](https://doi.org/10.1021/ja00046a032). 23
- [180] Project E, Nachliel E, and Gutman M. Parameterization of Ca^{+2} -protein interactions for molecular dynamics simulations. *J Comput Chem*, 29(7):1163–9, May 2008. doi: [10.1002/jcc.20876](https://doi.org/10.1002/jcc.20876). 23, 36
- [181] Hess B, Kutzner C, van der Spoel D, and Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput*, 4(3):435–447, February 2008. doi: [10.1021/ct700301q](https://doi.org/10.1021/ct700301q). 24
- [182] Ewald P P. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys*, 369(3):253–287, 1921. ISSN 1521-3889. doi: [10.1002/andp.19213690304](https://doi.org/10.1002/andp.19213690304). 24
- [183] Darden T, York D, and Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys*, 98(12):10089–10092, 1993. ISSN 10897690. doi: [10.1063/1.464397](https://doi.org/10.1063/1.464397). 24
- [184] Essmann U, Perera L, Berkowitz M L, et al. A smooth particle mesh Ewald method. *J. Chem. Phys*, 103(19):8577–8593, 1995. ISSN 10897690. doi: [10.1063/1.470117](https://doi.org/10.1063/1.470117). 24
- [185] Berendsen H C, Postma J P, van Gunsteren W F, and Hermans J. Interaction models for water in relation to protein hydration. In: *Interaction Forces*, Pullman, B. ed. . D. Reidel Publishing Company Dordrecht, 1981. 25
- [186] Berendsen H J, Grigera J R, and Straatsma T P. The missing term in effective pair potentials. *J. Phys. Chem*, 91(24):6269–6271, 1987. doi: [10.1021/j100308a038](https://doi.org/10.1021/j100308a038). 25
- [187] Hess B and van der Vegt N F. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J. Phys. Chem. B*, 110(35):17616–17626, August 2006. doi: [10.1021/jp0641029](https://doi.org/10.1021/jp0641029). 25
- [188] Lazaridis T and Karplus M. Thermodynamics of protein folding: a microscopic view. *Biophys. Chem*, 100(1-3):367–395, 2003. ISSN 0301-4622. doi: [10.1016/S0301-4622\(02\)00293-4](https://doi.org/10.1016/S0301-4622(02)00293-4). 26
- [189] Honig B and Nicholls A. Classical electrostatics in biology and chemistry. *Science*, 268(5214):1144–1149, 1995. doi: [10.1126/science.7761829](https://doi.org/10.1126/science.7761829). 26
- [190] Still W C, Tempczyk A, Hawley R C, and Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112(16): 6127–6129, 1990. doi: [10.1021/ja00172a038](https://doi.org/10.1021/ja00172a038). 26
- [191] Hermann R B. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem*, 76(19):2754–2759, 1972. doi: [10.1021/j100663a023](https://doi.org/10.1021/j100663a023). 26
- [192] Fogolari F, Brigo A, and Molinari H. The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit*, 15(6):377–392, 2002. ISSN 1099-1352. doi: [10.1002/jmr.577](https://doi.org/10.1002/jmr.577). 26
- [193] Shestakov A I, Milovich J L, and Noy A. Solution of the Nonlinear Poisson-Boltzmann Equation Using Pseudo-transient Continuation and the Finite Element Method. *Journal of Colloid and Interface Science*, 247(1):62–79, 2002. ISSN 0021-9797. doi:

