eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

TESIAREN ZUZENDARIAREN BAIMENA TESIA AURKEZTEKO

Fernando P. Cossío jaunak

Study of Protein-Protein Interactions Involved in Cancer Metastases and Autoimmune Diseases. Design of New Antimetastatic Molecules.

izenburua duen doktorego-tesiaren zuzendari naizenak, tesia aurkezteko baimena ematen dut, defendatua izateko baldintzak betetzen dituelako. **Eider San Sebastian** doktoregai andereak egin du aipaturiko tesia, **Kimika Organikoa I** sailean.

_____, _____ (e)ko _____ ren _____ a

TESIAREN ZUZENDARIA

Izpta.: Fernando P. Cossío

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

TESIAREN ZUZENDARIAREN BAIMENA TESIA AURKEZTEKO

Xabier López Pestaña jaunak

Study of Protein-Protein Interactions Involved in Cancer Metastases and Autoimmune Diseases. Design of New Antimetastatic Molecules.

izenburua duen doktorego-tesiaren zuzendari naizenak, tesia aurkezteko baimena ematen dut, defendatua izateko baldintzak betetzen dituelako. **Eider San Sebastian** doktoregai andereak egin du aipaturiko tesia, **Kimika Organikoa I** sailean.

_____, _____ (e)ko _____ ren _____ a

TESIAREN ZUZENDARIA

Izpta.: Xabier López Pestaña

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

SAILAREN ADOSTASUNA

Kimika Organiko I Saileko Kontseiluak, _____(e)ko bileran,

Study of Protein-Protein Interactions Involved in Cancer Metastases and Autoimmune Diseases. Design of New Antimetastatic Molecules.

izenburua duen doktorego-tesia aurkeztearen alde dagoela adierazi du. **Fernando P. Cossío** eta **Xabier López Pestaña** jaunen zuzendaritzapean egin den tesi hori **Eider San Sebastian** andereak aurkeztu du sail honetan.

_____, _____ (e)ko _____ ren _____ a

O.E. SAILEKO ZUZENDARIA SAILEKO IDAZKARIA

Izpta.: _____ Izpta.: _____

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Study of Protein-Protein Interactions Involved in Cancer Metastases and Autoimmune Diseases. Design of New Antimetastatic Molecules.

TESI DOKTORALA

Eider San Sebastian Larzabal

Donostia 2007

Tesiaren Laburpena

Minbizia eta gaixotasun autoinmuneetan parte hartzen duten proteina-proteina elkarrekintzen azterketa. Minbiziaren aurkako molekula berri txikien diseinua.

LFA-1 eta ICAM-1 proteinen elkarrekintza garrantzia handikoa da bai minbiziaren hedapenean bai eta gaixotasun autoinmuneetan. Hauen haurkako tratamenduetariko bat, molekula txiki antagonisten ekintzan oinarritzen da. Tesi proiektu honetan zehar, elkarrekintza honen afinitatearen mekanismoak ikertu dira, eta aminoazido garrantzitsuenak izendatu. Azken honek, antagonista txiki berrien diseinua erraztu digu. Lortutako emaitzak eta erabilitako teknikak ondorengo parrafoetan laburtzen dira.

1) Kimika teoriko eta konputazionala: LFA-1/ICAM-1 elkarrekintzaren azterketa mekanika kuantikoaren bitartez.

o Metal Ion Dependent Adhesion Site: Kalkulu hauen emaitzek ondorengoa adierazten dute: i) I-domeinuan dagoen metalari koordinaturik dauden aminoazidoen ezaugarri kimikoek era oso garrantzitsuan kontrolatzen dute ligandoarekiko afinitatea. ii) Manganeso ioiak elkarrekintza hobetzen du, Zink edo Magnesio metalek baino gehiago. Kaltzio ioiak ordea, ICAM-1ekiko afinitatea gutxitzen du. iii) Inguru hidrofobiko batek eta nitronato eta karboxilo taldeak dituzten ligandoek, ICAM-1ekiko afinitatea haunditzen dute.

o Manganesoaren spin egoera: Mangenesoaren trantsizio metal egoerak, beharrezkoa egiten du spin egoera desberdinen energia erlatiboa berkalkulatzea, inguru proteikoetan daukan egoera zehazteko Tesi honetan haurreko ataleko kalkulu teorikoak berregin dira "Difussion Monte Carlo" izeneko metodologia erabiliz. Lortutako emaitzek honakoa adierazten dute: i) Manganesoaren spin egoera inguru proteikoetan 5/2 da, hau da, spin altuko egoera; ii) Spina zenbat eta altuagoa izan, ligandoarekiko afinitatea gutxitzen da; iii) DFT mailako kalkulu teorikoek spin egoera desberdinak ondo deskribatzen dituzte erlatiboki, baina Diffusion Monte Carlo metodoak, horrez gain, emaitza experimentalekin bat datozen baloreak sortzen ditu.

2) Kimika teoriko eta konputazionala: LFA-1/ICAM-1 elkarrekintzaren azterketa mekanika klasikoaren bitartez.

o Afinitate gutxi, erdi eta altuko LFA-1/ICAM-1 komplexuak: Metodologia konputazionalak erabiliz, afinitate gutxiko eta altuko LFA-1/ICAM-1 konplexuak sortu dira tesi honetan, eta ezaguna den afinitate erdiko komplexuarekin konparatu. Dinamika molekularrak eta Energia Librea aminoazido bakoitzaren kontribuzioetan dekonposatzen duen protokolo bat jarraituz, honako ondorioetara iritsi gara: LFA-1 integrinaren aktibazioak i) vdW elkarrekintzen garrantziaren txikitzearekin batera datorrela, ii) I-domeinu inaktiboak ez duela complexu egonkor bat sortzen ICAM-1 ekin, iii) elkarrekintza kopurua txikitzen dela, baina beroien efizientzia haunditu, iv) elkarrekintza garrantzitsuenetan hurrengo aminoazidoek hartzen dute parte: GLU34, THR35, PRO37, LYS39 eta MET64 ICAM-1en eta Mg katioia, GLU241,

THR243 y HSE264 I-domeinuan; v) aldaketa konformazional garrantzitsua ikusi da $\alpha 7$ elizean.

o LFA-1 / HC0303 konplexua: HC0303 izeneko inhibidore sintetikoa LFA-1en kontrato antagonista hoberena zen tesi hau hasi aurretik. Tesi honetan egindako kalkuluek, bi inhibidore famili berrien diseinua gidatu dute. "in vivo" eta "in vitro" experimentuek erakutsi dute, bi famili berri hauen parte diren molekula gehienek, aktibitate antimetastasikoa hobetu dutela.

3) Erresonantzia Magnetiko Nuklearra: Inhibidoreen LFA-1en duten lotura gunea ezagutzen.

o Leku alosterikoa: RMNren bitartez eginiko ikerketek ondorengoa adierazten dute: I domeinua LFA-1 proteinatik isolatzen denean, "in vivo" egoeran dauzkan ezaugarriak galtzen dituela eta molekula txikien lotura gunea ikertzeko ez duela balio. Ondorio hauek, hurrengo arrazoiengan oinarritzen da: i) bai LFA-1 inhibidoreak bai eta beste integrinen inhibidoreak leku berdinera lotzen dira I-domeinu isolatuan; ii) elkarekintza mikromolar indarrekoa da kasu guztietan, baina in vivo eta "in vitro" emaitzek konposatu batzu besteak baina hobeak direna erakutsi zuten; iii) inhibidoreak, behin leku alosterikora lotuta, ez dira LFA-1/ICAM-1 elkarrekintza murrizteko gai, in vivo emaitzetan kontrakoa ikusi delarik.

Resumen de la Tesis

Estudio de las interacciones proteína-proteína involucradas en metástasis cancerosa y enfermedades autoinmunes. Diseño de nuevas moléculas antimetasatásicas.

La interación entre el dominio I de la proteína LFA-1 y su ligando natural ICAM-1 es clave en procesos de metástasis cancerosa y enfermedades autoinmunes, que pueden ser tratadas con moléculas de pequeño tamaño que evitan su unión. Durante la realización de esta tésis, se han estudiado exhaustivamente los mecanismos de regulación y los aminoácidos clave de esta unión, lo que ha permitido diseñar dos familias prometedoras de inhibidores. Los resultados obtenidos y las técnicas utilizadas se resumen ha continuación:

1) Química Teórica y Computacional: Estudio de la unión LFA-1/ICAM-1 por medio de herramientas mecánico cuánticas.

o Metal Ion Dependent Adhesion Site: Los resultados de esta tésis indican que i) La naturaleza de los aminoácidos que se unen al metal presente en el dominio I regulan de forma crítica la afinidad por el ligando. ii) El ión Manganeso favorece la unión a ICAM-1 en mayor medida que el Zinc o el Magnesio. El ión Calcio, reduce la afinidad por ICAM-1. iii) Un entorno hidrofóbico y iii) ligandos basados en grupos carboxílicos y nitrotanos, incrementan la afinidad por ICAM-1.

o Estado de multiplicidad del Manganeso: El carácter de metal de transición que caracteriza al ión Manganeso hizo necesario un estudio exhaustivo de la energía relativa de los diferentes estados de spín del catión para poder determinar su verdadero estado el entornos protéicos y su afinidad por ICAM-1. Cálculos teóricos usando la metodología conocida como Diffusion Monte Carlo indican que i) El Manganeso en entornos protéicos es un complejo de spin 5/2. ii) A mayor spin, menor es la afinidad por el ligando.iii) La teoría del funcional de la densidad (DFT) permite obtener una buena caracterización de los distinos estados de spin, pero los resultados obtenidos con el estudio de Diffusion Monte Carlo, generan datos ms equiparables con los experimentales.

2) Química Teórica y Computacional: Estudio de la unión LFA-1/ICAM-1 por medio de Dinámicas Moleculares.

o Complejos de LFA-1 con ICAM-1 de baja, media y alta afinidad: Herramientas computacionales nos han permitido crar los complejos de LFA-1 con ICAM-1 de baja y alta afinidad y compararlos con el de media afinidad ya existente. Por medio de Dinámicas Moleculares y un protocolo de descomposición por aminoácidos de la energía libre de unión nos ha permitido concluir que la activación de la integrina LFA-1 desde la conformación de baja afinidad a la de media y alta afinidad respectivamente, i) implica una disminución de las interacciones de vdW y un aumento considerable de las interacciones electrostáticas; ii) el dominio I en conformación inactiva no genera un complejo estable con ICAM-1, sugiriendo la necesidad de un cambio

conformacional previo a su unión a ICAM-1; iii) el número de interacciones disminuye, pero aumenta significativamente su eficacia; iv) las interacciones clave están protagonizadas por los residuos GLU34, THR35, PRO37, LYS39 y MET64 en ICAM-1 y por el catión Mg, GLU241, THR243 y HSE264 en el dominio I; v) se ha observado un cambio conformacional significativo en alfa-hélice del extremo C-terminal del dominio I.

o Complejo de LFA-1 con el inhibidor HC0303: El compuesto denominado HC0303 es el mejor candidato a antagonista de LFA-1 generado por miembros anteriores del grupo. Dinámicas Moleculares del complejo LFA-1/HC0303 indican que sólo una de las dos interaciones claves está teniendo lugar. Esta información nos ha permitido diseñar dos nuevas familias de inhibidores. Ambas familias fueron sometidas a estudios in vitro e in vivo, demostrando un aumento significativo de su capacidad antimetastásica.

3) Resonancia Magnética Nuclear: Caracterización de la unión de las nuevas familias de inhibidores a LFA-1.

o Unión al sitio alostérico: El estudio por RMN de la unión de diferentes inhibidores al dominio I tanto inactivo como activo indica que el dominio I aislado no representa adecuadamente a la integrina completa y no conviene utilizarlo para estudiar la unión a ligandos de pequeño peso molecular. Estas conclusiones vienen avaladas por los siguientes resultados obtenidos: i) tanto inhibidores de LFA-1 como de otras integrinas parecen unise inesperadamente a un bolsillo hidrofóbico conocido como IDAS; ii) la unión se da a escala micromolar, pero no se corresponde a lo observado en ensayos celulares y ensayos in vivo, que destacan unos compuetos frente a otros; iii) los antagonistas, una vez unidos al sitio alostérico, no son capaces de inhibir la unión del dominio I a ICAM-1, al contrario de lo observado in vivo.

To Dani

Acknowledgements

The work presented in this thesis book has been carried out in the Organic Chemistry Department I of the Chemistry Faculty in San Sebastian (University of the Basque Country) and was economically supported by Spanish Ministerio de Educación y Ciencia.

I want to dedicate the first lines to Prof. Fernando P. Cossío and Prof. Xabier López, cosupervisors of this thesis project. I am also glad to extend my gratitude to Prof. Jesus Mari Ugalde, who kindly hosted me in his group and let me use his human and non-human resources. The three of them may probably collect all the existing knowledge in Organic and Theoretical Chemistry. It has been a pleasure to learn from the three of them. Mila esker hiruroi.

I could also have included three other excellent researchers in the paragraph above: Prof. Roland H. Stote (Universite Louis Pasteur) and Prof. Annick Dejaegere (École Supérieure de Biotechnologie de Strasbourg) from Strasbourg, and Dr. Francisco Blanco (Centro Nacional de Investigaciones Oncológicas) from Madrid. I was given the chance to visit their research groups and the three of them devoted important part of their time and knowledge to our collaborative projects. Merci beaucoup Roland and Annick. Gracias, Paco.

Both in Madrid and Strasbourg I have met a lot of interesting people. Special thanks to Martin, Sergei, Polina, Elyette and Virginie, from Strasbourg. We have spent such good moments together that I must admit that I miss you all a bit now that all is over. Michael, Thoma, Cedric, Martina, Nicolas, Herve, Mark, Emma, Marie-Aude. Merci pour tout. Special thanks to Tahl Zimmerman, from Madrid. We have laughed to death together. It was a real pleasure to meet you. Pascal, Maria, Sonia, Jaska, Bego, Pablo, Susana, Simone, Javi, Richard, Dani T, Bernar: They all made my stay in Madrid most fun. Gracias a todos.

I have shared my days with the people in the Theoretical Chemistry group: Elena, Eli, Ioni, Julen, Mario and Txoni. And Txema and Inaki. I can not put down in words my gratitude for you all. And the people in the Organic Chemistry group: Ana, Aizpea (in the middle of the nothing), Dorleta, Yosu, Tamara, Eneko, Mirari and Begoña. Both groups are excellent examples of what a working team should be. My special thanks to them all. Milaesker.

Ama y aita, por fin acabo la tesis (esta vez es de verdad). Gracias por vuestro apoyo incondicional durante estos cinco años. Thanks to my brother, my sisters and my friends. They always showed interest for the possible progresses in my thesis, despite the fact that nobody really understood a word of what I was doing. But I really appreciated the interest shown. Mila esker danori.

This thesis work is dedicated to Dani. For all the reasons.

Computer Resources

The SGI/IZO-SGIker UPV/EHU (supported by the National Program for the Promotion of Human Resources within the National Plan of Scientific Research, Development and Innovations, Fondo Social Europeo and MCyT) is greatfully acknowledged for generous allocation of computational resources in Spain.

The Swedish Science Research Council (VR) and the Wood Ultrastructure Research Center (WURC) at SLU are also gratefully acknowledged as are the CNRS and the University Louis Pasteur.

Contents

ABBREVIATIONS

ADME = Absorption, Distribution, Metabolism and Excretion
B3LYP = Becke's 3-parameter functional, with Lee, Yang and Parr's corrections for correlation
CASSCF = Complete Active Space Self Consistent Field
CAM = Cell Adhesion Molecule
CC = Coupled-Cluster methods
CI = Configuration Interaction methods
CNIO = Centro Nacional de Investigaciones Oncologicas
CSP = Chemical Shift Perturbation
DFT = Density Functional Theory
DMC = Diffusion Monte Carlo
ECM = Extracellular Matrix
ECP = Effective Core Potential
ELISA = Enzyme-Linked Immunosorbent Assay experiments
FDPB = Finite Difference Poisson Boltzmann
FID = Free Induction Decay
FT = Fourier Transformation
GTO = Gaussian-Type Orbital
ICAM = Immunoglobulin Cell Adhesion Molecule
H.A. = High Affinity
HF = Hartree-Fock
HIS = Histidine
HIV = Human immunodeficiency virus
HSQC = Heteronuclear Single Quamtum Correlation
IDAS = I-domain Allosteric Site
I-domain = Inserted domain or von Willebrand factor A domain Ig = Immunoglobulin
IgSF = Immunoglobulin Superfamily
IPTG = isopropyl-I-D-thio-galactoside L.A. = Low Affinity
LDA = Local Density Approximation
LSDA = Local Spin Density Approximation
LFA-1 = Leukocyte Function Associated Protein 1
mAbs = monoclonal Antibodies
MadCAM-1 = Mucosal addressin cell adhesion molecule-1 (MAdCAM-1)
M.A. = Medium Affinity
MC = Monte Carlo
MD = Molecular Dynamics
MIDAS = Metal Ion Dependent Adhesion Site
MM/PBSA = Molecular mechanics- Poisson Bolzmann Surface Area
MPn = n^{th} order Møller-Plesset
MRCI = Multi-Referential Configuration Interaction
NMR = Nuclear Magnetic Resonance
N-CAM = Neural Cell Adhesion Molecule
PB = Poisson Boltzmann
PBC = Periodic Boundary Conditions

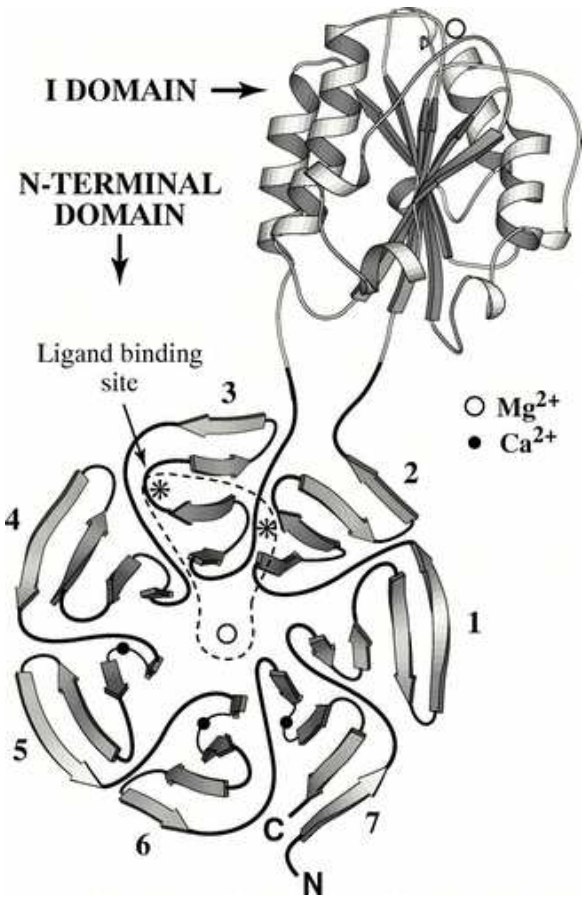
ABBREVIATIONS

PDB = Protein Data Bank
PECAM-1 = Patelet-endothelial cell adhesion molecule-1
PES = Potential Energy Surface
ppm = parts per million
RMS = Root Mean Squared
RMSD = Root Mean Squared Deviation
SAS = Solvent Accessible Surface
SCF = Self Consistent Field
SDS = Sodium dodecyl sulfate
SKBJ = Shared-exponent basis set of Stevens, Krauss, Basch and Jasien.
SPR = Surface Plasmon Resonance
STO = Slater-Type Orbital
UB = Urey-Bradley
VCAM-1 = Vascular Cell Adhesion Molecule-1
VLA-4 = Very late antigen 4

Chapter 1

General Introduction

Contents


I DOMAIN
N-TERMINAL DOMAIN
Ligand binding site
3
2
4
1
5
6
7
C
N
Mg2+
Ca2+

Introduction to Cell Adhesion Processes

Cells *in vivo* must interact both mechanically and chemically with their neighboring cells in order to form tissues and organs. In addition, most vertebrate cells cannot survive unless they are anchored to the extracellular matrix (ECM hereafter)[1]. These are known as cell-cell and cell-ECM interactions, respectively. Many of the major human diseases are either caused by defects in cell-ECM coordination, exacerbated by aberrant use of normal cell adhesive processes, or are potentially correctable by altering tissue structure or cell movement, as will be explained in the following paragraphs. To adhere either to other cells or to the ECM, most of the cells in vertebrates are provided with Cell Adhesion Molecules (CAMs hereafter) in their plasma membrane.

This thesis project was focused on the understanding of the main aspects of the interactions among specific CAMs, as well as on the application of this knowledge to the design of small molecular weight inhibitors that could serve as a therapy against the pathological processes these interactions are related to.

In the following sections the cell-ECM and cell-cell interactions will be briefly introduced to the reader[2], as well as the various CAMs that take part in them and the diseases or pathological processes CAMs are involved in. Finally, the reading will be focused on the particular proteins (CAMs) of interest for this thesis project as well as on what makes them a target for this study.

[1] Most cells in multicellular organisms are surrounded by a complex mixture of proteins, proteoglycans and, in the case of bones, mineral deposits, that constitute the extracellular matrix (ECM).

[2] For further reading on cell adhesion processes, see the following online text books [1, 2].