- [10.1006/jcis.2001.8033](https://doi.org/10.1006/jcis.2001.8033). 26
- [194] Lu B, Zhang D, and McCammon J A. Computation of electrostatic forces between solvated molecules determined by the Poisson–Boltzmann equation using a boundary element method. *J. Chem. Phys*, 122(21):214102–214107, 2005. ISSN 10897690. doi: [10.1063/1.1924448](https://doi.org/10.1063/1.1924448). 26
- [195] Baker N A, Sept D, Joseph S, Holst M J, and McCammon J A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98:10037–41, 2001. doi: [10.1073/pnas.181342398](https://doi.org/10.1073/pnas.181342398). 26, 29, 37, 63, 68
- [196] Berendsen H J, Postma J P, van Gunsteren W F, DiNola A, and Haak J R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys*, 81(8):3684–3690, 1984. ISSN 10897690. doi: [10.1063/1.448118](https://doi.org/10.1063/1.448118). 27
- [197] S Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys*, 52(2):255–268, 1984. doi: [10.1080/00268978400101201](https://doi.org/10.1080/00268978400101201). 27
- [198] Hoover W G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985. doi: [10.1103/PhysRevA.31.1695](https://doi.org/10.1103/PhysRevA.31.1695). 27
- [199] Bussi G, Donadio D, and Parrinello M. Canonical sampling through velocity rescaling. *J. Chem. Phys*, 126(1):014101–014107, 2007. ISSN 10897690. doi: [10.1063/1.2408420](https://doi.org/10.1063/1.2408420). 27
- [200] Parrinello M and Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*, 52(12):7182–7190, 1981. ISSN 10897550. doi: [10.1063/1.328693](https://doi.org/10.1063/1.328693). 27
- [201] Nosé S and Klein M L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys*, 50(5):1055–1076, 1983. doi: [10.1080/00268978300102851](https://doi.org/10.1080/00268978300102851). 27
- [202] Alsallaq R and Zhou H-X. Prediction of protein-protein association rates from a transition-state theory. *Structure*, 15(2):215–224, 2007. ISSN 0969-2126. doi: [10.1016/j.str.2007.01.005](https://doi.org/10.1016/j.str.2007.01.005). 28
- [203] Schreiber G, Haran G, and Zhou H-X. Fundamental aspects of protein–protein association kinetics. *Chem. Rev*, 109(3):839–860, 2009. doi: [10.1021/cr800373w](https://doi.org/10.1021/cr800373w). PMID: 19196002. 28
- [204] Smoluchowski M V. Versuch einer mathematischen theorie der koagulationskinetik kolloider loeschungen. *Z Phys Chem*, 92:129–168, 1917. 28
- [205] Northrup S H, Allison S A, and McCammon J A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys*, 80(4):1517–1524, 1984. ISSN 10897690. doi: [10.1063/1.446900](https://doi.org/10.1063/1.446900). 28, 30
- [206] Huber G A and Kim S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J*, 70(1):97 – 110, 1996. ISSN 0006-3495. doi: [10.1016/S0006-3495\(96\)79552-8](https://doi.org/10.1016/S0006-3495(96)79552-8). 28
- [207] Gabdoulline R R and Wade R C. Biomolecular diffusional association. *Current Opinion in Structural Biology*, 12(2):204 – 213, 2002. ISSN 0959-440X. doi: [10.1016/S0959-440X\(02\)00311-1](https://doi.org/10.1016/S0959-440X(02)00311-1). 29

- [208] Elcock A H. Molecular simulations of diffusion and association in multimacromolecular systems. *Numerical Comput. Methods, D: Methods Enzymol*, 383:166 – 198, 2004. ISSN 0076-6879. doi: [10.1016/S0076-6879\(04\)83008-8](https://doi.org/10.1016/S0076-6879(04)83008-8). 29
- [209] Elcock A H, Sept D, and McCammon J A. Computer simulation of protein–protein interactions. *J Phys Chem B*, 105(8):1504–1518, 2001. doi: [10.1021/jp003602d](https://doi.org/10.1021/jp003602d). 29
- [210] Lin J and Beratan D N. Simulation of Electron Transfer between Cytochrome c2 and the Bacterial Photosynthetic Reaction Center: Brownian Dynamics Analysis of the Native Proteins and Double Mutants. *J Phys Chem B*, 109(15):7529–7534, 2005. doi: [10.1021/jp045417w](https://doi.org/10.1021/jp045417w). PMID: 16851864.
- [211] Spaar A, Dammer C, Gabdoulline R R, Wade R C, and Helms V. Diffusional Encounter of Barnase and Barstar. *Biophys. J*, 90(6):1913–1924, 2006. ISSN 0006-3495. doi: [10.1529/biophysj.105.075507](https://doi.org/10.1529/biophysj.105.075507). 29
- [212] Fernandes M X and García de la Torre J. Brownian dynamics simulation of rigid particles of arbitrary shape in external fields. *Biophys. J*, 83(6):3039 –3048, 2002. ISSN 0006-3495. doi: [10.1016/S0006-3495\(02\)75309-5](https://doi.org/10.1016/S0006-3495(02)75309-5). 29
- [213] Beard D A and Schlick T. Unbiased rotational moves for rigid-body dynamics. *Biophys. J*, 85(5):2973–2976, 2003. ISSN 0006-3495. doi: [10.1016/S0006-3495\(03\)74717-1](https://doi.org/10.1016/S0006-3495(03)74717-1). 29
- [214] Luty B A, McCammon J A, and Zhou H-X. Diffusive reaction rates from Brownian dynamics simulations: Replacing the outer cutoff surface by an analytical treatment. *J. Comput. Phys*, 97(8):5682–5686, 1992. ISSN 10897690. doi: [10.1063/1.463777](https://doi.org/10.1063/1.463777). 30
- [215] Hubbard S and Thornton J M. NACCESS: program for calculating accessibilities. *Department of Biochemistry and Molecular Biology, University College of London.*, 1992. 33, 63
- [216] Lee B and Richards F M. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol*, 55(3):379 – IN4, 1971. ISSN 0022-2836. doi: [10.1016/0022-2836\(71\)90324-X](https://doi.org/10.1016/0022-2836(71)90324-X). 33
- [217] McDonald I K and Thornton J M. Satisfying hydrogen bonding potential in proteins. *J Mol Biol*, 238(5):777–93, May 1994. doi: [10.1006/jmbi.1994.1334](https://doi.org/10.1006/jmbi.1994.1334). 33, 63
- [218] Lander A D, Q N, and Wan F Y M. Do Morphogen Gradients Arise by Diffusion? *Dev Cell*, 2(6):785 – 796, 2002. ISSN 1534-5807. doi: [10.1016/S1534-5807\(02\)00179-X](https://doi.org/10.1016/S1534-5807(02)00179-X). 35
- [219] Eldar A, Rosin D, Shilo B Z, and Barkai N. Self-Enhanced Ligand Degradation Underlies Robustness of Morphogen Gradients. *Dev Cell*, 5(4):635 – 646, 2003. ISSN 1534-5807. doi: [10.1016/S1534-5807\(03\)00292-2](https://doi.org/10.1016/S1534-5807(03)00292-2). 35
- [220] Incardona J P, Lee J H, Robertson C P, Enga K, et al. Receptor-mediated endocytosis of soluble and membrane-tethered Sonic hedgehog by Patched-1. *Proc Natl Acad Sci U S A*, 97(22):12044–12049, 2000. doi: [10.1073/pnas.220251997](https://doi.org/10.1073/pnas.220251997). 35
- [221] Krivov G G, Shapovalov M V, and Dunbrack R L Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 77(4):778–95, Dec 2009. doi:

- 10.1002/prot.22488. 36, 62
- [222] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0, <http://www.R-project.org>. 37, 63
- [223] Jakushev S and Hoffmann D. A Novel Algorithm for Macromolecular Epitope Matching. *Algorithms*, 2(1):498–517, 2009. ISSN 1999-4893. doi: 10.3390/a2010498. 37, 51, 64, 74
- [224] Dolinsky T J, Nielsen J E, McCammon J A, and Baker N A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–7, Jul 2004. doi: 10.1093/nar/gkh381. 37, 63
- [225] Wang J, Cieplak P, and Kollman P A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21(12):1049–1074, September 2000. ISSN 0192-8651. doi: 10.1002/1096-987x(200009)21:12<%3C1049::aid-jcc3%3E3.o.co;2-f. 37, 63
- [226] Søndergaard C R, Olsson M H M, Rostkowski M, and Jensen J H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput*, 7(7):2284–2295, 2011. doi: 10.1021/ct200133y. 38
- [227] Notredame C, Higgins D G, and Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, 302(1):205–17, Sep 2000. doi: 10.1006/jmbi.2000.4042. 38
- [228] Gouy M, Guindon S, and Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, 27(2):221–4, Feb 2010. doi: 10.1093/molbev/msp259. 38
- [229] Irwin J J, Sterling T, Mysinger M M, Bolstad E S, and Coleman R G. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model*, 52(7):1757–68, Jul 2012. doi: 10.1021/ci3001277. 38
- [230] Pettersen E F, Goddard T D, Huang C C, Couch G S, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–12, Oct 2004. doi: 10.1002/jcc.20084. 38
- [231] Trott O and Olson A J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, 31(2):455–61, Jan 2010. doi: 10.1002/jcc.21334. 38
- [232] Kumar S, Balczarek K A, and Lai Z C. Evolution of the hedgehog Gene Family. *Genetics*, 142(3):965–972, 1996. URL <http://www.genetics.org/content/142/3/965.abstract>. 52
- [233] Eijsink V G H, Matthews B W, and Vriend G. The role of calcium ions in the stability and instability of a thermolysin-like protease. *Protein Sci*, 20(8):1346–55, Aug 2011. doi: 10.1002/pro.670. 55
- [234] Li Y, Zhang H, Litingtung Y, and Chiang C. Cholesterol modification restricts the