1.1 Cell-Extracellular Matrix and Cell-Cell Interactions

Adhesion of cells to the ECM and to other cells is done via CAMs. There are dozens of different CAMs (for review articles on CAMS see [3, 4, 5, 6]) but, at least so far, they fall into four major families (Figure 1.1):

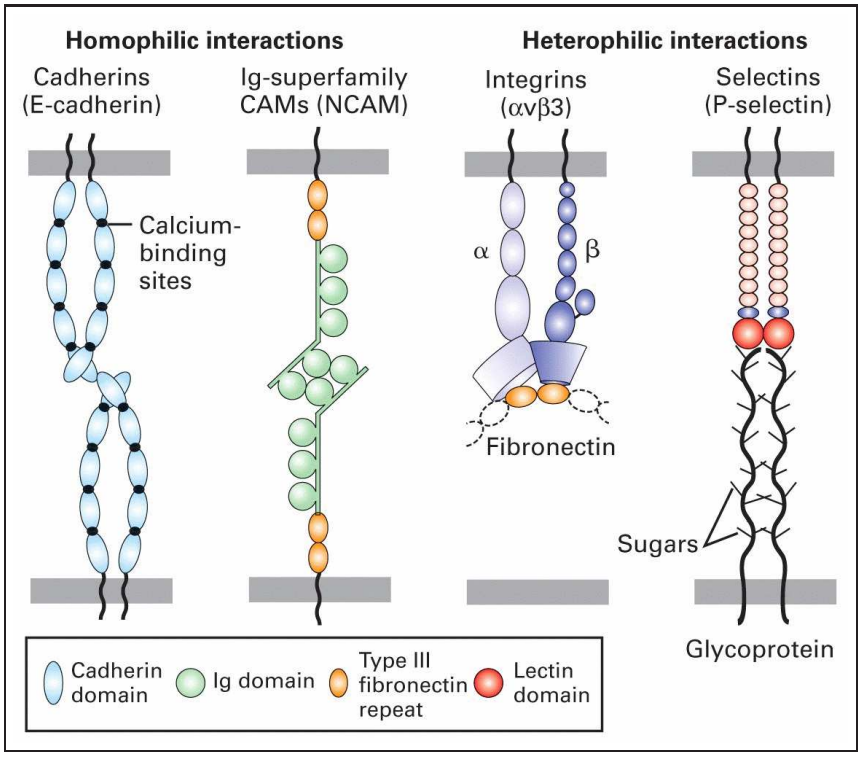


Fig. 1.1: The four major families of Cell Adhesion Molecules, and the types of interactions they form (Picture obtained from "Molecular Biology of Cells" [1]).

Cadherins: Cadherins (For a review see [7]) are adhesion molecules that are differentially expressed during development and in adult organs. Many cell types express multiple cadherin subclasses simultaneously (the combination differs with cell type), therefore, it can be inferred that the adhesion properties of individual cells are thus governed by varying the combinations of cadherins [8]. There is evidence that cadherins are involved in invasion and metastasis of tumor cells, in embryonic morphogenesis and in the formation of stable interactions among the cells in solid tissues and in epithelia [9, 10]. In adhesion (Figure 1.1), cadherin dimers on one cell make homophilic contacts with cadherin dimers on another cell [11]. Both dimerization and adhesion are calcium-dependent [12].

Immunoglobulin-like adhesion molecules: The immunoglobulin - like (Ig hereafter) large family of adhesion molecules have components that resemble Immunoglobulins[3]. The group (for a review on Immunoglobulin-like adhesion molecules see [13]) includes molecules involved in antigen recognition by lymphocytes and other cells as well as molecules involved in the development of the nervous system and regulation of the immune system [13]. They function collectively by both homophilic and heterophilic binding. The best-studied members of this group are the Neural Cell Adhesion Molecules (N-CAMs, Figure 1.1), which are expressed predominantly in nervous tissue, and the Intercellular Cell Adhesion Molecules (ICAMs).

Integrins: Integrins constitute a large family of cell adhesion molecules expressed on most cell types in human beings. They are implicated in biological functions such as embryogenesis, the immune response, haemostasis, inflammation, and maintenance of tissue integrity. They bind not only to components of the ECM (Figure 1.1), but also to other cell adhesion molecules [14], and, in some cases, are exploited as cell surface receptors for certain viruses [15]. Integrins are competent to signal response to both adhesive interactions outside the cell (outside-in) [16, 17] and events inside the cell (inside-out) [14, 18, 19]. Outside-in signaling triggers processes within the cell mediated by the cytosolic domains of the integrins, that ultimately affect the cell cycle. Inside-out signaling is manifest through what is believed to be a conformational switch in the extracellular region between active (adhesive) and inactive states (non-adhesive), and regulates ligand binding.

Selectins: The selectin family of adhesion molecules (For a Review see Tedder et al [20]) are expressed primarily in leukocytes and endothelium, and mediate the initial attachment between these two types of cells during the immune response. In contrast to other cell adhesion molecules, selectins bind to carbohydrate ligands on cells (Figure 1.1). The resulting binding forces are relatively weak, which allows leukocytes to roll along the venular wall [21, 22]. In contrast to integrins and immunoglobulin superfamily members that mediate an array of cell-cell interactions throughout the body, selectin function is uniquely restricted to the vascular system.

In summary, adhesion of cells to the ECM via CAMs permits connection of the intracellular cytoesqueleton to the extracellular environment, which is crucial for bidirectional signal transduction (inside-out and outside-in) between extracellular and intracellular compartments. In addition, adhesion of cells to other cells via CAMs is responsible for tissue integrity and is essential in the immunological response. As a whole, ECM and CAMs are involved in a large range of developmental processes and diseases, either through an increase or a decrease in adhesion, as diverse as the differentiation of tissues, morphogenesis and development of metastasis, common colds, Duchenne muscular dystrophy, Human Immunodeficiency Virus disease (HIV hereafter), malaria, leprosy, cancer, graft rejection, asthma, atherosclerosis, some inflammatory diseases and viral infections [23, 24, 25, 26, 27, 28, 29, 30, 31].The development of strategies to correct cell-ECM adhesion and cell-cell adhesion dysfunctions, therefore, has enormous promise as a route for improving treatment of many important clinical conditions.

[3]An immunoglobulin or antibody is a large Y-shaped protein used by the immune system to identify and neutralize foreign objects like bacteria and viruses.

Among the known CAMs, only proteins of the integrin family participate in both cell-cell and cell-ECM interactions and are therefore related to an enormous number of pathologies in living organisms. For this reason, integrins became the main character of this and other theses projects in our group, where an important effort has been done to understand the nature and regulation of the cell-cell and cell-ECM interactions mediated by proteins of this family. Additionally, a successful attempt has been done to design and synthesize small molecular weight inhibitors that mimic the electronic and structural features of CAMs and bind to integrins thus, blocking the pathological processes integrins and their ligands are related to.

Before going into detail with the studies carried out during this thesis project, the reader will find in the following sections a brief introduction to the structure and function of integrins and their counterreceptors, and the type of interactions they form. In later sections, the reading will focus on two specific integrins (and counterreptors), Leukocyte-Function associated Antigen-1 (LFA-1 hereafter) and Very Late Antigen-4 (VLA-4 hereafter), as these are the therapeutic targets on which the group has worked during previous and the current thesis projects. In Chapters 2 through 7, the methodology used, the results obtained and the conclusions from the different studies will be explained.

1.2 Integrins

Like other cell adhesion molecules, integrins (discovered by Takagi and Hynes [32, 33]), differ from other cell-surface receptors (hormones) and other extracellular soluble signal molecules (chemokines) in that they usually bind their ligands with comparatively very low affinity and are present at about tenfold to a hundredfold higher concentration on the cell surface. The reason for the weak binding could be found in the fact that many integrin-dependent biological processes in our body, as diverse as cell migration, the organization of the extracellular matrix and platelet aggregation, need to be transient in nature. If the binding were too tight, cells would presumably become irreversibly glued to the matrix or to other cells and would be unable to move and fulfill their roles. This problem does not arise if attachment depends on large numbers of weak adhesions, also known as the Velcro principle.

Like other transmembrane cell adhesion proteins, however, integrins do more than just attach a cell to its surroundings; they also activate intracellular signaling pathways that communicate to the cell the character of the extracellular matrix or cell to which it is bound, and vice versa. I.e, integrins transduce signals across the cell plasma membrane bidirectionally, which makes them unique [19].

For the sake of brevity, only those details about integrins considered to be of special significance for this thesis project will be provided, whereas for deeper analysis the reader will be referred to excellent reviews on the subject.

1.2.1 Structure of Integrins

Integrins are transmembrane glycoproteins composed of two noncovalently associated subunits, α and β. Nineteen different integrin α subunits and eight different β subunits have been reported in vertebrates [14, 34]. A given β subunit can combine with one or several different α subunits forming, 25 known α-β heterodimers (See Figure 1.2) and perhaps making of integrins the most structurally and functionally diverse family of cell adhesion molecules.

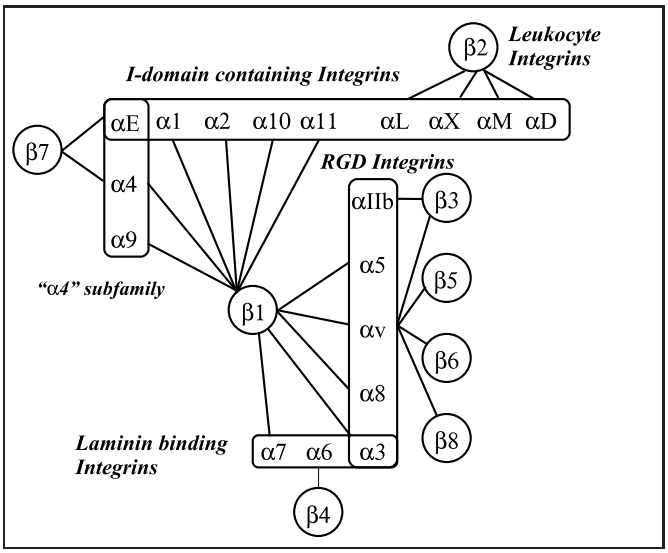


Fig. 1.2: Heterodimeric combinations reported to date between different integrin α and β subunits.

The α subunit: This subunit is about 1100-1200 residues long and contains a short cytoplasmic tail, a short transmembrane domain and a large extracellular part. The N-terminal extracellular region is used by most of the integrins to bind to extracellular ligands. This book will be focused on this domain. The N-terminal region of the integrin α subunit contains seven segments of about 60 amino acids each that have been predicted to fold into a seven-bladed β-propeller domain [35] (Figure 1.3) . Half of the integrin α subunits contain a domain of about 200 amino acids inserted between the second and third β-strands of the β-propeller, known as the Inserted domain, I-domain or von Willebrand factor A domain [36, 37] (Figure 1.3).

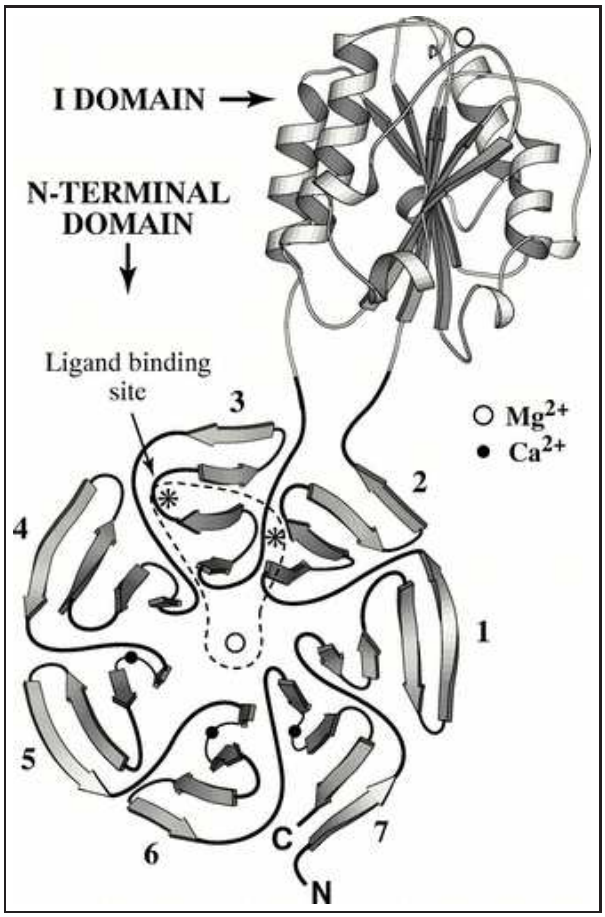


Fig. 1.3: Cartoon representation of the general model for a β propeller domain. Inserted between β-sheets 2 and 3, the I-domain is shown. As an empty circle and a black circle, respectively, the magnesium and calcium binding sites. (Figure taken from reference [38])

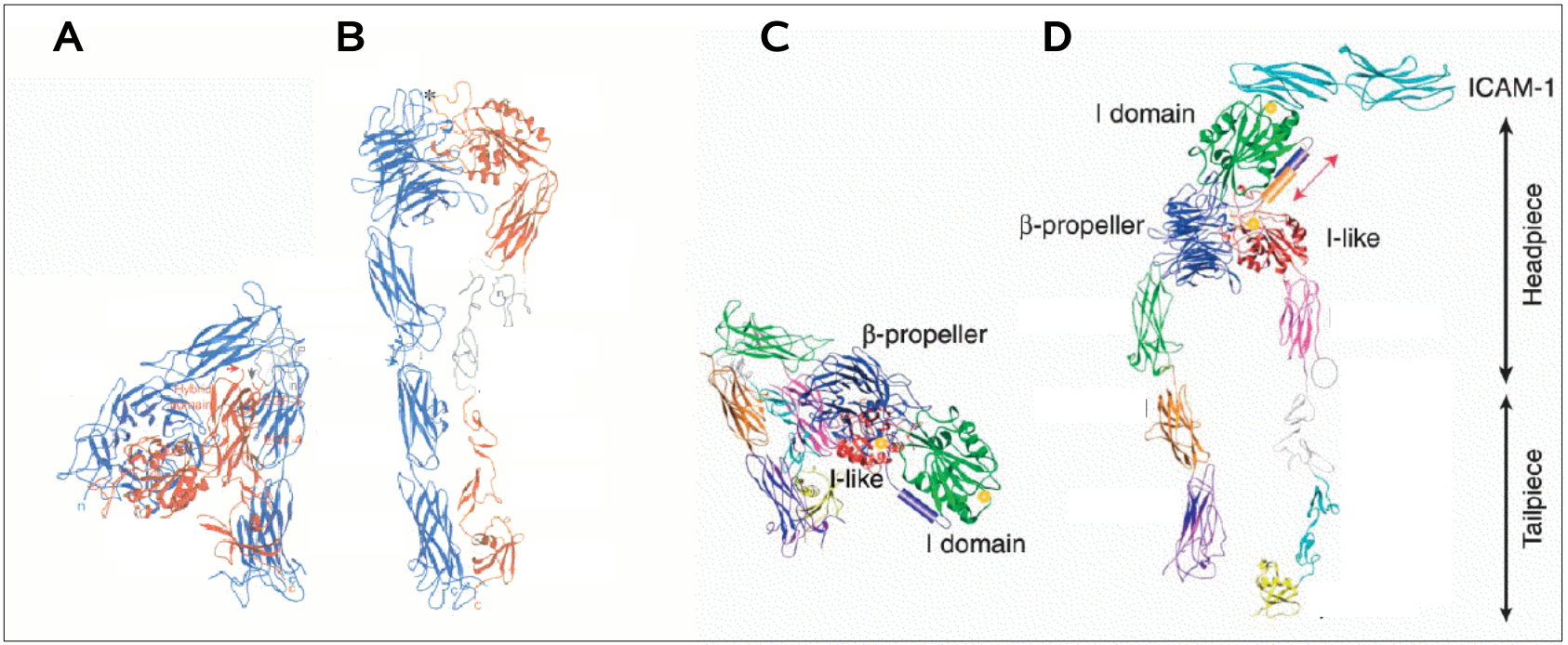


Fig. 1.4: A and B: Ribbon diagram of the three-dimensional structure of the unliganded αv (in blue) - β-3 (in red) integrin, in the inactive conformation (A) and active conformations (B). C and D: Model structure of an I-domain containing integrin (such as LFA-1), in the inactive and ligand-free conformation (C) and the active and ligand-bound conformation (D). (Figure obtained from Hynes et al [19]).

The three-dimensional structures of the αM I-domain [39, 40, 41], αL I-domain [42, 43, 44, 45], α2 I-domain [46, 47] and the α1 I-domain [48, 49], show that the I-domain adopts the dinucleotide-binding or Rossmann fold (Figure 1.3), with a divalent Mg^{2+}/Mn^{2+} cation octahedrically coordinated to a conserved DxSxS sequence at the "top", designated the metal ion-dependent adhesion site (MIDAS hereafter). The metal ion binds negatively charged residues in extracellular ligands [46, 47].

The β-subunit: The N-terminal region of the integrin β subunits contain an evolutionarily conserved domain of about 240 residues that contains a putative metal-binding DxSxS sequence motif similar to that of the MIDAS in the I domain, a similar secondary structure [47], and weak but detectable sequence homology to the I-domain [50]; therefore, it has been termed the I-like domain. This region is a hotspot for point mutations that result in a lack of association of the integrin β2 subunit with α subunits, or loss of function. The I-like domain appears to bind ligand directly in integrins that lack I domains (such as VLA-4), and to indirectly regulate ligand binding by integrins that contain I-domains.

For further reading on the α and β subunits of integrins, see published works of Shimaoka and coworkers [51], Springer and coworkers [35] and Gahmberg and coworkers [52].

1.2.2 Integrin Binding to Ligands. Bidirectional Signaling

As mentioned above, most mature integrins are composed of a large $\alpha\beta$ extracellular domain, a $\alpha\beta$ transmembrane domain, and a short $\alpha\beta$ cytoplasmic domain. Integrin extracellular domains bind to ligands in the extracellular matrix or on the surface of other cell types or viruses. These interactions outside the cell induce within cells changes in cytoskeletal organization, protein phosphorylation and gene expression. This signaling is termed 'outside-in' signaling (For reviews on 'outside-in' signaling see [16, 17]). In addition, integrin intracellular domains interact with molecules inside the cell, a process known as integrin activation. Following activation, two events may occur: integrin clustering on the cell surface and/or a conformational change on the integrin extracellular domains. The former is known as 'avidity' regulation of integrin ligand binding, whereas the latter is known as 'affinity' regulation. Both 'avidity' and 'affinity' regulations are consequences of what is known as 'inside-out' signaling (For reviews on 'inside-out' signaling see [14, 18, 19]).

In this thesis project, there was a particular interest on the affinity modulation events (inside-out signaling), i.e., on the conformational changes on the integrin domains that regulate affinity for extracellular ligands. Alternatively, it will also be discussed the possibility of the extracellular ligands being able to provoke the conformational changes on the integrin extracellular domains, which has also been widely discussed [16, 17].

Current experimental data suggest that integrins are present on the cell surface in a default inactive state, and that intracellular ligands provoke conformational changes on the integrin extracellular domains, that changes affinity for extracellular ligands. This 'inactive' to 'active' conformational changes are very important in many biological processes such as platelet aggregation [53] and cell migration [54], which need to be transient in nature. In fact, usually, following activation, ligand binding is regulated by a combination of conformational changes and receptor clustering [53, 55, 56]. Currently it is not clear how does integrin activation take place, understood as binding of intracellular ligands to the cytoplasmic tails of integrins and

traversing the signal across the integrin transmembrane domains to the ligand-binding region in order to mediate changes in ligand-binding affinity. One proposed mechanism to account for this transmission of information is through long-range conformational changes. While the nature of the structural changes responsible for the switch between the inactive-bent and active-extended conformations is not known, it seems pretty probable that the extracellular I-domain of integrins can adopt a number of distinct conformations [57, 58]. Thus far, it has been possible to solve the three-dimensional structures of isolated I-domains in both activated and resting conformations [59].

Cation dependency: Binding of integrins to their ligands depends on extracellular divalent cations, such as Ca^{2+}, Mn^{2+} or Mg^{2+}, depending on the integrin [44, 45, 60, 61, 62, 63, 64, 65, 66]. There are divalent-cation-binding domains in the extracellular part of both the α and the β subunits. The nature of the divalent cation can influence both the affinity and the specificity of the interaction, but the specific role of different cations is not clear yet, as the experimental results in this aspect of the modulation of integrin affinity for ligand are sometimes contradictory. Except for the $\alpha 4$ subunit, the α subunits that lack an I-domain, contain three putative cation binding sites. Those α subunits that have an I-domain, however, have an extra divalent cation site, the previously mentioned MIDAS motif (Figure 1.3). In chapters 3 and 4, the reader will find the theoretical results and conclusions obtained in this thesis work with respect to the cation dependency of integrin-ligand interactions. The most significant results of these studies have been published in [67, 68].

1.3 Proteins of the Immunoglobulin Superfamily

This family of proteins includes a large number of cell surface proteins with diverse biological functions, such as the mediation of cell surface recognition in immunological processes [69, 70, 71], or serving as host cells viral receptors, with the most notable example being CD4 for HIV [72]. Immunoglobulin Superfamily (IgSF hereafter) molecules contain one or more Ig-related domains that share a common folding pattern, displaying a sandwich structure of two β-sheets that consist of antiparallel β-strands containing 5-10 amino acids, such the example in Figure 1.5, VCAM-1.