- spread of Shh gradient in the limb bud. *Proc Natl Acad Sci U S A*, 103(17):6548–53, Apr 2006. doi: [10.1073/pnas.0600124103](https://doi.org/10.1073/pnas.0600124103). 55, 59
- [235] Wright C S, Li S C, and Rastinejad F. Crystal structure of human GM2-activator protein with a novel beta-cup topology. *J Mol Biol*, 304(3):411–22, Dec 2000. doi: [10.1006/jmbi.2000.4225](https://doi.org/10.1006/jmbi.2000.4225). 55
- [236] Rossmann M, Schultz-Heienbrok R, Behlke J, Remmel N, et al. Crystal structures of human saposins C and D: implications for lipid recognition and membrane interactions. *Structure*, 16(5):809–17, May 2008. doi: [10.1016/j.str.2008.02.016](https://doi.org/10.1016/j.str.2008.02.016). 55
- [237] Rebolledo-Rios R, Bandari S, Wilms C, Jakushev S, et al. Signaling domain of Sonic Hedgehog as cannibalistic calcium-regulated zinc-peptidase. (in revision), 2014. 56
- [238] Ma G, Yu J, Xiao J, Chan D, et al. Indian hedgehog mutations causing brachydactyly type A1 impair Hedgehog signal transduction at multiple levels. *Cell Res*, 21(9):1343–57, Sep 2011. doi: [10.1038/cr.2011.76](https://doi.org/10.1038/cr.2011.76). 58
- [239] Boshier S K and Warren R L. Very low calcium content of cochlear endolymph, an extracellular fluid. *Nature*, 273(5661):377–8, 1978. ISSN 0028-0836. doi: [10.1038/273377a0](https://doi.org/10.1038/273377a0). 59
- [240] I Silver. Microelectrode studies on the acid microenvironment beneath adherent macrophages and osteoclasts. *Exp Cell Res*, 175(2):266–276, 1988. doi: [10.1016/0014-4827\(88\)90191-7](https://doi.org/10.1016/0014-4827(88)90191-7). 59
- [241] Egelman D M and Montague P R. Calcium Dynamics in the Extracellular Space of Mammalian Neural Tissue. *BIOPHYS. J*, 76:1856–1867, 1999. doi: [10.1.1.39.604](https://doi.org/10.1.1.39.604). 59
- [242] Rusakov D A and Fine A. Extracellular Ca²⁺ Depletion Contributes to Fast Activity-Dependent Modulation of Synaptic Transmission in the Brain. *Neuron*, 37(2):287 – 297, 2003. ISSN 0896-6273. doi: [10.1016/S0896-6273\(03\)00025-4](https://doi.org/10.1016/S0896-6273(03)00025-4). 59
- [243] Pepinsky R B, Rayhorn P, Day E S, Dergay A, et al. Mapping sonic hedgehog-receptor interactions by steric interference. *J Biol Chem*, 275(15):10995–1001, Apr 2000. doi: [10.1074/jbc.275.15.10995](https://doi.org/10.1074/jbc.275.15.10995). 62, 77
- [244] Chothia C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol*, 105(1):1 – 12, 1976. ISSN 0022-2836. doi: [10.1016/0022-2836\(76\)90191-1](https://doi.org/10.1016/0022-2836(76)90191-1). 63
- [245] Chothia C and Janin J. Principles of protein-protein recognition. *Nature*, 256(5520):705–8, Aug 1975. doi: [10.1038/256705a0](https://doi.org/10.1038/256705a0). 65
- [246] Fukuda M and Mikoshiba K. Mechanism of the calcium-dependent multimerization of synaptotagmin VII mediated by its first and second C2 domains. *J Biol Chem*, 276(29):27670–6, Jul 2001. doi: [10.1074/jbc.M100851200](https://doi.org/10.1074/jbc.M100851200). 74
- [247] Martínez-Gil M, Romero D, Kolter R, and Espinosa-Urgel M. Calcium causes multimerization of the large adhesin LapF and modulates biofilm formation by *Pseudomonas putida*. *J Bacteriol*, 194(24):6782–9, Dec 2012. doi: [10.1128/JB.01094-12](https://doi.org/10.1128/JB.01094-12). 74
- [248] Capila I, Hernandez M J, Mo Y D, Mealy T R, et al. Annexin V-Heparin Oligosaccharide Complex Suggests Heparan Sulfate-Mediated Assembly on Cell Surfaces. *Structure*, 9(1):57–64, 2001. ISSN 0969-2126. doi: [10.1016/S0969-2126\(00\)00549-9](https://doi.org/10.1016/S0969-2126(00)00549-9). 75