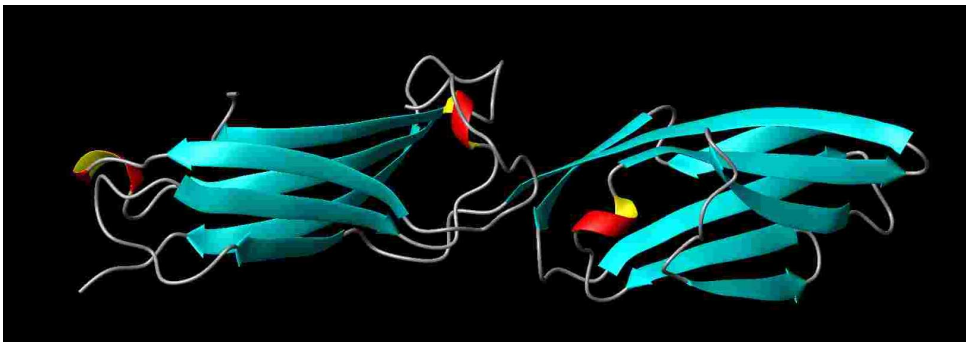


Fig. 1.5: Cartoon representation of the crystal structure of domain 1 and domain 2 of VCAM-1 (PDB ID=1IJ9 [73]) .

IgSF protein-protein interactions can be either homophilic [74, 75, 76] or heterophilic [77, 78]. Though not all Immunoglobulin-Superfamily adhesion molecules mediate cell-cell interactions, many which belong to this family do function as adhesion receptors, including intercellular adhesion molecule-1 (ICAM-1; CD54), intercellular adhesion molecule-2 (ICAM-2), vascular cell adhesion molecule-1 (VCAM-1; CD106), platelet-endothelial cell adhesion molecule-1 (PECAM-1; CD31) and the mucosal addressin cell adhesion molecule-1 (MAdCAM-1). In heterophilic interactions, IgSF proteins constitute the main group of ligands for integrins, and mediate cell-cell adhesion.

The ligand binding repeat is localized to different domains, depending on the IgSF member (Domain 1 in ICAM-1 and VCAM-1, Domain 2 in ICAM-2...) They are constitutively expressed in the biologically active conformation in many cell types, such as leukocytes, endothelium and muscle cells. Some of them, such as ICAM-1, can be overexpressed by cytokines, but some others, such as ICAM-2 can not. Among the IgSF members, ICAM-1 and VCAM-1 are involved in the adhesion of T cells to endothelial cells by serving as surface ligands for the integrins LFA-1 (Leukocyte-Function associated Antigen-1; $\alpha L\beta 2$) and VLA-4 (Very Late Antigen 4; $\alpha 4\beta 1$), respectively.

1.4 VLA-4 and LFA-1 as Therapeutical Targets

The immune system, normally efficient in defeating external threats from the microbial world, at times directs its potent arsenal against the body's self-constituents, causing autoimmunity and chronic immflamatory diseases. The latter, often involve distinct anatomic regions, such as junctions, and involve homing of leukocytes to specific tissues. Molecules key in homing leukocytes to sites of inflammation (integrins and selectins) display a high degree of tissue diversity, thus, the particular integrin or selectin expressed is critical for entry to a particular anatomic site. As a consequence, it is conceivable that blocking a particular ligand/receptor interaction would be sufficient to abolish pathologic homing, yet leave lymphocytes free to move elsewhere. This has indeed turned out to be a very promising avenue, at least in preclinical models of autoimmunity [79]. Integrin VLA-4 [79, 80, 81], as well as integrin LFA-1 [82, 83, 84, 85, 86, 87], have become targets of the treatments of autoimmune and inflammatory diseases [88].

In addition, integrins VLA-4 and LFA-1 play an important role in virtually every stage of several cancer progression and, more importantly, are required for survival by cancerous cells, but not by normal cells. Interaction of VCAM-1 with VLA-4 integrin expressed on certain tumor types has been suggested to be an important mechanism for the development of metastasis in melanoma [89, 90, 91], renal cancer [92], bone cancer [93], stomach cancer [94] and lymphoma [95]. Interaction of LFA-1 with ICAM-1 is involved in the metastasis of gastrointestinal carcinoma [96, 97], melanoma [98, 99, 100], lymphoma [101, 102, 103, 104] and colon cancer [105].

Previous works in the Organic Chemistry group ([106, 107, 108, 109]), were aimed at designing and synthesizing inhibitors against the VLA-4/VCAM-1 interaction. This was done by mimicking the geometric and electronic features of the binding motif in VCAM-1, so that the molecules could compete with VCAM-1 binding to the β propeller (Figure 1.3) in the $\alpha 4$ subunit of VLA-4. A series of molecules were designed, synthesized and tested in biological assays and showed to have and excellent inhibitory capacity. They were also subjected to *in vivo* tests in murine

cancer models, and were able to substantially reduce the number and size of metastases. These molecules are derivatives of pyrrolidine scaffolds, and some of them are shown in Figure 1.6.

Due to the existing similarity between VCAM-1 and ICAM-1 integrin binding domains (see later), the following studies were performed: i) try to predict how would these molecules interact with the I-domain of LFA-1 and ii) study the way these molecules should be modified in order to get a more specific and efficient family of inhibitors against LFA-1. Fortunately, not only has the isolated αL I-domain been crystallized, but an 'active' mutant of this I-domain has also been cocrystallized with ICAM-1, which provides some information about hotspots in the binding of these two proteins. Based on the electronic and geometric features of ICAM-1, a new familiy of LFA-1 antagonists was designed by former member of the group (see chapter 7), closely related to the family of inhibitors against integrin VLA-4.

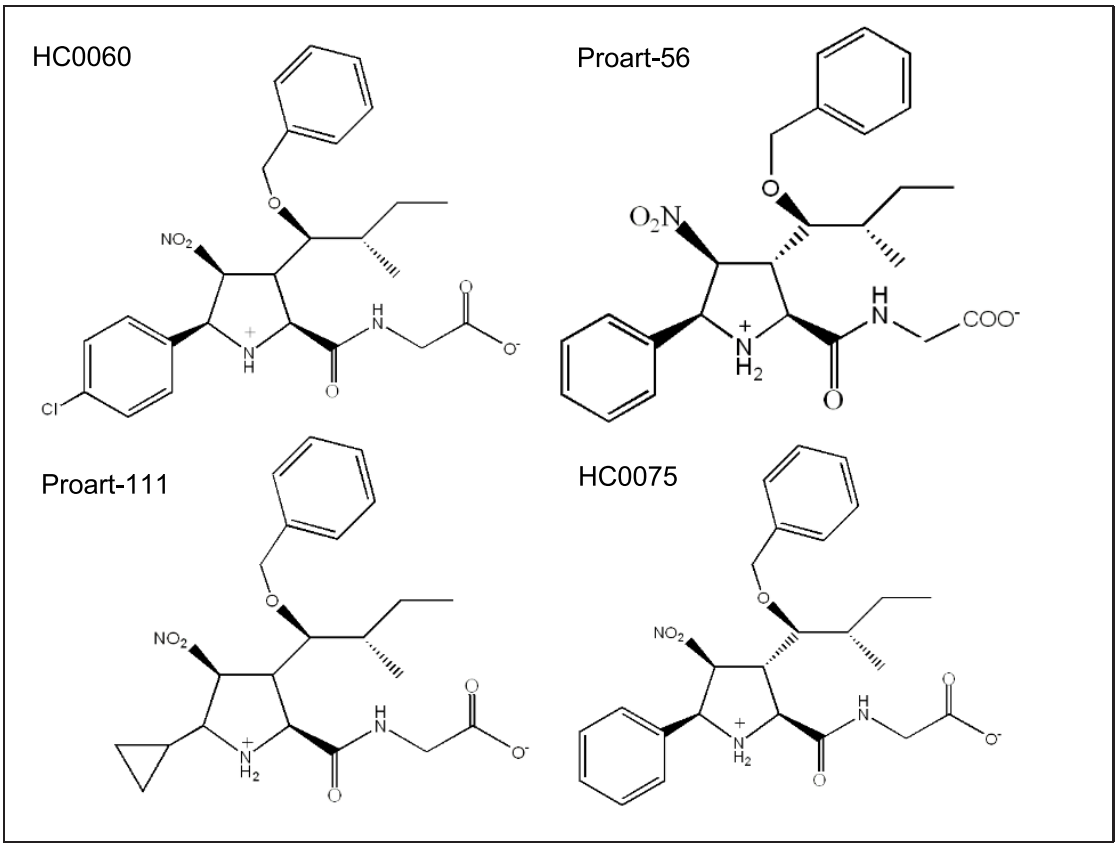


Fig. 1.6: Some of the compounds designed, synthesized and tested against integrin VLA-4 by other members of the group.

1.5 Aims of This Thesis Project

The goal of this thesis project was to solve some unanswered questions with respect to the nature of the cell adhesion processes in which integrin LFA-1 and its extracellular ligand ICAM-1 are involved. In addition, small molecular weight LFA-1 antagonists have been designed, synthesized and tested *in vitro* and *in vivo*. The studies performed during this thesis work are summarized in the paragraphs below, and will be presented in Chapters 3 to 6.

Chapter 3.
Quantum Mechanics: A Density Functional Theory study on the MIDAS motif

Since integrins possess several ligand-affinity modulating mechanisms, it has been proved difficult under experimental conditions to determine the relative contributions of different individual phenomena on the affinity of the MIDAS motif for ligands. Density Functional Theory (DFT hereafter) has been applied in this thesis work to estimate to which extent do different factors regulate affinity of the MIDAS motif toward a glutamate-type ligand that mimics GLU34 in ICAM-1. Among different possible factors that influence affinity the following have been considered: i) the changes in the coordination of the metal in the MIDAS motif, ii) the nature of the metal ion present in the MIDAS motif (Mg^{2+}, Ca^{2+}, Zn^{2+} and Mn^{2+}), iii) the nature of the incoming ligand and iv) the dielectric permittivity of the environment.

Chapter 4.
Density Functional Theory vs Diffusion Monte Carlo:
Combined study on Mn^{2+}

As the reader will find out later in this thesis book, results summarized in Chapter 3 suggest that Mn^{2+} is the cation that provokes a stronger interaction between the MIDAS motif and extracellular ligands. Due to the transition metal character of this ion, it was considered necessary to refine results in Chapter 3 using a more accurate methodology, such as the Diffusion Monte Carlo method (DMC hereafter), which is better than DFT at estimating relative electronic energies of different spin states of transition metal complexes such as the one under study. The goals of this work are threefold: i) to check using different high level theoretical methods whether the experimental assumption that Mn-protein complexes are high spin species is correct. ii) To analyze the behavior of B3LYP method in comparison to the more sophisticated DMC. iii) To check to which extent do the nature of the coordinated ligands unbalance the relative stability of high-spin complexes.

Chapter 5.
Locating Hotspots: LFA-1/ICAM-1 and LFA-1/Inhibitor

X-ray structures have been solved for three different conformers of the αL I-domain, the so-called low, medium and high affinity structures. Neverthess, only the medium affinity I-domain has been cocrystallized in complex with its natural ligand ICAM-1. In this thesis project a "low" and "high" affinity complexes were also generated with computational tools, and key interactions in each of the three complexes were located by means of a combination of Molecular Dynamics Simulations and a Free Energy Decomposition protocol. The nature and magnitude of the most important interactions are analyzed, which represent a key piece of information that is used in this study for the design of a new family of LFA-1 antagonists.

Chapter 6.
Nuclear Magnetic Resonance experiments to locate the binding pocket

In this experimental part of the present thesis work, the binding site of LFA-1 antagonists were located on the isolated αL I-domain and binding strength estimated by means of Nuclear Magnetic Resonance (NMR hereafter) techniques. In addition, the experiments revealed unknown details on the structure of the active mutant αL I-domain as reported in the literature. A summary of the *in vivo* and *in vitro* cell adhesion assays performed with in-house compounds is also provided in this chapter.

In this project, Theoretical Chemistry computer tools have been used, such as software based on Quantum Mechanics (DFT and DMC), classical mechanics, Molecular Dynamic Simulations and Free Energy Decomposition protocols, and Nuclear Magnetic Resonance techniques. The basis of all these methodologies are summarized in Chapter 2. In addition, Chapter 7 collects all the conclusions obtained in each of the studies performed and summarized in Chapters 3 to 6.

Chapter 2

Methods

Contents


OH ALICE.. YOU'RE THE ONE FOR ME
BUT BOB.. IN A QUANTUM WORLD HOW CAN WE BE SURE?
Physics World

2.1 Quantum Theory

Computational chemistry and molecular modeling are used to characterize and to predict the behavior of small and macro-molecules by evaluating their energy. The most precise method to calculate the energy of molecular systems is by solving the equation of Schrödinger, the basis of the Quantum Theory, as will be explained below. This methodology, treats the electrons in a molecule explicitly, which is computationally expensive; the latter provokes that the method is unfortunately currently limited to systems with a small number of atoms. For large biological molecules like proteins, the description of energetics of the system relies on simpler methods, namely molecular mechanics. These methods are based on the principles of classical and not quantum mechanics, treating the positions of atoms and not of electrons. Quantum theory will be the central part of the present section, whereas molecular mechanics will be treated in the next section.

Quantum Theory describes the arrangement of nuclei (protons and neutrons) and electrons with a mathematical function called *wavefunction*, Ψ. To obtain the wavefunction, the so-called Schrödinger equation needs to be solved (Equation 2.1) for that system,

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where $\hat{H}$ denotes the Hamiltonian operator and E the energy of the system. For a system with N_e electrons and N_N nuclei, the Hamiltonian in atomic units would be expressed as in Equation 2.2, aside from relativistic effects [1] and spin-orbit coupling [2].

$$\begin{aligned}
\hat{H} &= \hat{T}_e + \hat{T}_N + \hat{V}_{e-e} + \hat{V}_{e-N} + \hat{V}_{N-N} \\
&= -\sum_{i}^{N_e} \frac{\nabla_i^2}{2} - \sum_{A}^{N_N} \frac{1}{2M_A} \frac{\nabla_A^2}{2} \\
&\quad -\sum_{A}^{N_N} \sum_{i}^{N_e} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N_e} \frac{1}{r_{ij}} + \sum_{A>B}^{N_N} \frac{Z_A Z_B}{R_{AB}}
\end{aligned} \tag{2.2}$$

In Equation 2.2 $\hat{T}_e$ and $\hat{T}_N$ represent the kinetic energy of each electron and each nucleus, respectively, and $\hat{V}_{e-e}$, $\hat{V}_{e-N}$ and $\hat{V}_{N-N}$ represent the electron-electron, electron-nucleus and nucleus-nucleus interaction terms, respectively. Also in Equation 2.2, N_e and N_N represent the number of electrons and the number of nuclei in the system, respectively, M_A is the mass of nucleus A, Z_A and Z_B are the atomic number of nucleus A and B, respectively, R_{AB}, r_{ij} and r_{iA} are the distance between nuclei A and B, and electrons i and j, and the electron i and the nucleus A, respectively.

[1] When a mass moves at a speed approaching the speed of light, it's inherent properties are changed in surprising ways; these are known as the Relativistic Effects which for the quantum description of the heaviest atoms in the periodic table, may become crucial

[2] Magnetic interaction between the spin of the electron and the angular momentum of the orbital

Since nuclei are so heavy with respect to electrons, their movement will be much slower and hence, the nuclear kinetic term ($\hat{T}_N$), can therefore be neglected. As a consequence, the $\hat{V}_{N-N}$ nucleus-nucleus potential energy term can be regarded as constant, since this term only depends on internuclear distances, which are fixed in the approximation. This is referred to as the Born-Oppenheimer approximation, and yields the approximate Hamiltonian shown in Equation 2.3.

$$\begin{aligned} \hat{H} &= \hat{T}_e + \hat{V}_{e-e} + \hat{V}_{e-N} + \hat{V}_{N-N} \\ &= -\sum_{i}^{N_e} \frac{\nabla_i^2}{2} - \sum_{A}^{N_N} \sum_{i}^{N_e} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N_e} \frac{1}{r_{ij}} + \sum_{A>B}^{N_N} \frac{Z_A Z_B}{R_{AB}} \end{aligned} \tag{2.3}$$

Many mathematical methods have been devised to provide a way of finding the electronic wavefunction for any given atomic arrangement: Hartree-Fock (HF hereafter), nth order Møller-Plesset (MPn hereafter), Coupled-Cluster methods (CC hereafter), Configuration Interaction methods (CI hereafter), Complete Active Space Self Consistent Field (CASSCF hereafter), Multi-Referential Configuration Interaction (MRCI hereafter)... These methods are called *ab initio* methods, because they calculate the wavefunction from first principles, according to the raw Quantum Theory. A more modern approach, and highly promising, is the Density Functional Theory. Although not strictly *ab initio*, it is nonetheless very closely related to the HF method above (simplest of *ab initio* methods), but modified in such a way that its accuracy is substantially improved. Since DFT was the method chosen in the quantum studies carried out in this thesis project, the complete following section is devoted to the explanation of its foundations.

2.1.1 DFT Methods

The Density Functional Theory is a formalism that tries to be an alternative to the wavefunction one. As mentioned above, the Quantum Theory proposes that all the properties of an N-electron system can be calculated from a mathematical entity called *wavefunction*, which is obtained solving the famous Schrödinger equation for that given system. The wavefunction formalism, thus, tries to find (approximate) solutions to that equation. The problem here is that the sought wavefunction Ψ_e is $3N_e$-dimensional, being N_e the number of electrons in the system, and the mathematical complexity increases exponentially with the number of dimensions.

It is known that such "wavefunction" could well be the total electronic density, $\rho(\mathrm{r})$. In 1964 the Hohenberg-Kohn theorem was published [110], which states that there is a one-to-one mapping between the ground state of a system (the electronic wavefunction Ψ_0 which gives the lowest E_0 electronic energy among all the possible Ψ_i that are a solution to $\hat{H}_e\Psi_i = E_i\Psi_i$), and its total electronic density $\rho(\mathrm{r})$. Such a mapping might in principle allow us to calculate the ground state energy, provided that we have the ground-state electronic density as well as the mathematical functional $\mathrm{E}[\rho(\mathrm{r})]$ that gives the electronic energy from the density function.

In a second theorem it was postulated that if $\mathrm{E}[\rho(\mathbf{r})]$ is known, then a variational principle will hold for any trial function $\tilde{\rho}(\mathrm{r})$, as long as $\int \tilde{\rho}(\mathbf{r})d\mathbf{r} = N$. The variational principle states that $E_0 \leq E[\tilde{\rho}(\mathbf{r})]$, so that if we find the $\tilde{\rho}(\mathrm{r})$ trial density that minimizes the energy functional, then $\tilde{\rho}(\mathrm{r}) = \rho_0(\mathrm{r})$. Since the density can be now obtained, the functional is only left to be known.

An indirect method to get an approach to this functional is known as Kohn-Sham method, that takes advantage of the fact that the exact ground-state electronic energy E_0 of an N-electron system is given by Equation 2.4

$$\begin{aligned} E_0 &= -\frac{1}{2}\sum_{i=1}^{N}\left\langle \psi_i(1) \mid \nabla_1^2 \mid \psi_i(1)\right\rangle + \int v(r)\rho(1)d\mathbf{r}_1 \\ &\quad +\frac{1}{2}\int\int \frac{\rho(1)\rho(2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho(\mathbf{r})] \end{aligned} \tag{2.4}$$

In Equation 2.4, $v(r) = \sum_{\alpha}\frac{Z_\alpha}{r_{1\alpha}}$ is the *external potential* exerted by the nuclei, ψ_i are the *Kohn-Sham orbitals*, and $E_{xc}[\rho]$ is the *exchange-correlation energy*. The ground-state density $\rho(\mathrm{r})$ can be constructed from the Kohn-Sham orbitals according to Equation 2.5. The Kohn-Sham orbitals, in turn, are obtained from the set of one-electron equations 2.6.

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} \mid \psi_i \mid^2 \tag{2.5}$$

$$\hat{F}_{KS}(1)\psi_i(1) = \epsilon_i\psi_i(1) \tag{2.6}$$

The $\hat{F}_{KS}$ in Equation 2.6 is the Kohn-Sham operator, and in turn depends on a term called the *exchange-correlation potential*, V_{xc}. This V_{xc} is the functional derivative of the exchange-correlation energy with respect to the electronic density, namely $V_{xc} = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}$.

The procedure to find out ρ(r) is rather similar to the *self-consistent field* (SCF hereafter) employed in the Hartree-Fock method. Here one proposes a trial density $\tilde{\rho}_0$(r), and uses it to construct a $\hat{F}_{KS}$, which has two terms that depend on the density. The first one is a term called *Coulomb operator*, and the second one is the aforementioned V_{xc}. The latter is obtained from functional derivation of E_{xc}, which in turn is defined to depend on ρ(r) as given by Equation 2.7. We can then obtain a set of Kohn-Sham orbitals from Equation 2.6, and consequently a new density from Equation 2.5. This new density can be fed into a new $\hat{F}_{KS}$, and rinse and repeat until self-consistency is found, i.e., the input density is equal to the output one. In Equation 2.7, the ϵ_{xc} function corresponds to the exchange-correlation energy per electron, and is an unknown functional of ρ.

$$E_{xc}[\rho] = \int \rho(\mathbf{r}) \epsilon_{xc}[\rho(\mathbf{r})] d\mathbf{r} \tag{2.7}$$

One approach can be made assuming that, if the electronic density fluctuates slowly from one point of the space to another one, the local exchange-correlation energy per electron (its value at a given point) is given by the same equation that would rule a constant density (homogeneous) electron gas. This approximation is named Local Density Approximation (LDA hereafter), and an accurate ϵ_{xc} is given by an expression by Vosko, Wilk and Nussair [111]. When the LDA is applied taking into account different functions for the α and β spin electrons, then it is called Local Spin Density Approximation (LSDA hereafter).

Since the true molecular electronic densities are far from smoothly-varying, as the LDA assumes, many more sophisticated formulations of ϵ_{xc} have been proposed and tested. One route to increasing the accuracy of the functional is to add to it terms including not only the value of the density in a given point of space, but also the value of its gradient. Since including a gradient gives a touch of non-locality, these modifications are named non-local gradient corrections.