-
- [249] Shao C, Zhang F, Kemp M M, Linhardt R J, et al. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J Biol Chem*, 281(42):31689–95, Oct 2006. doi: [10.1074/jbc.M604502200](https://doi.org/10.1074/jbc.M604502200). 75
- [250] Fromm J R, Hileman R E, Caldwell E E O, Weiler J M, and Linhardt R J. Pattern and Spacing of Basic Amino Acids in Heparin Binding Sites. *Arch Biochem Biophys*, 343(1):92–100, 1997. ISSN 0003-9861. doi: [10.1006/abbi.1997.0147](https://doi.org/10.1006/abbi.1997.0147). 75
- [251] Lortat-Jacob H, Grosdidier A, and Imberty A. Structural diversity of heparan sulfate binding domains in chemokines. *Proc Natl Acad Sci U S A*, 99(3):1229–34, Feb 2002. doi: [10.1073/pnas.032497699](https://doi.org/10.1073/pnas.032497699). 75
- [252] Gandhi N S, Coombe D R, and Mancera R L. Platelet endothelial cell adhesion molecule 1 (PECAM-1) and its interactions with glycosaminoglycans: 1. Molecular modeling studies. *Biochemistry*, 47(17):4851–62, Apr 2008. doi: [10.1021/bi702455e](https://doi.org/10.1021/bi702455e). 75
- [253] Taubenheim J. Interactions between Hedgehog proteins and heparan sulfate/heparin. Master's thesis, University of Duisburg-Essen, 2013. 76

Journal Articles

- [1] **Rebollido-Rios R**, Bandari S, Wilms C, Jakushev S, et al. Signaling domain of Sonic Hedgehog as cannibalistic calcium-regulated zinc-peptidase. (accepted), 2014.
- [2] Rolo-Naranjo A, **Rebollido-Rios R**, Melchor-Rodriguez K, and Codorniu-Hernandez E. Pseudo-phase portrait applied to pattern recognition in flavonoid-protein interactions. *Applied Mathematics and Computation*, 215(1):156–167, 2009. ISSN 0096-3003. doi: [10.1016/j.amc.2009.04.070](https://doi.org/10.1016/j.amc.2009.04.070).
- [3] **Rebollido-Rios R**, Martinez J, Aguilera Y, Melchor K, Koerner I, and Stegmann R. Microbial populations during composting process of organic fraction of municipal solid waste. *Applied Ecology and Environmental Research*, 6(3):61– 67, 2008. ISSN 1589-1623. URL http://www.ecology.kee.hu/pdf/0603_061067.pdf.

Posters

- [1] **Rebollido-Rios R**, Wilms C, Jakushev S, Dybowski J N, Hoffmann D. 2013. ISMB/ECCB. Berlin, Germany. *Molecular modeling of Sonic Hedgehog binding suggests novel regulatory mechanism*.
- [2] **Rebollido-Rios R**, Wilms C, Hoffmann D. 2012. Forschungstag der Medizinischen Fakultät. University of Duiburg-Essen. *Calcium ions involved in Hedgehog multimerization*.
- [3] **Rebollido-Rios R**, Wilms C, Hoffmann D. 2012. ECCB. Basel, Switzerland. *Molecular models suggest that calcium ions enhance Hedgehog multimerization*.
- [4] **Rebollido-Rios R**, Jakushev S, Hoffmann D. 2012. RECOMB 2012. Barcelona, Spain. *How does the calcium binding could affect the suspected protease activity of Sonic Hedgehog?*
- [5] **Rebollido-Rios R**, Hoffmann D. 2011. German Conference on Bioinformatics. Germany. *The role of calcium ions for the stability of Sonic Hedgehog proteins*.

Declarations

Erklärung:

Hiermit erkläre ich, gem. § 6 Abs. (2) g) der Promotionsordnung der Fakultäten für Biologie zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema “Computational study of Sonic hedgehog proteins: insights into the role of metal ions” zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Rocio Rebolledo-Rios befürworte und die Betreuung auch im Falle eines Weggangs, wenn nicht wichtige Gründe dem entgegenstehen, weiterführen werde.

Essen, den _____

Unterschrift eines Mitgliedes der Universität Duisburg-Essen

Erklärung:

Hiermit erkläre ich, gem. § 7 Abs. (2) d) + f) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient, bei der Abfassung der Dissertation nur die angegebenen Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Essen, den _____

Unterschrift des Doktoranden

Erklärung:

Hiermit erkläre ich, gem. § 7 Abs. (2) e) + g) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den _____

Unterschrift des Doktoranden