A parallel route would be to add a part of the HF exchange (E_x^{HF}), the formula of which is exactly known. This gives rise to hybrid methods. These hybrid methods give astonishingly good results, provided that they have been fine-tuned by hand previously. This, however, is a serious drawback, since it breaks the ab initio compliance, and makes these methods semiempirical, strictly speaking.

Nevertheless, the most used DFT functional by far, and main responsible of the success of DFT, is the one using Axel D. Becke's corrections [112] for the exchange, and Chengteh Lee, Weitao Yang and Robert Ghormley Parr's correlation formula [113]. The expression for this functional is shown in Equation 2.8, and, since it depends on three parameters, it is known as Becke's 3-parameter functional, with Lee, Yang and Parr's corrections for correlation (B3LYP hereafter).

$$E_{xc} = (1 - a_0)E_x^{LSDA} + a_0 E_x^{HF} + a_x E_x^{B88} + a_c E_c^{LYP} + (1 - a_0)E_c^{VWN} \tag{2.8}$$

The DFT methods are close to HF when it comes to computational cost, but since they include some sort of corrections for taking into account the exchange and correlation, the accuracy of their results is comparable to MP2, and sometimes better. As the reader will find out later on, most of the results in this thesis work have been obtained by DFT calculations.

All-electron basis sets and Pseudopotentials

How exactly does one represent the wavefunctions and densities mathematically? It turns out that only the simplest of atoms, the hydrogen-like ones (H, He^+, Li^{2+}, etc.), can have their Schrödinger equation analytically solved, because they have a single electron under the influence of a central external potential (the nuclear charge). The resulting hydrogen-like wavefunctions all consist on a spherical harmonic angular part and a radial part of the Slater type. Slater type functions are of the kind $f_{Slater} = Ar^l e^{-\alpha(\mathbf{r}-\mathbf{R})}$, where r is the position of the electron, R the position of the nucleus, α a constant that depends on the nuclear charge, l an integer number accounting for the angular momentum, and A a normalization constant, which makes $\int f^2 d\mathbf{r} = 1$. Each such Slater type function that describes the state of the electron, is called an orbital.

Assuming that in multi-electronic systems, each electron will dwell an orbital that will resemble one of those hydrogen-like wavefunctions, discrete sets of different hydrogen-like wavefunctions could be used as basis sets, that is, elements whose proper combination gives rise to a good approximation of the actual total wavefunction. Such basis sets are called Slater-Type Orbitals (STO hereafter).

In practice, the mathematical operations required to find the coefficients of this "proper" combination are much more efficiently carried out if instead of STOs one makes use of Gaussian-Type Orbitals (GTOs hereafter). These functions are similar to Slater-type functions, except their exponents are squared: $f_{\mathsf{Gauss}} = Ar^l e^{-\alpha(\mathbf{r}-\mathbf{R})^2}$.

Since a single GTO is a much worse fit to a hydrogen-like wavefunction (and hence, also to a multi-electron wavefunction orbital), modern basis sets consist of functions that are formed, each of them, by Gaussian-type function combinations (typically 3-6 of them) that resemble single STOs (See Equation 2.9).

$$f_{\mathsf{GTO}} = \sum_i^N c_i f_{\mathsf{Gauss}}(\alpha_i) \tag{2.9}$$

All-electron basis sets are defined as basis sets in which all the electrons of our system have their orbital described by one or more GTO of the set. As an example, the 3-21G all-electron basis set for carbon is shown , as obtained from the EMSL Basis Set Library (http://www.emsl.pnl.gov/forms/basisform.html).

$$
\begin{aligned}
f_1(S) \rightarrow \quad & c_i = (0.0617669, 0.358794, 0.700713) \\
& \alpha_i = (172.256, 25.9109, 5.53335) \\
f_2(S) \rightarrow \quad & c_i = (-0.395897, 1.21584) \\
& \alpha_i = (3.66498, 0.770545) \\
f_3(S) \rightarrow \quad & c_1 = 1.0 \\
& \alpha_1 = 0.195857 \\
f_1(P) \rightarrow \quad & c_i = (0.23646, 0.860619) \\
& \alpha_i = (3.66498, 0.770545) \\
f_2(P) \rightarrow \quad & c_1 = 1.0 \\
& \alpha_1 = 0.195857
\end{aligned}
$$

Since each P function consists actually of three functions (P_x, P_y and P_z), the 3-21G basis set (as the name implies in a cryptic nomenclature), devotes 3 Gaussian-type functions for the orbitals in the core of the atom (just the 1s orbital in the case of carbon), that is $f_1(S)$, which is built by the linear combination of three Gaussians.

The valence orbitals are described by two functions each: $f_2(S)$ and $f_3(S)$ for the 2s, a $f_1(P)$ and a $f_2(P)$ for each of the 2p. The accuracy loss caused by the construction of the GTOs with less than three Gaussians is more than compensated by the fact that each orbital in the valence can be described by the arbitrary linear combination of two GTOs.

In the case of pseudopotentials, the core electrons are replaced by model potentials, and the valence electrons are treated by regular GTOs. They are therefore also called Effective Core Potentials (ECPs hereafter). This way, the chemically active valence electrons are considered to move in the potential caused by the chemically inactive core electrons. The fact that a whole lot of electrons is not considered into the wavefunction coefficient optimization makes that much less GTOs are required into the basis set, and thus the calculations are importantly sped up.

During this thesis project, for the geometry optimizations and subsequent frequency calculations the selection was the following (unless otherwise indicated): shared-exponent basis set of Stevens, Krauss, Basch and Jasien (SKBJ hereafter) effective core potential basis set [114], as defined in the Extensible Computational Chemistry Environment Basis Set Database. Such basis set was selected for performance considerations, since it has all of the core electrons described by pseudopotential functions, and only the valence shell of each element is described with Gaussian basis functions, thus reducing the computational expense considerably. For the single point energy calculations widely known 6-311++G(2df,2p) basis set was used, as implemented in Gaussian. This basis set will be referred to as TZ basis set throughout this thesis.

Search of Stationary Structures

As explained in the paragraphs above, the Born-Oppenheimer approximation leads to the concept of a potential energy surface (PES hereafter). The PES is the function that describes how the potential energy changes as the nuclei move relative to one another. Once the six degrees of freedom for overall rotation and translation of the molecules are discarded, and for a nonlinear molecule, 3N-6 *internal* degrees of freedom are left. Thus, the PES is a 3N-6 dimensional hypersurface, where the minima and saddle points have to be located.

Minima and Saddle Points

A complete mapping of the full PES at a quantum level of theory is tractable only for small molecules, but a great deal about a molecule and its behavior can be learned from knowledge of a few selected points of the PES. Local minima on the PES are fundamental points since they represent stable isomers of a molecule. They determine the molecular structure and the moments of inertia by which the rotational spectra can be estimated. The curvature of the many-dimensional PES about the local minima, the steepness of the walls, determines the vibrational properties of the molecule including the normal modes and harmonic vibrational frequencies.

Another fundamental point on the PES is the lowest-energy *saddle point* connecting two local valleys or minima. The saddle point is the highest point on a lowest total energy pathway connecting these minima. It has the property that it is a maximum with respect to one degree of freedom and is a minimum with respect to all other degrees of freedom. The significance of the saddle point is that it is an approximate dynamic bottleneck, a point of no return, for a transformation from the vicinity of one local minimum or valley to another. The probability (or rate) of reaction is very strongly influenced by the height in energy of this saddle point.

Both the minima and the saddle points are *stationary points* on the PES; the derivative of the energy with respect to each of the geometry coordinates, and thus the force on each of the atoms is zero as shown in Equation 2.10,

$$\frac{dE(q_1, q_2, ..., q_m)}{dq_i} = 0 \tag{2.10}$$

Location and characterization stationary points

There are various methods to locate stationary points on a PES, also called *optimization methods*. The interested reader will find good descriptions in the bibliography [115, 116]. Once the stationary point is found, its nature should be tested by calculating vibrational frequencies [117]. The stationary point is a minimum if all frequencies are real. If there is one imaginary frequency, the stationary point is a saddle point and the transition vector will give the sense of the distortion toward the reactants in one direction an towards products in the other. It must be remembered that the number of imaginary frequencies (and therefore the nature of the stationary point) is valid only at the level of theory used to calculate the frequencies. Calculations at higher levels of theory may present a different picture. For example, if a calculation on a saddle point with a larger basis set and including electron correlation gives a substantially different energy barrier than the lower level of theory, then the surface may be qualitatively different at higher level. When this situation occurs, it is best to reoptimize all stationary points at a higher level of theory.

Thermodynamic Quantities

Apart from characterizing the nature of the stationary points, vibrational analysis can also be used for the evaluation of several thermodynamic quantities. The breakdown of the key thermodynamics relations and energy components is summarized below:

$$G = H - T \cdot S \tag{2.11}$$
$$H = U + R \cdot T \tag{2.12}$$
$$U = E_0 + E_{vib} + E_{rot} + E_{trans} \tag{2.13}$$
$$E_0 = (E_{elec} + E_{NN}) + E_{ZPV} = E + E_{ZPV} \tag{2.14}$$

where G, U, H, S and T are the Gibbs free energy, internal energy, enthalpy, entropy and temperature, respectively, R is the universal gas constant, and E_{elec}, E_{NN}, E_{ZPV}, E_{vib}, E_{rot} and E_{trans} are the electronic energy, nuclear-nuclear repulsion energy, zero-point vibrational energy, thermal vibrational energy correction, rotational and translational energy components, respectively. The expression for the enthalpy (Equation 2.14) assumes the ideal gas law for a mole of particles. The internal energy and entropy are derived from standard statistical mechanical expressions for separable vibrational, rotational and translational contributions within the harmonic oscillator, rigid rotor, ideal gas/particle-in-a-box models in the canonical ensemble [118]. The standard state is for a mole of particles at $T = 298$ K and 1 atm pressure ($V = R \cdot T/P$). All quantities above except E_0, E_{elec}, E_{NN} and E_{ZPV} have explicit temperature dependence.

2.1.2 Quantum Montecarlo Methods

The correct prediction of the relative energies between different spin states in transition metal complexes require the use of accurate ab initio methods, and it can be a delicate task. In fact, attempts to predict the high spin state of such a metal, Mn^{2+}, have failed to do so [119]. This is not surprising, since a variety of effects have to be taken into account accurately to yield reliable predictions of relative energies among different electronic states in transition-metal complexes. In the context of Density Functional Theory reliable exchange and correlation potentials are needed. In addition, these systems can show important near-degeneracies, also known as non-dynamical electron correlation, which can be problematic in the context of Density Functional Theory. In fact, pure DFT functionals have been seen to yield incorrect ordering of spin states [120]. On the other hand the hybrid B3LYP functional, that includes a fraction of the exact exchange, describes better systems with large exchange [121]. In general, although DFT is widely applied in organometallic systems, one should be very careful with those metals, like Mn^{2+}, in which different spin states and near-degenerate orbitals are present. Quantum Monte Carlo methods, where the name comes from the fact that Monte Carlo technique is used for the evaluation of the integrals, and the functions involved are quantum wave functions, are known to recover 95% of the electron correlation in chemical systems. Moreover, it seems that these methods are able to describe properly non-dynamical electron correlation, leading to a more balanced description among different spin states. In particular, the Diffusion Monte Carlo method (DMC) is becoming widely used in last years, and it can be the method of use for this type of difficult cases for DFT. The wavefunctions used in the DMC method and the method itself are described below.

Trial Wave Functions

The use of accurate trial wavefunction is important, although not crucial, for DMC. However, since the calculation of the trial wave function are time consuming, it appears necessary the use of wavefunctions that are both accurate and easy to evaluate. The wavefunctions used in DMC are Slater-Jastrow type wavefunctions, as expressed in Equation 2.15:

$$\Psi(X) = D(X)e^{J(X)} \tag{2.15}$$

where $D(X)$ is a Slater-type determinant and $J(X)$ is the Jastrow correlation factor. $X = (x_1, x_2, \ldots, x_N)$ contains the spatial and spin coordinates of all electrons. $x_i = (r_i, \sigma_i)$. Usually, monodeterminantal wavefunctions are used, and the orbitals of this determinant are mostly calculated using HF or DFT methods, described earlier in this Chapter. HF orbitals give the lowest energy monodeterminantal wavefunction, and combined with a good jastrow factor usually gives accurate results.

The trial wavefunctions used in this work have been built by using HF [122, 123] method for the determinantal part and Jastrow factors recently developed [124]. All the DMC calculations were carried out by Dr. Ion Mattin Matxain from the Theoretical Chemistry Group at the Chemistry Faculty of the University of the Basque Country (San Sebastian).

For a deeper reading the reader is referred to more specialized papers [125, 126].

Diffusion Monte Carlo

The DMC method [125, 127, 128] is a stochastic projector method for solving the imaginary-time many-body Schrödinger equation. It is based on the similarity between the Schrödinger equation in imaginary time $\tau = it$ 2.16 and the generalized diffusion equation 2.17

$$\frac{\partial \Phi(r,\tau)}{\partial \tau} = \frac{1}{2m}\nabla^2 \Phi(r,\tau) - V(r)\,\Phi(r,\tau) \tag{2.16}$$

$$\frac{\partial f(r,t)}{\partial t} = D\nabla^2 f(r,\tau) - k(r)\,f(r,t) \tag{2.17}$$

D is the diffusion constant in Fick's second law, and $k(r)$ is the position-dependent rate constant of a first-order rate equation. Fermi suggested that a random walk in which a particle diffuses and simultaneously multiplies based on the rate constant would eventually give the ground-state wave function. Starting from the formal solution of equation 2.16 shown in equation 2.18, considering equation 2.19 and spanding the initial wave function in eigenfunctions of H as shown in equation 2.20 the time-dependent solution in terms of the eigenfunctions is obtained (equation 2.21).

$$\Phi(r,\tau) = e^{-\tau H}\Phi(r,0) \tag{2.18}$$

$$H = -\frac{1}{2m}\nabla^2 + V(r) \tag{2.19}$$

$$\Phi(r,0) = \sum_i a_i \psi_i \tag{2.20}$$

$$\Phi(r,\tau) = \sum_i a_i e^{-E_i \tau}\psi_i \tag{2.21}$$

The contributions from excited states decay exponentially compared with the ground state. When a random walk that satisfies the diffusion equation is built, after sufficient time, the ground state wave function is obtained exactly. Exact imaginary-time evolution would lead to the exact ground state wave function, provided it has a non-zero overlap with the initial state. This is a fundamental property of the projector $e^{-\tau H}$, which is the basis of Diffusion Monte Carlo methods.

Due to the antisymmetry of the wave function, it has positive and negative regions. Unfortunately, DMC can only handle positive values. There are several ways to overcome the so called sign problem, being the **fixed-node approximation** [127, 129, 130, 131] the most common one. This method is not exact, but provides a variational upper bound on the ground state energy, and usually is very accurate. The fixed-node variational principle was proved by Moskowitz and coworkers [131] and Reynolds and coworkers [130], and the reader is referred to those works for the demonstration.

Pseudopotentials

Although the computational effort of a DMC calculations scales as the cube of the number of electrons, the scaling with the atomic number, Z, of the atoms is approximately $Z^{5.5-6.5}$ [132]. Many properties such as the interatomic bonding and low-energy excitations are determined by the behavior of the valence electrons. It is therefore very advantageous to use pseudopotentials in DMC calculations, which reduces the effective value of Z. Errors are introduced, but the gain in computational efficiency is huge and makes applications to heavy atoms possible. The idea of pseudopotentials is to create an effective potential (pseudopotential) in order to reproduce the effect of both the nucleus and the core electrons on the valence electrons. For each angular momentum states this is done separately, and hence the pseudopotential contains angular momentum projectors, which are nonlocal operators, and the pseudopotentials are called nonlocal pseudopotentials. Conventionally the pseudopotential $V_l^{ps}(r)$ is divided into two different parts. A local part, $V_{loc}^{ps}(r)$, common to all angular momentum, and a nonlocal part, $V_{nl,l}^{ps}(r)$, different for each angular momentum l.

The use of pseudopotentials in VMC is not very problematic, and will not be explained in this thesis book. In DMC, however, the use of pseudopotentials is more problematic. If the Hamiltonian contains the nonlocal operator $V_{nl,l}^{ps}(r)$, the propagator contains matrix elements of the form $\left\langle R \left| e^{-\tau \widehat{V}_{nl}} \right| R' \right\rangle$, which may be positive or negative for any R, R', τ. Therefore, as the population of walkers evolve according to equation 2.22, the sign of a walker can change as the time evolves.

$$\partial_t f = \frac{1}{2}\nabla^2 f - \nabla \cdot (v_D f) - \frac{\left(\widehat{H} - E_T\right)\Psi_T}{\Psi_T} f + \left\{ \frac{\widehat{V}_{nl}\Psi_T}{\Psi_T} - \frac{\widehat{V}_{nl}\Phi}{\Phi} \right\} f \tag{2.22}$$

This is a problem similar to the sign problem presented before. In order to overcome this problem, the terms containing $V_{nl,l}^{ps}(r)$ are neglected, making in this way 2.22 formally equivalent to an imaginary-time Schrödinger equation with local potentials. This procedure is called pseudopotential localization approximation. If the trial wave function is a good approximation to the exact wave function, then the error is proportional to $(\Psi_T - \Psi_0)^2$ [133]. Therefore, it is important to use accurate trial wave functions. The Slater-Jastrow type wave functions introduced earlier are appropriate because they are accurate and easy to evaluate.

2.1.3 Dielectric Permitivitty and Solvent Effects.

Molecular properties and chemical reactions often vary considerably between the gas phase and in solution. For example, low lying conformations can have quite different energies in the gas phase and in solution (and in different solvents), conformation equilibria can differ, and reactions can take significantly different paths.

During this thesis project the Polarizable Continuum Model (PCM hereafter) was used for modeling system in solution. This approach represents the solvent as a polarizable continuum and places the solute in a cavity within the solvent. The Polarizable Continuum Model (PCM) by Tomasi and coworkers is one of the most frequently used continuum solvation methods and has seen numerous variations over the years. The PCM model calculates the molecular free energy in solution as the sum over three terms:

$$G_{sol} = G_{es} + G_{dr} + G_{cav} \tag{2.23}$$

These components represent the electrostatic (es) and the dispersion-repulsion (dr) contributions to the free energy, and the cavitation energy (cav). All three terms are calculated using a cavity defined through interlocking van der Waals-spheres centered at atomic positions. The reaction field is represented through point charges located on the surface of the molecular cavity (Apparent Surface Charge (ASC) model). The particular version of PCM that will be discussed here is the one using the United Atom for Hartree-Fock (UAHF) model to build the cavity. In this model the vdW-surface is constructed from spheres located on heavy (that is, non-hydrogen) elements only (United Atom approach). The vdW-radius of each atoms is a function of atom type, connectivity, overall charge of the molecule, and the number of attached hydrogen atoms. In evaluating the three terms in equation 2.23 this cavity is used in slightly different ways.

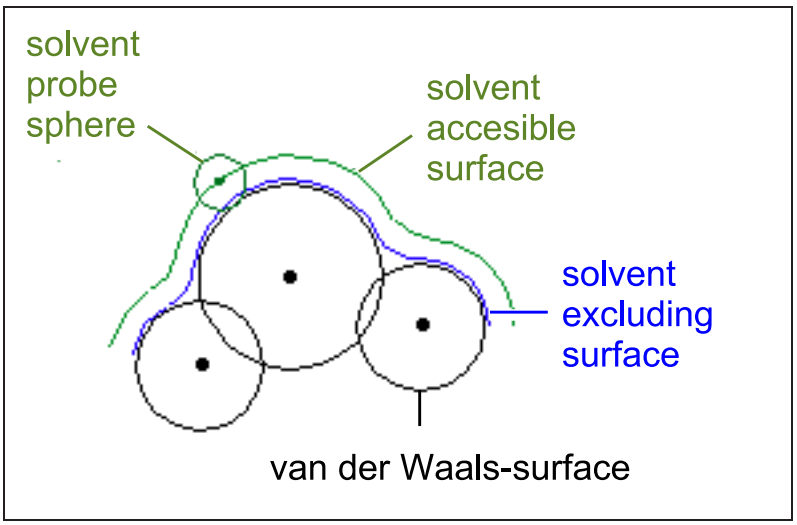


Fig. 2.1: Eschematic representation of the PCM model for modeling the solvent. Solvent probe sphere roles over the surface of the atoms, generating the solvent accesible and the solvent excluding surfaces.

While calculation of the cavitation energy G_{cav} uses the surface defined by the van der Waals-spheres, the solvent accessible surface is used to calculate the dispersion-repulsion contribution Gdr to the solution free energy. The latter differs from the former through additional consideration of the (idealized) solvent radius. The electrostatic contribution to the free energy in solution Ges uses an approximate version of the solvent excluding surface constructed through scaling all radii by a constant factor (e.g. 1.2 for water) and then adding some more spheres not centered on atoms in order to arrive at a somewhat smoother surface. Localization and calculation of the surface charges is approached through systematic division of the spherical surface in small regions (tesserae) of known area and calculation of one point charge per surface element.

2.2 Molecular Mechanics

As mentioned before, for large biological molecules like proteins, the description of energetics of the system relies on methods simpler than DFT, namely molecular mechanics. In these methods, the electronic structure is not explicitly calculated, but it is implicitly built up in the model by a set of fixed parameters (atomic charges, van der Waals parameters, force constants...), carefully obtained to reproduce experimental or ab-initio data. Molecular mechanical methods provide an analytical form for the potential energy of a given system as a function of the atomic coordinates, which can be efficiently computed and, therefore, opens the possibility to treat much bigger systems than those attainable by the ab-initio or DFT methods presented in the previous section.

2.2.1 Force Fields

The specific form of the analytical function and the corresponding set of parameters used in Molecular Mechanics are referred to as the force field. A common pictorial representation of a force field is one in which atoms are represented by spheres and the covalent bonds by springs. The value of the energy is calculated as a sum of several terms: i) internal or bonded terms (E_{bond}) and ii) external or nonbonded terms ($E_{nonbond}$). The former describes the covalent bonds, the angular deformations between covalently bound atoms and torsional energies in a molecule. The latter accounts for the interactions between non-covalently bonded atoms or atoms separated by 3 or more covalent bonds. In the molecular mechanics programme used in this work, the CHARMM program [134] (from Chemistry at HARvard Molecular Mechanics), the potential energy function ($V(r^N)$) is a function of the positions (r) of the N atoms of the system, and can be written as in Equation 2.24:

$$
\begin{aligned}
V(r^N) &= E_{bond} + E_{nonbond} \\
V(r^N) &= \sum_{\text{bond}} \frac{k_l}{2} (l_i - l_{eq})^2 \\
&\quad + \sum_{\text{angles}} \frac{k_\theta}{2} (\theta_i - \theta_{eq})^2 + \sum_{UB} \frac{k_{UB}}{2} (u_i - u_{eq})^2 \\
&\quad + \sum_{\text{torsions}} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] + \sum_{\text{improper}} \frac{k_\omega}{2} (\omega_i - \omega_{eq})^2 \\
&\quad + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\underbrace{4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]}_{\text{Lennard-Jones Potential}} + \underbrace{\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}}_{\text{Coulomb Potential}} \right)
\end{aligned}
\tag{2.24}
$$

Among the seven terms shown in equation 2.24, the first five correspond to the so-called "bonded" terms, whereas the last two terms correspond to the so-called "non-bonded" terms. All of them are explained below.

Bonded terms

The first term accounts for the change in the potential energy when the distance between covalently bounded atom pairs is changed. Since it is modeled through a harmonic potential, if the distance deviates (l_i) from the equilibrium bond length (l_{eq}), a quadratic penalty is introduced. The magnitude of this penalty is controlled through the k_l force constant, specific for each bond type, and determines its flexibility or rigidity.

The second term, or angle term, describes the changes in energy of the system upon variations in the angle (θ_i) formed by three atoms, $A - B - C$, where A and C are covalently bonded to B. Since this term is also modeled by a harmonic potential, it is required the definition of an equilibrium angle θ_{eq}, and k_θ is the force constant associated to a specific angular term. The third term, or Urey-Bradley (UB hereafter) term, is also associated with the angles, and takes into account the harmonic stretching of the $A...C$ distance u_i. k_{UB} and u_{eq} are the force constant and the equilibrium distance between atoms A and C respectively.

The fourth term describes the energy required for a torsion around a bond. This potential is periodic, V_n being the barrier controlling the amplitude of the torsional motion, n the periodicity and γ the phase that determines the values of the angles at which the energy vanishes.

Improper dihedrals constitute the fifth term. They are used for the description of out of plane motions. ω_i is the improper dihedral angle, ω_{eq} the equilibrium angle and k_ω the force constant assigned to a particular improper dihedral angle.

Non-bonded terms

In the CHARMM potential function the energy term representing the contribution of non-bonded interactions to the energy of the system has two components, the Van der Waals interaction energy and the electrostatic interaction energy. Some other potential functions also include an additional term to account for hydrogen bonds. In the CHARMM potential energy function, these interactions are included in the terms for the electrostatic and van der Waals interactions.

The van der Waals interactions are one of the most important forces for the stability of the biological macromolecules, and CHARMM computes them according to the sixth term in Equation 2.24. The van der Waals interaction is most often modeled using the Lennard-Jones 6-12 potential which expresses the interaction energy using the atom-type dependent constants σ and ϵ [135]. This interaction between two atoms arises from a balance between repulsive and attractive forces (see Fig. 2.2). The repulsive force dominates at short distances (nucleus-nucleus repulsion, electron-electron Pauli repulsion, etc.). The attractive force can be considered as arising from fluctuations in the charge distribution in the electron clouds. The fluctuation in the electron distribution on one atom or molecule gives rise to an instantaneous dipole which, in turn, induces a dipole in a second atom or molecule giving rise to an attractive interaction.

Each of these two effects is equal to zero at infinite atomic separation and become significant as the distance decreases. The attractive interaction is longer range than the repulsion but as the distance becomes short, the repulsive interaction becomes dominant. This gives rise to a minimum in the energy. The value of the energy at the minimum ($E^\star$), and the optimal separation of atoms ($r^\star$), which is roughly equal to the sum of Van der Waals radii of the atoms, depend on the atom chemical type.

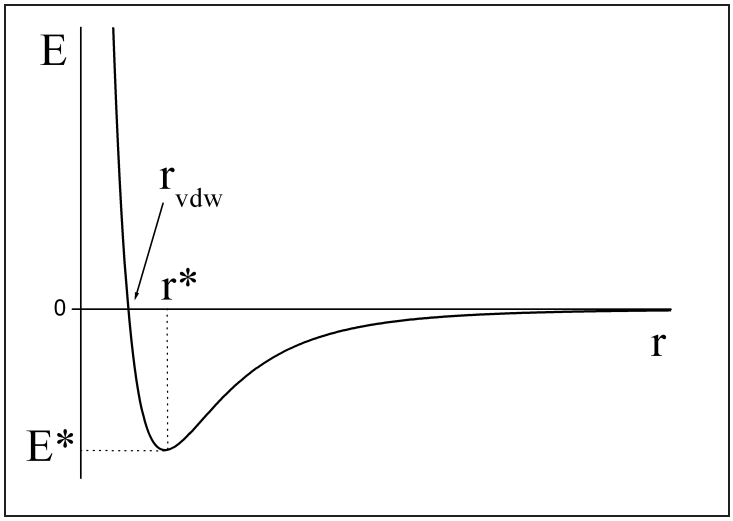


Fig. 2.2: Lennard-Jones potential for the van der Waals interaction. $r^\star$ is the optimal separation of two given atoms, and $E^\star$ is the energy at that distance.

The electrostatic interactions between a pair of atoms is represented by the Coulomb potential, as shown in the seventh term in Equation 2.24; $4\pi\epsilon_0$ is the term for the effective dielectric moment of the medium, r_{ij} is the distance between two atoms having charges q_i and q_j and ϵ_0 is the dielectric constant *in vacuo* [136].

2.2.2 Treatment of the Non-Bonded Interactions

The most time consuming part of an energy evaluation is the calculation of the non-bonded terms in the potential energy function. In principle, the non-bonded energy terms between every pair of atoms should be evaluated; in this case, the number of computation steps increases as the square of the number of atoms for a pairwise model (N^2). To speed up the computation, the interactions between two atoms separated by a distance greater than a predefined distance, the cutoff distance, are ignored, and are therefore equal to cero. Several different ways to terminate the interaction between two atoms have been developed over the years. Some of them will be explained in the paragraphs below.

Truncation: The interactions are simply set to zero for interatomic distances greater than the cutoff distance. This method can lead to large fluctuations in the energy and is therefore not often used.

Spherical truncation methods [137] represent a better option, in which the so-called cutoff functions are used. These functions are applied to each non-bonded contribution, as expressed

in Equation 2.25, where $V_{\mathsf{cutoff}(r)}$ is the non-bonded contribution with cutoff, V(r) is the non-bonded term without cutoff and F_{cutoff} is the cutoff function. There are several functional forms for the cutoff functions. The most important ones for the work developed in this thesis project are summarized below.

$$V_{\mathsf{cutoff}}(r) = V(r) \times F_{\mathsf{cutoff}} \tag{2.25}$$

The SHIFT cutoff method. The SHIFT function (SH in Equation 2.26) modifies the entire potential energy surface such that at the cutoff distance (r_{off}) the interaction potential is zero (Figure 2.3). The drawback of this method is that the potential well is slightly decreased.

$$\mathsf{SHIFT\ Function:} \left\{ \begin{array}{ll} \mathsf{for}\, r \leq r_{off}: & SH(r) = \left[1 - \left(\frac{r}{r_{off}}\right)^2\right]^2 \\ \mathsf{for}\, r > r_{off}: & SH(r) = 0 \end{array} \right\} \tag{2.26}$$

The SWITCH cutoff method. This method allows the interaction potential to taper over a predefined range of distances (Equation 2.27). The potential takes its usual value up to the first cutoff (r_{on}) and is then switched to zero when the two atoms considered are at a distance r between the first r_{on} and last cutoff r_{off} (Figure 2.3). This model suffers from strong forces arising from the switching region which can perturb the equilibrium structure. The SWITCH function (SW in Equation 2.27) is not recommended when using short cutoff distances.

$$\mathsf{SWITCH\ Function:} \left\{ \begin{array}{ll} \mathsf{for}\, r < r_{on} & SW(r) = 1 \\ \mathsf{for}\, r_{on} \leq r \leq r_{off} & SW(r) = \frac{\left(r_{off} - r\right)^2 \left(r_{off} + 2r - 3r_{on}\right)}{\left(r_{off} - r_{on}\right)^3} \\ \mathsf{for}\, r > r_{off} & SW(r) = 0 \end{array} \right\} \tag{2.27}$$

The FORCE SHIFT and Force SWITCH methods. Being similar to the functions specified above but they are applied to the forces rather than to the potential function.

Applying the Switch cutoff function the potential remains the same at short distances, therefore this method is principally used for the van der Waals interactions. For electrostatic interactions, the FSHIFT method was used in this work.

Finally, one should mention that there are several methods to go beyond the cutoff scheme in order to handle all long-range electrostatic interactions. These methods include the Ewald summation method [138, 139, 140, 141], the Extended Electrostatics approach [142] or Fast Multipole Methods [143, 144], but their study is beyond the goal of this thesis book.

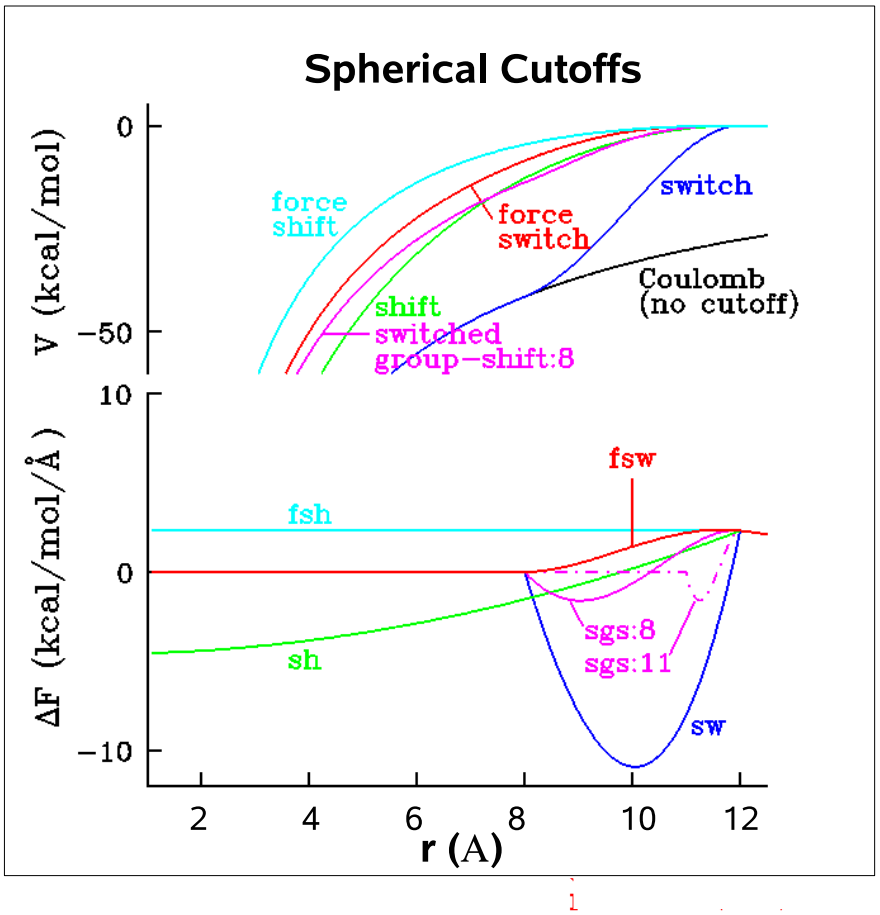


Fig. 2.3: The effects of the spherical cutoff functions on the electrostatic potential.

2.2.3 Limitations of the Empirical Potential Energy Function

In molecular mechanics the electronic structure is not explicitly taken into account, which leads to certain limitations. For instance, since covalent bonds are represented as harmonic springs, the creation or breaking of covalent bonds is not possible. Another limitation is due to the fixed set of atom types employed when determining the parameters for the force field. Atom types are used to define an atom in a particular bonding situation; an aliphatic carbon atom in an sp3 bonding situation has different properties than a carbon atom found in the Histidine (HIS hereafter) ring. Instead of presenting each atom in the molecule as a unique one described by unique set of parameters, there is a certain amount of grouping in order to minimize the number of atom types. This can lead to type-specific errors. The properties of certain atoms like aliphatic carbon or hydrogen atoms, are less sensitive to their surroundings and a single set of parameters may work quite well, while other atoms like oxygen, nitrogen and sulfur are much more influenced by their environment. These atoms require more types of parameters to account for the different bonding environments.

An approximation introduced to decrease the computational demand is the pair-wise additive approximation, i.e., interaction energy between one atom and the rest of the system is calculated as a sum of pair-wise (one atom to one atom) interactions, or as if the pair of atoms are invisible to other atoms in the system (see Equation 2.24). The simultaneous interaction between three or more atoms is not calculated in most contemporary force fields, so certain polarization effects are not explicitly included in the force field. This can lead to subtle differences between calculated and experimental results, for example, in the calculation of experimentally observable pK_a shifts of ionizable amino acid residue side chains induced by the electrostatic field of the whole protein. The development of polarizable force fields is currently an active area of research [145].

Another important point to take into consideration is that the energy based methods do not explicitly include entropic effects. Thus, a minimum value of energy does not necessarily correspond to the equilibrium, or the most probable state, which would only correspond to the minimum of free energy. Because of the fact that experiments are generally carried out under isothermal-isobaric conditions (constant pressure, constant system size and constant temperature) the equilibrium state corresponds to the minimum of Gibb's Free Energy, G. While just an energy calculation ignores entropic effects, these are included in a molecular dynamics simulations (see section 2.2.6, [135]).

Despite their limitations, current force fields are calibrated towards experimental results and quantum mechanical calculations of small model compounds. Force fields differ by the definition of their mathematical potential energy function and the values used for their associated parameters, such as: l_{eq}, k_l, θ_{eq}, k_θ, V_n, n and γ (see Equation 2.24). Among the most commonly used potential energy functions are CHARMM [146], Amber [147], GROMOS [148], OPLS [149] and MMFF [150] force fields. The ability of these force fields to reproduce physical properties measurable by experiment has been critically tested. The properties include structural data obtained from X-ray crystallography and NMR, dynamic data obtained from spectroscopy, inelastic neutron scattering and thermodynamic data. Force fields are also developed and adapted to different types of systems; there are for example all atom force fields for proteins [146], nucleic acids [151], organic molecules [152] and lipids [146] developed for the CHARMM program, similarly for the other force fields. The continuing development of force fields remains an intense area of research with implications for both fundamental research as well as for applied research in the pharmaceutical industry.

2.2.4 Energy Minimization

The potential energy plotted as a function of the atomic coordinates yields the potential energy surface. For a particular molecule, the energy minima on this surface correspond to stable conformations. Therefore, knowledge of the potential energy surface combined with determining the relative populations in the various energy minima provides an understanding of the energetics and eventually the thermodynamics of the molecular system. Thermodynamics provides an atomic scale understanding of molecular processes. Sampling the potential energy surface can be done by several different techniques.

For a given molecule one would like to find low energy conformations. For most molecules - except for very simple ones - the potential energy surface is a complicated multi-dimensional function that depends on the coordinates of the N atoms of the system and to find the global

potential energy minimum would require a search of the entire potential energy surface. All energy minimization methods are used to find local minima, thus the structures found by this approach are always close to the initial structure.

There are several different algorithms for minimizing the energy of the system [153, 154, 155]. They all involve calculating the first derivative and possibly the second derivative of the potential energy, and using this information to guide the coordinates to a lower energy conformation.

One of the simplest minimization algorithms is the *steepest descent* method [153, 155]. The coordinates are adjusted at each step of this iterative method in the negative direction of the gradient $\left(-\frac{dE}{dx}\right)$ where E is the energy and x refers to the coordinates. The step size is an adjustable parameter which determines how far the coordinates are shifted at each step. If the energy decreases, the step size is increased by 20 % to accelerate the convergence. If the energy rises, a minimum has been neglected and the step size thus halved. Although the steepest descent method does not converge readily for complicated functions, it rapidly improves very poor conformations [134].

A second method is the *conjugate gradient technique* which has better convergence characteristics [154] . The method is iterative and remembers the previous minimization steps as well as the current gradient to determine the next step. The conjugate gradient technique requires less energy and gradient evaluations to achieve the same reduction in energy as the steepest descent method.

During this thesis project the Steepest Descent method was used in most of the cases as minimization algorithm.

2.2.5 Monte Carlo Simulations

In the Monte Carlo (MC) method, the averages of the desired properties are calculated as in Equation 2.28, generating a representative sampling of the conformations of the system. <F> is the average of the desired property and is a function of the position of the N atoms of the system (r^N) and of the probability of the system to be in that specific configuration ($\rho_{NVT}(r^N)$). The probability of a configuration to exist in a canonical ensemble is calculated as in Equation 2.29, where V(r^N) is the potential energy of the system in a certain configuration.

$$< F > \;=\; \int dr^N F(r^N)\rho_{NVT}(r^N) \tag{2.28}$$

$$\rho_{NVT}(r^N) \;=\; \frac{e^{-\beta U(r^N)}}{\int dr^N e^{-\beta U(r^N)}} \tag{2.29}$$

Briefly, new conformations are randomly generated and then accepted or rejected depending on a certain criterion such that a Boltzmann distribution is obtained after the process is repeated for a sufficient number of steps. To do so, one uses the Metropolis algorithm [156], which can be summarized as in the scheme below.

---

1. Propose initial conformation of the atoms (r^N)
2. Calculate the potential energy $V(r^N)$ for the initial geometry.
3. Random generation of new geometry (r'^N)
   - Calculate potential energy for new configuration $V(r'^N)$
   - The new geometry is accepted (Acc) with probability,

     $$Acc(r^N \rightarrow r'^N) \;=\; min\left(1, e^{-\beta\left[V(r'^N)-V(r^N)\right]}\right)$$

     , i.e., a random number between $[0, 1]$ is generated (*rand*) and compared with the value of $Acc(r^N \rightarrow r'^N)$, the new configuration is accepted if,

     $$rand \;\leq\; Acc(r^N \rightarrow r'^N)$$

   - Calculate and accumulate desired properties
4. Repeat step 3 *NStep* times, where *NStep* is the number of required Monte Carlo steps that ensures a representative sampling of the configurational space r^N.
5. Calculate the average of desired properties.

After enough Monte-Carlo steps, the procedure ensures that the averages will correspond to the thermodynamic averages in the canonical ensemble (constant Volume, constant number of atoms and constant Temperature), i.e., the average of a certain property F will be expressed as in equation 2.28.

In this thesis work, the MC method was used for docking purposes, as explained in Chapter 5. Briefly, ICAM-1 was docked to the low affinity and high affinity αL I-domains, whereas the synthetic antagonist HC0303 was docked also to the high affinity αL I-domain. The resulting complexes were subjected to MD simulations and Free Energy Decomposition analysis.

2.2.6 Molecular Dynamics Simulations

One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics simulations. This computational method calculates the time dependent behavior of a molecular system. Like MC methods, MD provides a means to sample the potential energy surface of a molecule but in addition, this computational method calculates the time dependent behavior of the system. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. They are also used in the optimization of structures from X-ray crystallography and NMR experiments.

Molecular dynamics simulations are based on Newton's second law or the equation of motion, $\vec{F} = m\vec{a}$, where $\vec{F}$ is the force exerted on the particle, m is its mass and $\vec{a}$ is its acceleration. From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

More precisely, the trajectory is obtained by solving the differential equations embodied in Newton's second law (see Equation 2.30). This equation describes the motion of a particle of mass m_i along one coordinate (x_i) with F_{x_i} being the force on the particle in that direction. The force can also be expressed as the gradient of the potential energy (V), as shown in equation 2.31. Combining equations 2.30 and 2.31, equation 2.32 is obtained, where V is the potential energy of the system. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{2.30}$$

$$\vec{F_i} = -\vec{\nabla}_i V \tag{2.31}$$

$$\frac{-dV}{dx_i} = m_i \frac{d^2x_i}{dt^2} \tag{2.32}$$

The initial distribution of velocities is usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum (P) as expressed in equation 2.33. The velocities v_i are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability ($p(v_{ix})$) that an atom i has a velocity v_x in the x direction at a temperature T (Equation 2.34).

$$P = \sum_{i=1}^{N} m_i v_i = 0 \tag{2.33}$$

$$p(v_{ix}) = (\frac{m_i}{2\pi k_B T})^{1/2} exp^{-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}} \tag{2.34}$$

The temperature can then be calculated if we consider that at thermal equilibrium, the average kinetic energy per degree of freedom can be expressed as in equation 2.35, where v_α is the α component of the velocity of a given particle. We can use this relation to define an instantaneous temperature at time t, $T(t)$, as shown in equation 2.36.

$$\left\langle \frac{1}{2} m v_\alpha^2 \right\rangle = \frac{1}{2} k_B T \tag{2.35}$$

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_f} \tag{2.36}$$

where N_f is the number of degrees of freedom, which is 3N - 3 for a system of N particles with fixed total momentum. The relative fluctuations in the temperature will be of order $1/\sqrt{N_f}$. As N_f is typically on the order of $10^2 - 10^3$, the statistical fluctuations in the temperature are on the order of 5-10%.

Integration Algorithms

The potential energy is a function of the atomic positions (3N) of all the atoms in the system (N). Due to the complicated nature of this function, there is no analytical solution to the equations of motion, they must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion. The algorithm must conserve energy and momentum, should be computationally efficient and permit long integration time steps [135, 157]. All integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion (Equation 2.37, 2.38 and 2.39):

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2} a(t)\delta t^2 + \frac{1}{3!} b(t)\delta t^3 + ... \tag{2.37}$$

$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + ... \tag{2.38}$$

$$a(t+\delta t) = a(t) + b(t)\delta t + ... \tag{2.39}$$

where r is the position, v is the velocity (the first derivative with respect to time), a is the acceleration (second derivative with respect to time), b(t) is the third derivative and δt the time step in the Molecular Dynamics simulation. The algorithm used in the present work was mostly the *Velocity Verlet* algorithm. Among other integration algorithms the *Leap-Frog* or the *Beeman* algorithms [134, 135] are found.

Verlet Algorithm : The *Verlet* algorithm is derived as shown in equations 2.40 and 2.41. When summing these two equations we get equation 2.42

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \frac{1}{3!}b(t)\delta t^3 + \mathcal{O}(\delta t^4) \tag{2.40}$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 - \frac{1}{3!}b(t)\delta t^3 + \mathcal{O}(\delta t^4) \tag{2.41}$$

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 + \mathcal{O}(\delta t^4) \approx 2r(t) - r(t-\delta t) + \frac{f(t)}{m}\delta t^2 \tag{2.42}$$

The *Verlet* algorithm uses positions and accelerations at time t and the positions from time $t - \delta t$ to derive new positions at time $t + \delta t$, with an error that is of order δt^4. The Verlet algorithm uses no explicit velocities. The advantages of the Verlet algorithm are its straightforwardness and the modest store requirements. However, only moderate precision is obtained.

Leap-Frog Algorithm: In this algorithm, the velocities are first calculated at time $t + 1/2\delta t$. They are then used to calculate the positions r and velocities v at time $t + \delta t$ (equations 2.43 and 2.44). In this way, the velocities leap over the positions, then the positions leap over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time t can be approximated by the relationship shown in equation 2.45:

$$r(t+\delta t) = r(t) + v(t+\frac{1}{2}\delta t)\delta t \tag{2.43}$$

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + a(t)\delta t \tag{2.44}$$

$$v(t) = \frac{1}{2}[v(t-\frac{1}{2}\delta t) + v(t+\frac{1}{2})] \tag{2.45}$$

2.2.7 Boundary Conditions

A major obstacle for a simulation is the large fraction of molecules that lie on the surface of any small sample. For example, in the case of 1000 molecules arranged in a 10x10x10 cube, 488 molecules are located at the cube faces [157]. Molecules on the surface will experience completely different forces than those molecules inside the bulk.

Periodic Boundary Conditions

One way to overcome the problems linked to the finite arrangement in molecular dynamics simulations is the use of periodic boundary conditions (PBC hereafter). This method allows for the use of a relatively small number of particles in a simulation in such a way that the particles experience forces as though they were in a bulk solution. See, for example, the two dimensional box in Figure 2.4. The central box is surrounded by eight neighbors. The coordinates of the image particles, those found in the surrounding box are related to those in the primary box by simple translations. The simplest box is the cubic box. The box is replicated throughout space infinitely. Forces on primary particles are calculated from particles within the same box as well as in the neighboring image box. The cutoff is chosen such that a particle in the primary box is not affected by its image in the surrounding boxes. As a molecule moves in the original box during a simulation, the periodic image in each of the neighboring boxes moves in exactly the same way. If a molecule leaves the central box, one of its images will enter through the opposite face. There are no walls at the boundary of the central box and no surface molecules. The number density in the central box is conserved. It is not necessary to store the coordinates of all image boxes, only the the coordinates in the central box are stored [157].

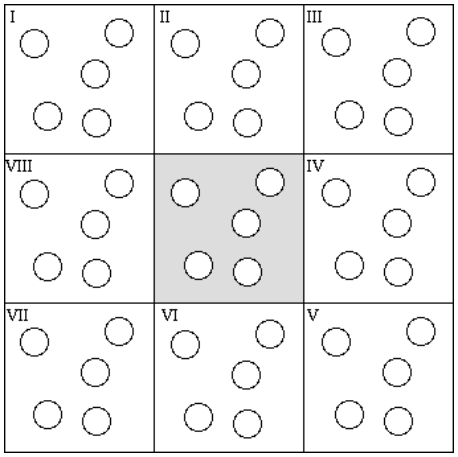


Fig. 2.4: Representation of a two-dimensional periodic system for a two dimensional box with only the first neighbors of the central box shown. Molecules can enter and leave each box across each of the four edges. In a three-dimensional example, molecules would be free to cross any of the cubic faces [157]. Figure from [135].

Limitations of this method are mainly patent for very small system size ($\sim$ 100 atoms) and for properties with important long-range contributions. PBCs however, have little effect on the equilibrium thermodynamic properties and structures [157].

In recent years, a number of models have been introduced which permit the inclusion of long-range electrostatic interactions in molecular dynamics simulation. For simulations of proteins

and enzymes in a crystalline state, the Ewald summation is considered to be the correct treatment for long range electrostatic interactions [157]. Variations of the Ewald method for periodic systems include the particle-mesh Ewald method [141].To treat non-periodic systems, such as an enzyme in solution methods based on multipole expansions have been developed. Many of these methods partition the electrostatic interaction into a long-range component and a short range component. The short range component is treated in the usual pairwise fashion while a multipole approximation is introduced to approximate the long-range electrostatic interaction; several such models have been developed [134, 142, 144]. Although these methods require more computer time than if one simply neglects the long-range part, they are significantly faster than if one does the N^2 summation of all the interactions and the results can be significantly improved.

2.2.8 Molecular Dynamics Simulations Protocols

To begin a molecular dynamics simulation, an initial conformation of the system has to be chosen. Most often, in simulations of biomolecules, an X-ray crystal structure or an NMR structure from the Brookhaven Protein Databank (http://www.rcsb.org/pdb/) is used as the initial structure. It is also possible to use a theoretical structure developed by homology modeling. The choice of the initial configuration must be done carefully as this can influence the quality of the simulation. It is often good to choose a configuration close to the state that you wish to simulate. The protocol followed in this work has generally been the following:

- Prior to starting a molecular dynamics simulation, an energy minimization of the structure is performed to remove any strong van der Waals interactions or steric clashes that may otherwise lead to local structural distortion and result in an unstable simulation.

- Explicit water molecules are added to solvate the protein. The solvating water molecules are usually obtained from a suitable large box of water that has been previously equilibrated. The entire box of water is overlayed onto the protein and those water molecules that overlap the protein are removed. At this point, another energy minimization is performed with the protein fixed in its energy minimized position to allow the water molecules to readjust to the protein molecule.

- Initial velocities at a low temperature are assigned to each atom of the system and Newton's equations of motion are integrated to propagate the system in time. For an explicitly solvated system, the protein positions are first fixed and the waters move to adjust to the presence of the protein. Once the waters are equilibrated, the constraints on the protein can be removed and the whole system (protein+water) can evolve in time. During the heating phase, initial velocities are assigned at a low temperature and the simulation is started. Periodically, new velocities are assigned at a slightly higher temperature and the simulation is allowed to continue. This is repeated until the desired temperature is reached (300 K).

- Once the desired temperature is reached, the simulation of protein/water system continues and during this phase several properties are monitored; in particular, the structure, the pressure, the temperature and the energy. The goal of the equilibration phase is to run the

simulation until these properties become stable with respect to time. If the temperature increases or decreases significantly, the velocities can be scaled such that the temperature returns to its desired value
.

- The final step of the simulation is to run the simulation in "production" phase for the time length desired. This can be from several hundred ps to 10 ns or more depending on the system size and the available computational resources. The main drawback of MD simulations arises from the duration attainable with current computational means with respect to the studied biological processes (see Table **2.1**).

Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)	Atomic fluctuations Sidechain Motions Loop Motions
Rigid Body Motions (1 to 10 Å, 10^{-9} to 1s)	Helix Motions Domain Motions (hinge bending) Subunit motions
Large-Scale Motions ($>$ 5 Å, 10^{-7} to 104 s)	Helix coil transitions Dissociation/Association Folding and Unfolding

Table **2.1**: Time scales for motions and processes in biological molecules.

2.2.9 Explicit and Implicit Solvation

The physiological environment of biomolecules is an aqueous solution. For the modeling of such molecules and their reactions, it is therefore important to account for solvation. Numerous approaches have been developed in the past to consider solvation in simulations. These approaches can be classified in two groups: *i*) an *explicit* introduction of water molecules in the simulation or *ii*) to account for bulk solvent effects *implicitly* through the use of a dielectric continuum model. The first method is computationally expensive, and for large systems, it can severely limit the length of the simulation. This method becomes important when the solvent interacts specifically with the solute by forming hydrogen bonds, for example. If the MD becomes to expensive, implicit solvent models are useful to account for the different dielectric properties of the solvent [158].

In this work, both explicit solvation and implicit solvation protocols have been followed. All MD simulations have been carried out with explicit solvation. The number of variables to consider in simulating a system is 3 x N (where N = number of solvent molecules). The complex between the high affinity αL I-domain and ICAM-1 contains 4000 atoms approximately. Solvating the system required on the order of 13500 water molecules, i.e, approximately additional 40000 atoms. In addition, the necessary amount of ions (Cl- or Na+) are placed in the cubic water box in order to neutralize the system. The positioning of the ions is guided by the biopolymer charge distribution.

The implicit solvation/finite difference model has been used in this thesis project in the following cases: i) For the docking of ICAM-1 and small antagonists to the αL I-domain, ii) for the calculation of the binding free energy between ligands and protein (see chapter 5) following the MM/PBSA protocol (see later), iii) for the determination of the protonation states of the titratable residues in proteins (see chapter 2.2.10). A detailed account of the finite difference method can be found in [135, 159]. The two principal methods for implicit solvation are based on the Poisson-Boltzmann [160] (chosen in this thesis project) and the generalized Born model [161]. In continuum electrostatic methods, the solvent is represented as a high dielectric continuum (for water $\epsilon \approx 80$) and the solute, i.e. the protein, is characterized as a low dielectric body (ϵ between 1-20) by its charge density $\rho(r)$. In such a system, the electric potential $\Phi(r)$ can be obtained from the Poisson equation shown in equation 2.46, where $\epsilon(r)$ is the dielectric varying in space as a result of the dielectric boundary between protein and solvent.

$$\nabla.\epsilon(r)\nabla\Phi(r) = -\rho(r) \tag{2.46}$$

Effects of mobile ions can be included into this equation. The Debye-Hückel theory establishes that the average charge density at any point arises from the competition between the electrostatic attraction of the central ion for its counterions and the disruptive effects of thermal motion [162]. The Boltzmann distribution then gives the proportion of ions at this location relative to the proportion in the bulk solution.The term is then $n(r) = Nexp(-V(r)/k_BT)$, where $n(r)$ is the number density of ions at location r,N is the bulk number density and $V(r)$ is the energy change necessary to bring the ion from infinity to the position r. The final equation is then referred as the *Poisson-Boltzmann* equation and is expressed as in equation 2.47, where q_i and n_i are the charge and the concentration of the ith ionic type, respectively [135, 159]. For simplification of the calculation, the exponential can be approximated by the linear term in the Taylor series expansion (first term). This yields the linear Poisson-Boltzmann equation (see equation 2.48), where κ is the inverse Debye-Hückel screening length $\kappa^2 = 2Ie^2/k_BT\epsilon(r)$. $I = \sum \frac{1}{2}(\frac{q_i^2}{e^2}n_i)$ is the ionic strength, and e is the unit electric charge.

$$\nabla.\epsilon(r)\nabla\Phi(r) - \sum_i q_i n_i exp(q_i\Phi/k_BT) = -\rho(r) \tag{2.47}$$

$$-\nabla.\epsilon(r)\nabla\Phi(r) + \epsilon\kappa^2\Phi(r) = \rho(r) \tag{2.48}$$

Unless the molecule's shape is spheric/elliptic or of another simple geometry, this equation can only be solved by numerical approaches [135, 163]. One of these approaches is the finite difference approach used in UHBD program [159]. It consists in discretizing the space into a cubic grid around the simulation system. Values for the electrostatic potential, the charge density, the dielectric constant and ionic strength are assigned to each grid point. The atomic charges do not usually coincide with a grid point and so the charge is allocated to eight surrounding grid points [135].

The solution of the Poisson-Boltzmann equation gives the potential $\Phi(r)$ in space. The free energy for assembling the charge distribution of protein in the presence of the dielectric discontinuity is the $\Delta G = \frac{1}{2}q_i\Phi_i$. This allows the estimation of thermodynamic parameters from Poisson-Boltzmann solution.

2.2.10 pK_a Calculations

The pK_as and pH-dependent ionization states of titratable groups play an essential role in the catalytic activity and stability of proteins [164, 165, 166]. The proper choice of protonation states for titratable groups can therefore have an important influence on the simulation.

Different methods exist for predicting the pK_a's. For small compounds, quantum mechanical methods can be applied. Simulation methods with clear physical and chemical concepts may be applied to systems with a limited number of protonation states. Protonation states of proteins with several titratable sites are often assigned based on visual analysis. With the development of continuum dielectric models it is now possible to calculate the pK_a values for proteins with several titration states. Using these methods, excellent results were obtained in the prediction of pK_a shifts induced by charge perturbations and the prediction of absolute pK_a values [167, 168]. In this work, the protocol based on the continuum electrostatics is used for the pK_a calculations, and the results were used to assign the protonation state of titrable groups for the simulations.

pK_a and Amino acids

One of the more frequent chemical reactions that occur in a protein-water solution is the uptake and release of protons by amino acids. Those fragments of amino acids that can absorb or release protons will be referred to as titratable protein groups. Thus, an Aspartic acid side chain contains a single titratable group (namely the carboxy group: COO-), and an N-terminal Lysine contains two titratable groups (the N-terminal amino group, and the amino group in the side chain).

Some amino acids in proteins possess acidic or basic titrable groups: Asp, Glu, Cys, Ser, Tyr, Thr and the C-terminus are acidic, while the N-terminus, His, Lys and Arg act like bases. The basic titratable protein groups are positively charged in their protonated state, and the acidic protein titratable groups are neutral in their protonated state.

The definition of acid / base always depends on the specific chemical reaction concerned. In the following example (equation 2.49), HA is an acid in reacting with water, but in the reaction with BH (equation 2.50), it behaves as a base:

$$HA + H_2O \rightarrow A^- + H_3O^+ \tag{2.49}$$

$$HA + BH \rightarrow H_2A^+ + B^- \tag{2.50}$$

In proteins, the equilibrium associated with specific protonation / deprotonation reactions depend on the prevalence of a certain species. For acids (HA), the reaction and the associated equilibrium constant (K_a) is expressed as in equation 2.51. The pK_a is defined as in equation 2.52.

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{2.51}$$

$$pK_a = -\log K_a \tag{2.52}$$

If the pK_a value of a protein titratable group is known, the charge or protonation state of this group can be predicted at a given pH value. Taking into account that $pH = -\log([H_3O^+])$, and rewriting equation 2.51, we obtain the Henderson-Hasselbach equation (equation 2.53) and the number of bound protons as a function of pH can be expressed as in equation 2.54 [165], where θ is the degree of proton occupancy. The pK_a is the pH at which 50 % of the proton site is occupied.

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \tag{2.53}$$

$$\theta(pH) = \frac{1}{1 + e^{-ln10(pK_a - pH)}} \tag{2.54}$$

The pK_a values of protein titratable groups in water have been estimated by comparing them with the pK_a values for model compounds in water (see Figure 2.5). When studying enzymatic mechanism and protein stability, pK_a calculation techniques provide a way of determining the effect of the protein environment on pK_a values of titratable groups in the protein.

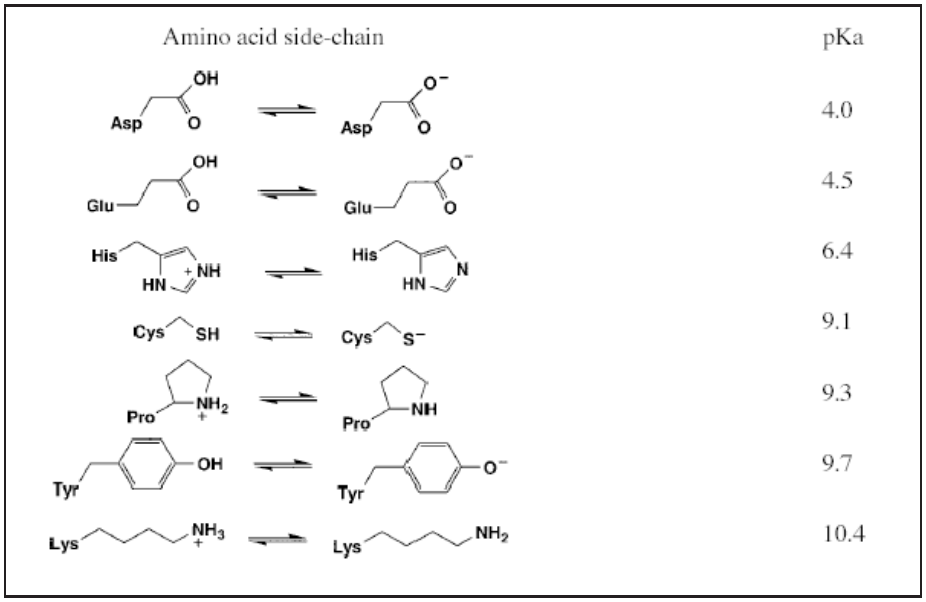


Fig. 2.5: Estimates of acid dissociation constants for amino acid residues in a polypeptide. Figure from [169].

Calculation of pK_a

The pK_a value of a titratable group is a measure of the free energy difference between the neutral and charged state of the group. It is therefore possible to calculate the pK_a value of a group if the free energy difference between the charged and neutral state of that group in the protein can be calculated. In continuum electrostatics for pK_a, it is assumed that electrostatic interactions with the protein are responsible for pK_a shifts of charged amino acids in solvent alone and in the protein. Therefore, to estimate the pK_a shifts, one uses continuum electrostatic calculations to evaluate the energetic effect of the protein environment on the neutral and charged state of each titratable amino acid. In what follows we briefly outline the method used for calculating titration curves; details are given in the work by Schaefer and coworkers [164].

The protonation state of a system with N titrating sites can be described with a state vector $\vec{s} = (s_1, ..., s_N)$, where s_i is an index that runs over all the protonation states of site i. In macromolecular systems, the titrating sites are interacting with each other, and their interaction with the solvent is altered as compared with the sites in the isolated model compounds (M_i) for which the standard pK_a's are known. This leads to *shifted* pK_a's and titration curves of individual sites that are perturbed from the isolated model compound in solution.

The pH-dependent probability $P(s_i')$ of finding site i in state s_i' is given by the ensemble average in equation 2.55, where $\Xi(\mathrm{pH})$ is the *generating function*. The macroscopic analogue of the partition function, $G(\vec{s}, \mathrm{pH})$) is the pH-dependent electrostatic free energy of the protonation state $\vec{s}$ of the protein and it is determined by equation 2.57, where $E_{el}(\vec{s})$ and $E_{el}^{M_i}(s_i)$ are the electrostatic free energy of the protein and of the model compound M_i in the protonation state $\vec{s}$ and s_i, respectively, $n_i(s_i)$ is the number of protons bound to site i in charge state s_i and $pK_{a,i}^{\mathrm{stnd}}(s_i)$ is the pK_a for the reaction.

$$P(s_i')(\mathrm{pH}) = \frac{1}{\Xi(\mathrm{pH})} \sum_{\vec{s}} \delta_{s_i s_i'} \exp(-\beta G(\vec{s}, \mathrm{pH})) \tag{2.55}$$

$$\Xi(\mathrm{pH}) = \sum_{\vec{s}} \exp(-\beta G(\vec{s}, \mathrm{pH}))\,; \quad \beta = \frac{1}{k_B T} \tag{2.56}$$

$$\begin{aligned} \Delta G(\vec{s}, \mathrm{pH}) &= ln(10) k_B T \sum_{i=1}^{N} (n_i(s_i)\mathrm{pH} - pK_{a,i}^{\mathrm{stnd}}(s_i)) + E_{el}(\vec{s}) - E_{el}(\vec{0}) \\ &\quad - \sum_{i=1}^{N} \left(E_{el}^{M_i}(s_i) - E_{el}^{M_i}(0) \right) \end{aligned} \tag{2.57}$$

The electrostatic free energies of the protein and the model compounds are calculated from the electrostatic potential ϕ. To determine ϕ, the solvent and protein are treated as polarizable continua with dielectric constants $\epsilon_s = 80$ and $\epsilon_i = 20$, respectively, and the linearized Poisson-Boltzmann (Eq. 2.48) is solved by finite-difference methods [159, 170] as explained in section 2.2.9.

In all calculations, the ionic strength was set equal to a physiological value of 145 mM. Given the potential $\phi(\vec{s})$ of the protein (or of a model compound) in protonation state $\vec{s}$, the electrostatic free energy is obtained as in equation 2.58 over all atoms k of the solute; $q_k(\vec{s})$ is the partial charge of atom k which depends on the protonation state for atoms belonging to titratable sites.

$$E_{el}(\vec{s}) = \sum_{k} q_k(\vec{s}) \phi(\vec{s}) \tag{2.58}$$

The potential ϕ is calculated as the sum of the contributions from the individual sites (including the non-titrating "background atoms"), $\phi(\vec{s}) = \sum_i \phi_i(s_i)$. A grid with spacing (distance between grid points) of 0.3 Å is used to calculate ϕ_i in the near-range from 0 to 3 Å from any atom of site i; in the long range, coarser grids with spacings of 0.6 Å (distances up to 10 Å), 1.2 Å (distances up to 20 Å), and 2 Å (distances greater than 20 Å) are used. This "focusing" technique leads to converged results for the potential while requiring considerably less computer time than a single calculation of the potential where a grid with spacing 0.3 Å spans the entire system [159].

Titration curves for all the sites are calculated according to Eq. 2.55. For the electrostatic potential and titration curve calculations, we require the standard pK_a for the site in a model system, and the partial atomic charges for the different protonation states of the site. For the amino-acid residues, the pK_a's and partial charges are taken from the work by Schaefer and coworkers [164]. The van der Waals parameters for the atoms are taken from the CHARMM 22 parameter set [146]. They are used in the Poisson-Boltzmann calculations to determine the interior of the protein defined as the volume that is inaccessible to a solvent probe sphere with radius 1.4 $\mathring{A}$.

2.2.11 Free Energy Decomposition

The MM-PBSA method

To analyze the free energy of binding of two proteins, the interface among them or to identify possible hotspots in the interaction, an analysis based on the Molecular mechanics-Poisson Boltzmann Surface Area method (MM-PBSA hereafter) [171, 172, 173] was performed in this work. To do so, a thermodynamic cycle is considered (Scheme 1), according to which the binding free energy is estimated as the sum of the gas-phase energies and solvation free energies.

$$\begin{array}{ccccc} \mathsf{LFA\text{-}1} & + & \mathsf{ICAM\text{-}1} & \overset{\Delta G^{0}_{\mathsf{bind}}}{\rightarrow} & [\mathsf{LFA\text{-}1/ICAM\text{-}1}]_{\mathsf{compl}} \\ \Delta G^{\mathsf{LFA\text{-}1}}_{\mathsf{solv}} \downarrow & & \Delta G^{\mathsf{ICAM\text{-}1}}_{\mathsf{solv}} \downarrow & & \downarrow \Delta G^{\mathsf{compl}}_{\mathsf{solv}} \\ \mathsf{LFA\text{-}1} & + & \mathsf{ICAM\text{-}1} & \overset{\Delta G^{s}_{\mathsf{bind}}}{\rightarrow} & [\mathsf{LFA\text{-}1/ICAM\text{-}1}]_{\mathsf{compl}} \end{array}$$

Scheme 1. Thermodynamic cycle used to estimate the free energy of binding of ICAM-1 to the αL I-domain.

The free energy of a molecule in the MM-PBSA approach [171] is estimated according to equation 2.59, where E_{MM} is the total molecular mechanical energy, G_{PBSA} the solvatation free energy and -TS_{MM} the entropic contribution at a given temperature. The solvatation free energy is divided in an electrostatic part and non-polar or hydrophobic contribution (equation 2.60).

$$G_{molecule} = E_{MM} + G_{PBSA} - TS_{MM} \tag{2.59}$$

$$G_{PBSA} = G_{PB} + Gnp \tag{2.60}$$

As an approximation, explicit conformational entropy contributions to the thermodynamics of binding are neglected, although the use of various structures generated in the MD simulations can partially account for entropy. The binding free energy was then calculated according to equation 2.61:

$$\Delta G = \Delta E^{vdw} + \Delta E^{elec} + \Delta G^{np}_{solv} + \Delta G^{elec}_{solv} \tag{2.61}$$

where ΔE^{vdw} and ΔE^{elec} are the electrostatic and van der Waals contributions to the energy of the complex formation in gas phase, respectively, and ΔG^{np}_{solv} and ΔG^{elec}_{solv} are the non polar and electrostatic contributions, respectively, to the solvation free energy. Each contribution to the binding free energy entering equation 2.61, was decomposed in a sum of individual contributions per residue, as described later in the text. Another simplification introduced concerns the neglect of conformational changes between the bound and unbound proteins; as the root mean squared (RMS hereafter) difference between the backbone atoms of free ICAM-1 and ICAM-1

bound to the medium affinity (M.A. hereafter) αL I-domain is small (data not shown), the results are still expected to be satisfactory in the context of identifying interaction energy hot spots in the protein-protein complex, which is the goal of the present study.

- *van der Waals interactions,* ΔE^{vdw}: Upon protein complexation, both attractive and repulsive van der Waals interactions are formed between amino acids. Attractive van der Waals forces involve the interactions among induced dipoles, which arise from fluctuations in the charge densities of adjacent uncharged non-bonded atoms. Repulsive van der Waals forces are the result of the electron-electron repulsion events that occur as two clouds of electrons begin to overlap when uncharged non-bonded atoms of different amino acids come very close together but do not induce dipoles. Although van der Waals forces are extremely weak relative to other forces governing complex formation, it is the huge number of such interactions that occur in large protein molecules that make them significant to the protein-protein interaction.

- *Non polar desolvation,* $\Delta G_{np}^{complex}$: The hydrophobicity of certain amino acid R-groups, is responsible for their unfavorable interaction with the water molecules in the environment. On protein complex formations such interactions are broken and favor the binding process.

- *Electrostatic interactions,* ΔE^{elec}: Typical charge-charge interactions that favor protein complex formation are those between oppositely charged R-groups such as K(Lys) or R(Arg) with D(Asp) or E(Glu). The opposite is true, i.e. formation of unfavorable contacts, for equally charged R-groups. Upon complexation, there is a change in the number and the environment of these pair-wise interactions, which should globally counterbalance the unfavorable electrostatic desolvation term explained below.

- *Electrostatic desolvation,* ΔG_{solv}^{elec}: This term arises from the disruption, before complex formation, of charge-dipole and dipole-dipole interactions that exist between ionized R-groups or the slight dipole moment in polar R-groups of amino acids, respectively, with the dipole of the water molecule. This is, then, a term that should unfavor complex formation.

ΔE^{vdw} and $\Delta G_{np}^{complex}$ were computed using the CHARMM22 force filed. For the former, ΔE^{vdw}, the 6-12 Lennard-Jones potential was used. The individual contributions to the van der Waals interaction energy of each residue of ICAM-1 was estimated by calculating the interaction with the set of LFA-1 residues located in a 12.5 Å cutoff distance, and vice-versa for LFA-1. For $\Delta G_{np}^{complex}$, as the energetics of desolvation non-polar surfaces has been observed to be proportional to the loss in Solvent Accessible Surface (SAS hereafter) of both proteins upon binding, the accessible surface was calculated with a probe radius of 1.4 Å. We determined the accessible surface of each residue in the free molecules and in the complex and then we calculated the accessible surface area difference between the free molecule and the complex for each residue. The constant γ used is 5 cal mol^{-1} [172, 174, 175].

$$\Delta G_{np}^{complex} = \gamma * (SAS^{complex} - (SAS^{protA} + SAS^{protB})) \tag{2.62}$$

To compute ΔE^{elec} and ΔG_{solv}^{elec} we used the UHBD program [159], that solves the linearized Poisson-Boltzmann equation using a finite-difference method (FDPB hereafter) . A shell script which includes interfaces to UHBD was used to compute electrostatic free energies of binding of the two molecules. The limit between the two regions with different dielectric constant is determined by the molecular surface. Van der Waals radii and charges of atoms are obtained from parameters of the force field CHARMM22. A constant dielectric of 80 was used for the solvent and 1 for the protein. The mesh grid over which charges are located during calculations is scaled down to 0.35 . We followed the implementation proposed for macromolecular complexation by Hendsch and Tidor [176].

Introduction to the experimental part of this thesis

The goals of the experimental part of this thesis projet were the following:

- To confirm the theoretical predictions and *in vivo* results that suggest the existence of binding between our compounds and the αL I-domain.
- If this binding can be observed, to structurally characterize the binding by proposing a binding site in the αL I-domain.
- To measure the strength of the observed binding.
- To confirm the prediction about the ability of the compounds to inhibit interaction of the αL I-domain with ICAM-1.

To achieve these goals Nuclear Magnetic Resonance was chosen among other techniques, i.e, x-ray crystallography or electronic microscopy. The main reasons for having chosen it are that i) the "small-to-medium" size of the I-domain makes of NMR a suitable technique; if the protein were much larger, other techniques may be best suited; ii) the type of experiment necessary to localize the binding site is relatively fast and is also used in the second step of measuring the binding Kd, as will be explained later in the text.

Since the technique of NMR suffers from an intrinsic low sensitivity, it was necessary to purify a large amount of protein from bacterial cells. A brief summary of the purification protocol is therefore presented in this section, together with a brief summary of the basis of the NMR methodology, as well as the protocols used to acquire, process and analyze the NMR spectra. The strength of the binding was estimated by NMR using Chemical Shift Perturbation experiments and complemented by an alternative technique such as Surface Plasmon Resonance (SPR hereafter). The reader will therefore find a brief explanation of the basis of these two methodologies and the specific protocols followed.

All the work presented in this section and in chapter 6, was carried out in the NMR group at the Centro Nacional de Investigaciones Oncológicas (CNIO, National Institute for Cancer Research) in Madrid, under the supervision of Dr. Francisco Blanco and the PhD student Tahl Zimmerman. Some of the results may not be conclusive yet, and further work is still needed to confirm the hypotheses proposed. This work is still being carried out by the researchers at the CNIO.

2.3 Nuclear Magnetic Resonance

2.3.1 Nuclear Spin and Resonance

Nuclear Magnetic Resonance is a powerful tool employed in many branches of science to understand and characterize the relations between chemical structure, molecular dynamics and the macroscopic properties of the systems. Along with the mass and the charge, the atomic nuclei posses a property called spin moment ($\vec{I}$) that generates a magnetic moment ($\vec{\mu}$) parallel and proportional to the spin. The spin moment is expressed as

$$\vec{\mu} = \gamma * \vec{I} \tag{2.63}$$

with γ being the magnetogiric ratio, an intrinsic characteristic of every atom, which may be viewed as a measure of how 'strongly magnetic' a nucleus is. The spin moment of a nucleus may have values greater than or equal to zero, but always multiple of 1/2. Those with I=0 do not exhibit nuclear magnetic resonance and are therefore termed 'NMR silent'. When a nucleus with a spin different from zero is placed in a magnetic field (Bo), its magnetic moment interacts with the field, producing a split of its energy levels (Zeeman levels). In the case of spin-half nucleus (such as ^{1}H, ^{13}C or ^{15}N), the splitting of all the nuclei in the system is done into two energy levels (Figure 2.6), which classically corresponds to some nuclei taking up one of two possible orientations with respect to the static field, either parallel or antiparallel, the former being lower in energy.

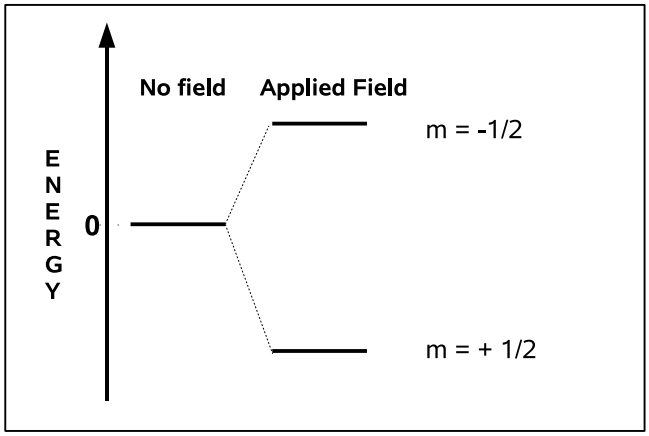


Fig. 2.6: Zeeman Levels: Splitting of the energy levels of the magnetic moments of spin-half nuclei, under the effect of an external magnetic field (Bo).

To produce transitions between the Zeeman levels, the nuclei are irradiated with a electromagnetic radiation with a frequency (Larmor frequency) equivalent to the energy difference between the levels. In principle, the absortion frequency or frequency of resonance (ω_o) only depends on the type of nucleus and the external magnetic field (Equation 2.3.3). Nevertheless, the electronic environment around each nucleus generates a local field that interferes with the principal magnetic field (Bo), which provokes that nuclei of the same type emit signals of different frequencies. In addition, the nuclear spins can also interact with each other through the chemical bonds (scalar coupling J) or through the space (dipolar coupling D). The different lines observed in the NMR spectra are caused by the differences in the frequency of resonance of each chemically distinct nucleus, as well as by the scalar or dipolar couplings between the different nuclei in a system. The big amount of information collected in the spectra makes of NMR a fundamental tool for solving structural and dynamic problems of molecules.

$$\omega_0 = \gamma * B_0 \tag{2.64}$$

2.3.2 The NMR Signal and the Spectrum

The excited nuclei of a molecule relax with emission of energy, which is ultimately detected by the NMR coil as tiny electrical signals, known as the NMR signal. The latter decays exponentially with time, producing the observed Free Induction Decay (FID hereafter) which carries information of the chemical environment of each nucleus (Figure 2.7). The FID signal is converted into the frequency-domain by a mathematical algorithm known as Fourier Transformation (FT hereafter). The new spectrum generated contains one or several signals in the frequency domain per nucleus excited (Figure 2.7). The position of each signal with respect to a reference compound which is assigned the chemical shift of zero is known as the Chemical shift (δ). The chemical shift is usually expressed in parts per million (ppm hereafter), is calculated as in equation 2.65 and its value does not depend on the operating frequency of the magnet. The chemical compound mostly used as reference in proton- and carbon-related NMR spectroscopy of proteins is 2,2-Dimethyl-2-silapentane-5-sulfonate (DSS hereafter).

$$\delta = \frac{\text{Frequency of nuclei - Frequency of reference}}{\text{Frequency of reference}} 10^6 \tag{2.65}$$

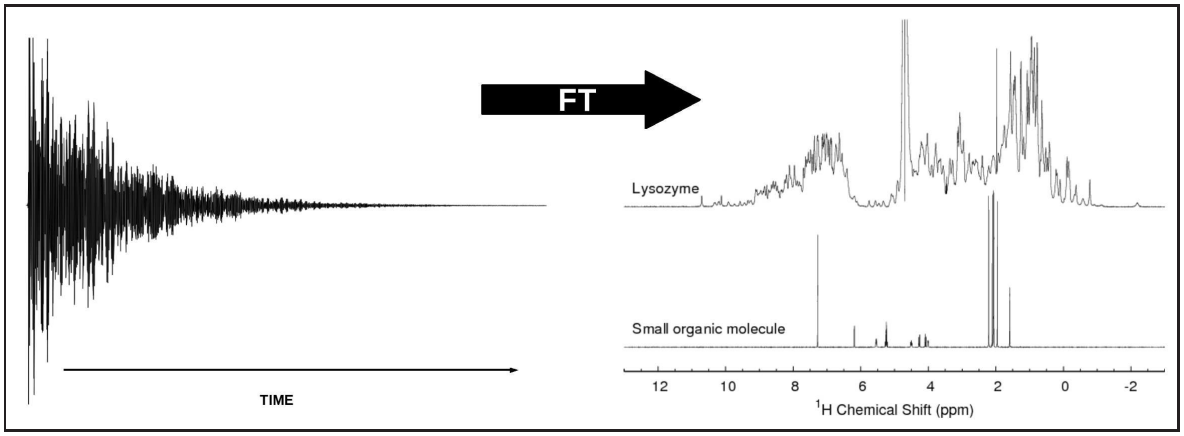


Fig. 2.7: (Left) Typical FID signal of the relaxation of the nuclei (protons, for example) in a NMR sample and (right) two examples of a Fourier transformed spectrum in the frequency domain. The ^{1}H-monodimensional spectrum of a protein such as lysozyme (top-right) is significantly more complex than that of a small organic molecule (bottom-right).

2.3.3 Heteronuclear Shift Correlation

As shown in Figure 2.7 (right) whereas a ^{1}H monodimensional spectrum of a small organic molecule is relatively simple (since it contains few protons), for a bigger molecule such as the protein lysozyme (which may contain hundreds of protons), the complexity of the spectrum increases to such an extent that the use of additional dimension(s) is necessary to handle the information provided. One example of such an experiment would be the so-called Heteronuclear Single Quantum Correlation (HSQC hereafter) spectrum. The latter is a 2D experiment where one of the axis usually corresponds to ^{1}H and the second axis to a heteronucleus. During this thesis project, ^{1}H^{15}N HSQC spectra were usually acquired (Figure 2.8), which is especially useful in the case of proteins, as will be explained in the following lines. The signals in a ^{1}H^{15}N HSQC spectrum correspond to a correlation between the amide proton and the amide nitrogen of each amino acid. Their position in the 2D spectrum is given by the ^{1}H frequency of the amide proton and the ^{15}N frequency of the N atom bound to it. The advantages of a ^{1}H^{15}N HSQC spectrum with respect to a ^{1}H monodimensional spectrum are the following: i) the number of signals (peaks) is substantially reduced, ii) the overlapping of signals is mostly broken and iii) the ^{1}H^{15}N HSQC can be regarded as a fingerprint of each protein. The number of signals is reduced because a ^{1}H^{15}N HSQC spectrum only contains a signal per proton attached to a nitrogen. Since each residue of a protein (except proline) generally has a unique amide proton the number of peaks in a ^{1}H^{15}N HSQC spectrum is close to the number of residues in the protein. This number may be slightly increased by signals coming from amino acid sidechains that contain nitrogen, such as ARG or GLN. This reduced number of signals is spread in a 2D map, which is representative of each protein (fingerprint), and provides information with respect to the quality of the purified construct. If the protein is folded, the amide bonds in the protein would have a particular chemical environment, and the peaks will usually be well dispersed. The NH amide groups of an unfolded protein all have a similar chemical environment (solvent/NMR buffer), and most of the peaks may lay on a similar area of the spectrum or even be overlapped.

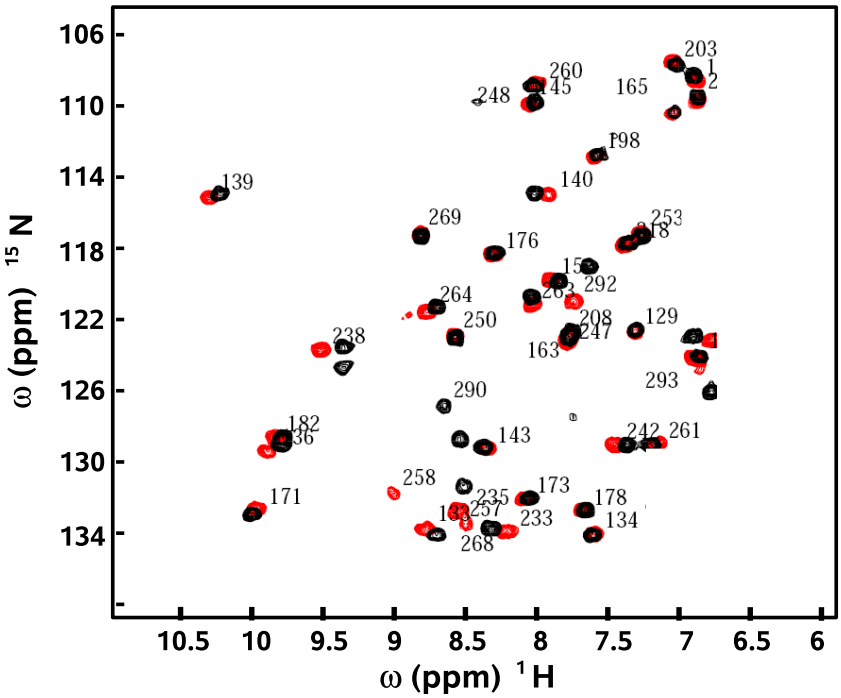


Fig. 2.8: Overlay of two $^{1}H^{15}N$ HSQC spectra of a well folded protein, before(red) and after (black) addition of ligand. Labels represent the sequence number of the amino acids in the protein. As observed, some residues undergo a significant chemical shift perturbation upon addition of the ligand.

Localizing binding sites: The success of the NMR technique is mainly due to the high sensitivity of the chemical shift of a nucleus to changes in its chemical environment. In this sense, a small conformational change in a protein is reflected in a change on the chemical shift of the signals in the $^{1}H^{15}N$ HSQC spectrum. A HSQC experiment can be used, therefore, as a relatively fast way of detecting possible interactions of a protein with ligands, such as other proteins, peptides or small molecular weight compounds. By comparing the $^{1}H^{15}N$ HSQC spectrum of a ligand-free protein with the $^{1}H^{15}N$ HSQC spectrum of the ligand-bound protein, it is sometimes possible to find changes in the chemical shifts of certain peaks, which is most likely to occur to the residues in the binding interface (Figure 2.8).

Estimating binding constants: If multiple HSQC spectra are acquired by gradually increasing the amount of ligand, the chemical shift of the signals that have gradually moved will reach a maximum perturbation (Figure 6.17, right signal). In other words, addition of extra amount of ligand will not provoke a chemical shift perturbation of the ligand binding residues anymore. Plotting the CSP versus the ligand concentration, the strength of the binding of the ligand to the protein can be estimated. In sumary, both the ligand binding site and the strength of the binding can be obtained by acquiring a series of relatively short $^{1}H^{15}N$ HSQC spectra.

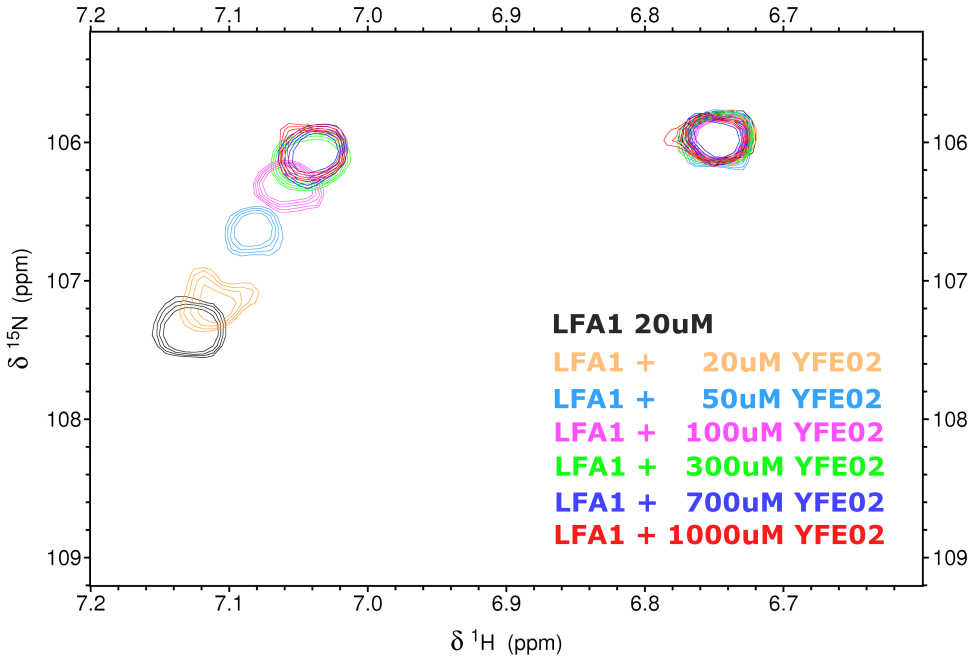


Fig. 2.9: Titration experiment: Overlay of multiple $^{1}H^{15}N$ HSQC spectra of protein LFA-1 in the presence or absence of a small molecular weigth inhibitor (YFE02). For clarity, only a small area of the spectra is shown, which contains two signals that correspond to two residues with a completely opposite behavior. The one to the right is unaffected by the presence of compound YFE02, whereas the the signal to the left displays a chemical shift perturbation pattern where the signal moves upon addition of increasing amounts of compound, up to a certain concentrarion at which the binding pocket is saturated. At that point, the chemical shift perturbation value will not vary any more.

2.4 Surface Plasmon Resonance

A second way of determining the binding constant between a protein and its ligand is by means of Surface Plasmon Resonance [177] (SPR). SPR is a phenomenon which occurs when light is reflected off thin metal films. A fraction of the light energy incident at a sharply defined angle can interact with the delocalised electrons in the metal film (plasmon) thus reducing the reflected light intensity. This phenomenon is the basis for the bioassays presented below.

The experiments require that protein A (ICAM-1) is coated to a gold surface (also called "chip") (Figure 2.10) to which a beam of light is directed. A solution with protein B (the αL I-domain) is flown over the chip. If binding occurs, the local refractive index changes, leading to a change in SPR angle, which can be monitored in real-time by detecting changes in the intensity of the reflected light, collected in a sensogram. The rates of change of the SPR signal can be analyzed to estimate apparent rate constants for the association and dissociation phases of the binding. The ratio of these values gives the apparent equilibrium constant (affinity). The size of the change in SPR signal is directly proportional to the mass being immobilised and can thus be interpreted crudely in terms of the stoichiometry of the interaction. Signals are easily obtained from sub-microgram quantities of material.

An advantage of this technique with respect to CSP experiments is that it is possible to determine the Kd of the binding of two relatively big proteins. By CSP experiments, the interaction between two proteins provokes a broadening of the peaks of the HSQC spectrum, which does not permit to estimate the chemical shift perturbation and therefore, the binding constant. CSP experiments are therefore usefull to localize the binding site and Kd of small molecular weight molecules or peptides, whereas for bigger molecules, SPR is more suitable.

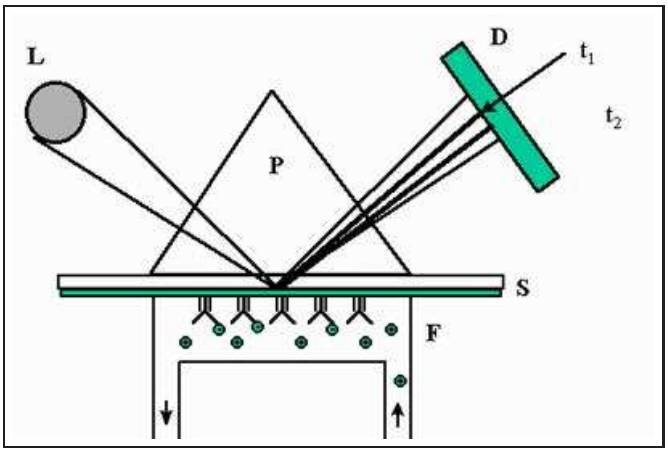


Fig. 2.10: Surface plasmon resonance detection unit. L: light source, D: photodiode array, P: prism, S: sensor surface, F: flow cell. The two dark lines in the reflected beam projected on to the detector symbolize the light intensity drop following the resonance phenomenon at time = t1 and t2. The line projected at t1 corresponds to the situation before binding of ligands to the protein on the surface and t2 is the position of resonance after binding.

2.5 Experimental Protocols

2.5.1 Cloning, Expression, and Purification of the I-domain

The sequence of the wild type LFA-1 αL I-domain (residues 128-307) was cloned into the pET 14d vector [3]. Alternatively, the active I-domain sequence was produced by introducing two mutations, K287C and K294C, by site-directed mutagenesis, as described elsewhere [60]. Both constructs were expressed at 37 C in the insoluble fraction of BL21(DE3) cells [4]. Purification of the wild type and active mutant forms of the LFA-1 αL I-domain from the insoluble fraction was performed as described in Huth and coworkers [178] except that a single column (a size-exclusion Superdex75) was utilized after the refolding step. Single (^{15}N) and doubly (^{13}C^{15}N) uniformly labeled forms of the wild type and active mutant αL I-domains were produced by expression in minimal medium E.coli [5] containing ^{15}N NH_4Cl and both ^{15}N NH_4Cl and ^{13}C D-glucose, respectively.

In Figure 2.11 several sodium dodecyl sulfate (SDS hereafter) electrophoresis gels are shown[6], which summarize the results of the different steps during the purification protocol of the αL I-domain (Molecular Weight = 20kDa). In Figure 2.11 A, it is observed that upon induction of protein synthesis with isopropyl--D-thio-galactsido (IPTG hereafter) [7] (lane 3), cells were enriched with our construct. The uninduced cells do not show a prevalence of I-domain in its protein content (lane 2). Induced cells were broken by sonication[8] and, as observed in Figure 2.11 B, most of the protein was found in the insoluble fraction or pellet (lane 4). The pellet, was washed four times (Figure 2.11, C), resolubilized and refolded ("before"), and concentrated ("after") (Figure 2.11, C, lane 7). The αL I-domain was found to be significantly enriched, such that it was purified using a single size-exclusion chromatography step with a Superdex 75, as observed in Figure 2.12. The chromatogram in Figure 2.12 shows that, as observed in the SDS-gel attached, the first peak corresponds to the I-domain (20 KDa), whereas the second peak corresponds to other proteins or impurities. To verify that the protein purified was homogeneous and contained the expected sequences, the sizes of the constructs were confirmed by electrospray ionization mass spectrometry (data not shown). Equivalent results were obtained for the mutant αL I-domain (data not shown).

[3] A vector is a "vehicle" for delivering genetic material such as DNA to a cell.

[4] *Escherichia Coli* are bacterial cells commonly used for protein expression and purification.

[5] A minimal medium solution contains only those nutrients that a cell needs to fulfill its basic activities: grow and multiply.

[6] SDS-Gel electrophoresis is the most commonly used method for determining the size and amount of proteins in a sample

[7] The generally low expression level of a protein is overcome by the addition of IPTG to the cell culture, which binds to the DNA promoting a dramatic increase on the production of the protein of interest.

[8] Sonication consists of applying ultrasound energy to biological samples in order to disrupt cell membranes and release cellular contents.

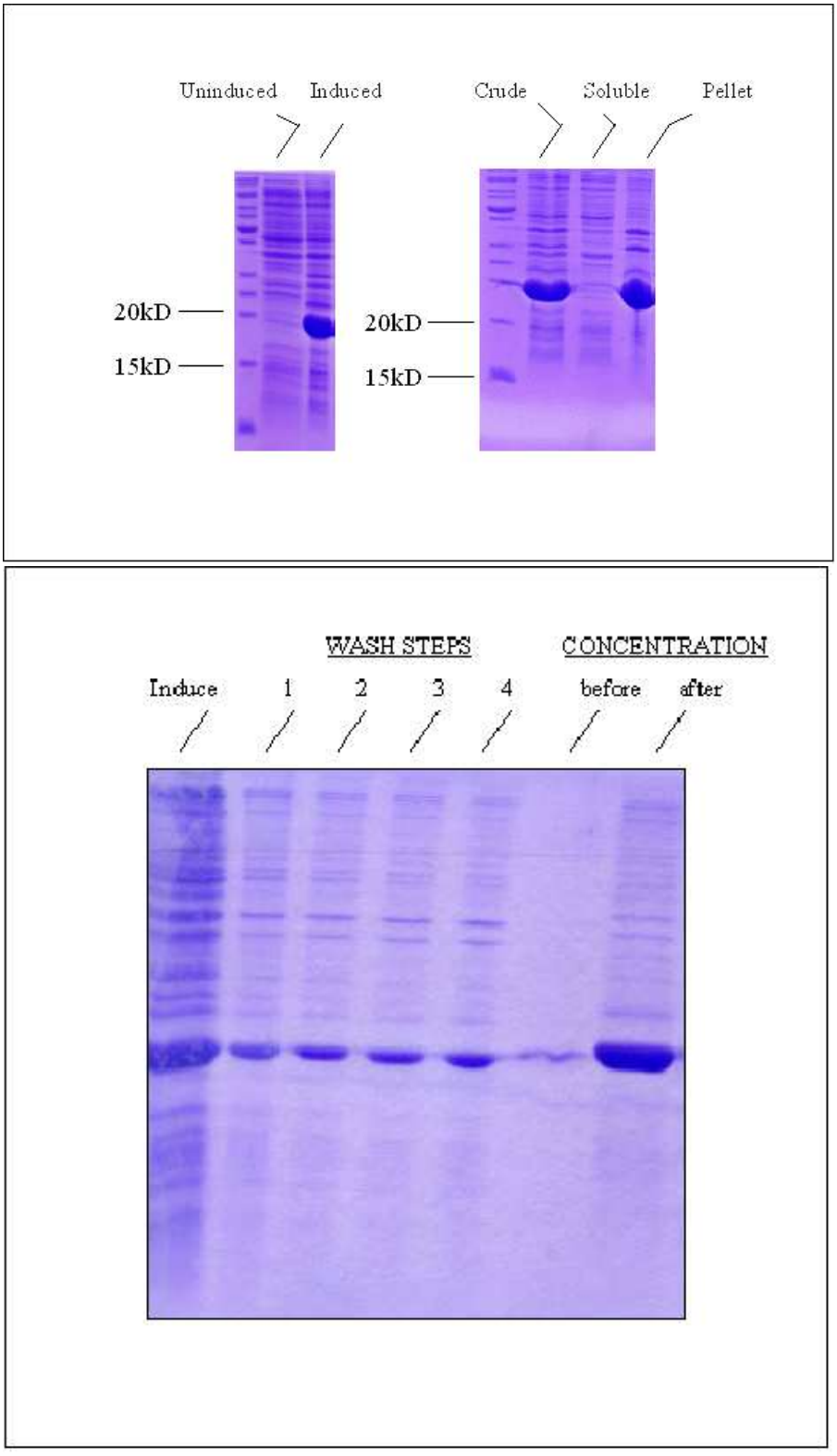


Fig. 2.11: SDS-Gel Electrophoresis. Expression (Top) and purification (Bottom) of the $^{15}N^{13}C$ labeled wild type I-domain. Top) Lane 1: Protein Markers; Lane 2: Non-Induced cell extract; Lane 3: Induced cell extract; Bottom-Left) Lane 1: Protein Marker; Lane 2: Crude ; Lane 3: Soluble Fraction; Lane 4: Insoluble Fraction; Bottom-Right; Lane 1: Induced extract; Lane 2: 1st washing step; Lane 3: 2nd washing step; Lane 4: 3rd washing step; Lane 5: 4th washing step; Lane 6: Refolding; Lane 7: Concentration

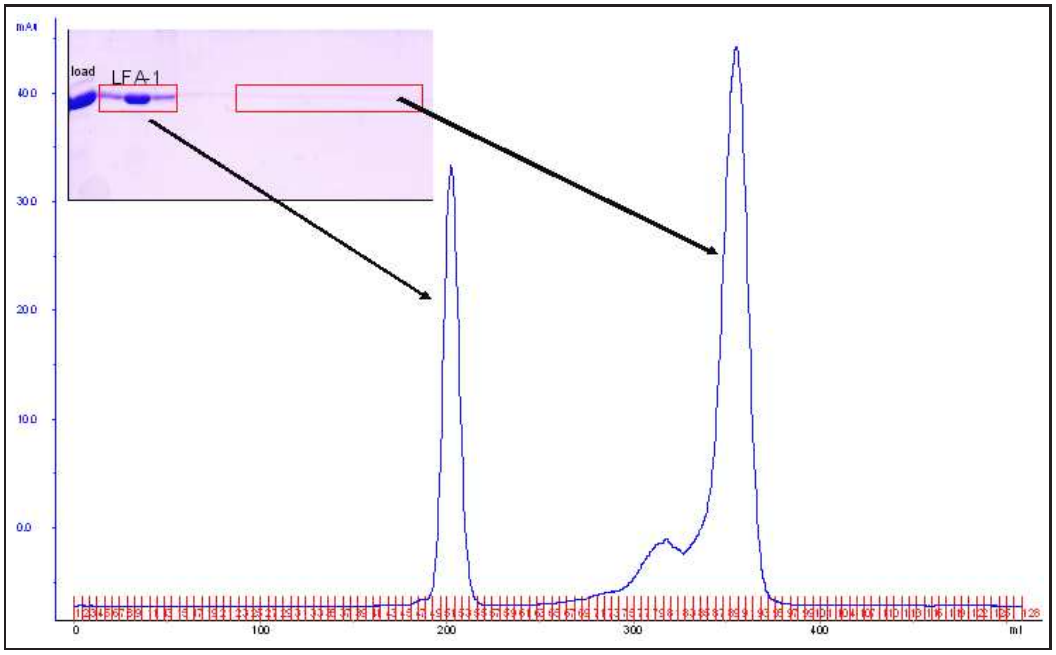


Fig. 2.12: Sensogram obtained upon purification of the wild type αL I-domain using the size exclusion chromatography column Superdex 75. Two peaks were mainly observed. The first one appearing corresponds to our protein (see the SDS-GEL in Figure 2.11); the second peak was discarded since the SDS-Gel showed that it consisted of other protein impurities (not shown).

2.5.2 Acquisition and Assignment of NMR Spectra

All NMR spectra were recorded at 22C on a 600 MHz Bruker Advance Spectrometer with a cryoprobe. Spectra used for backbone assignment of the wild type I-domain were acquired using a 460-700 μM samples in NMR sample buffer (10mM $NaPO_4$, 150mM NaCL, 1mM $MgCl_2$, 0.01% NaN_3, 33 μM DSS, pH 7.4). A 2D $^1H^{15}N$ HSQC, ^{15}N edited 3D HNHA, TOCSY, and NOESY were recorded with a uniformly labeled ^{15}N sample, and a 2D $^1H^{13}C$, 3D CBCA(CO)NH, HNCACB, HNCO, and CC(CO)NH were acquired utilizing a $^{15}N/^{13}C$ labeled protein. All spectra were processed with the NMRPipe software package [179] and analyzed with the NMRviewJ [180] programme. The assignment was performed combining both $^1H^1H$ correlations and heteronuclear $^1H^{13}C$ and $^1H^{15}N$ correlations.

2.5.3 Determination of Compound Solubility

Compound solubility was measured in order to determine the maximum concentration of compound that it would be used in the NMR experiments (Figure 2.13). A 500 μM sample of a previously calibrated DSS solution was used to find out the real concentration of a compound sample with an apparent concentration of 500 μM. A 1D spectrum was acquired of the compound YFE06 in NMR buffer with 500 μM DSS; comparing the DSS signal with that of a well resolved YFEO6 signal its solubility was found to be 200 μM. This maximum was a applied to the other members of the compound family due to their structural similarities and was the concentration used for each compound in subsequent assays.

2.5.4 Chemical Shift Perturbation Experiments

Protein-compound solutions of each potential inhibitor with the ^{15}N labeled wild type or active mutant I-domain were prepared at a ratio of 50 μM protein : 200 μM compound in NMR sample buffer + 0.13 % Dimethyl sulfoxide (DMSO hereafter). A compound-free reference was also prepared in the above solution. A 4hr ^{15}N-HSQC spectra was recorded with each solution. Spectra were analyzed using NMRview and Sparky [181] software packages.

2.5.5 Surface Plasmon Resonance Experiments

The high-affinity mutant or the wild-type I domain was perfused into a CM-5 chip bearing immobilized ICAM-1 or BSA in Tris-buffered saline containing 2mM Mnl_2 at a flow rate of 10uL/min at 25C as described [60]. Both type of I domains were preincubated with 100 μM of the antagonists, or 0.55 μM or 1.1 μM of soluble ICAM-1 in the same buffer for 30 min at room temperature and perfused onto the chip of a BIAcore 2000 instrument. The background response on the BSA surface was substracted from each sensogram.

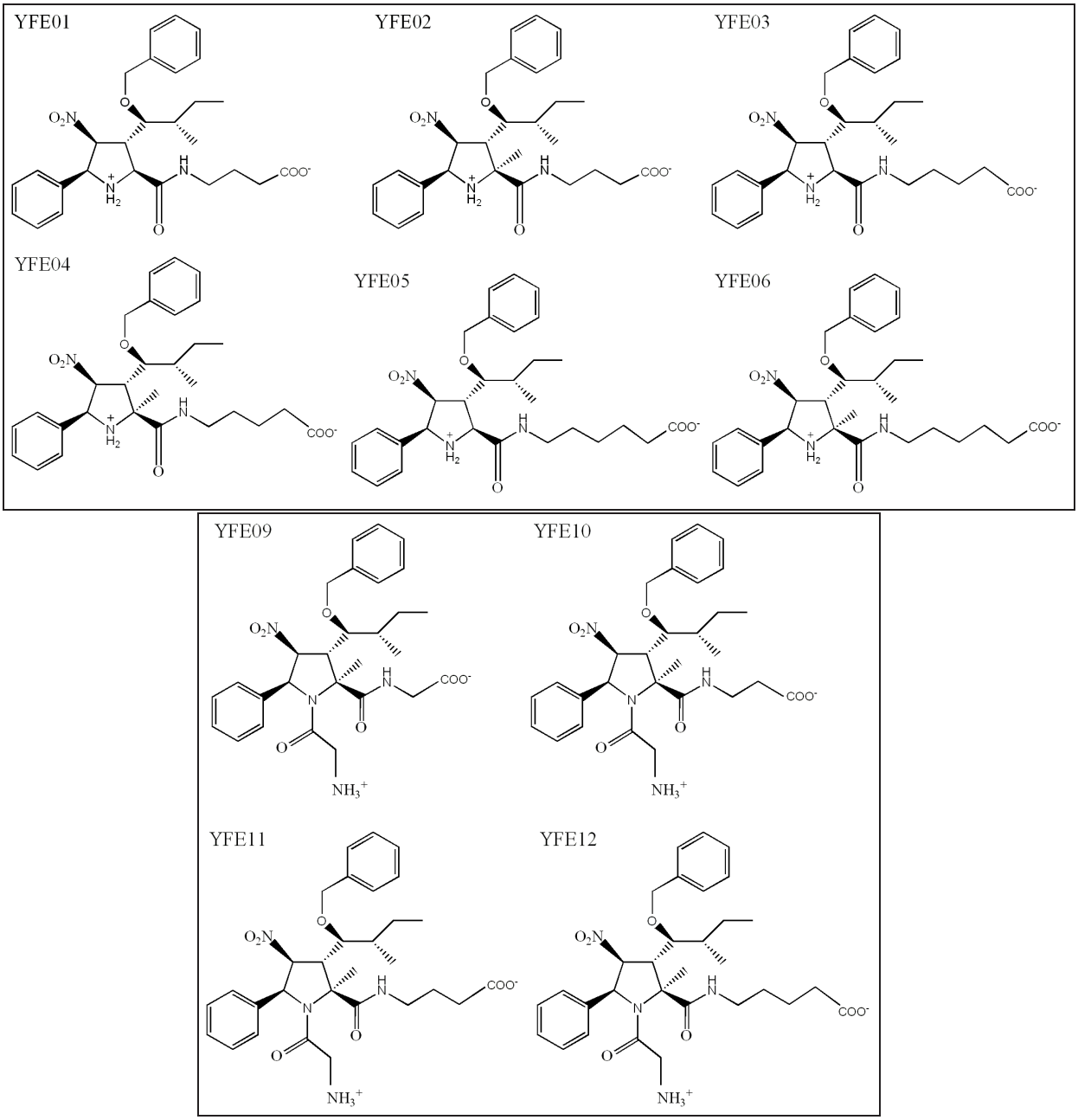


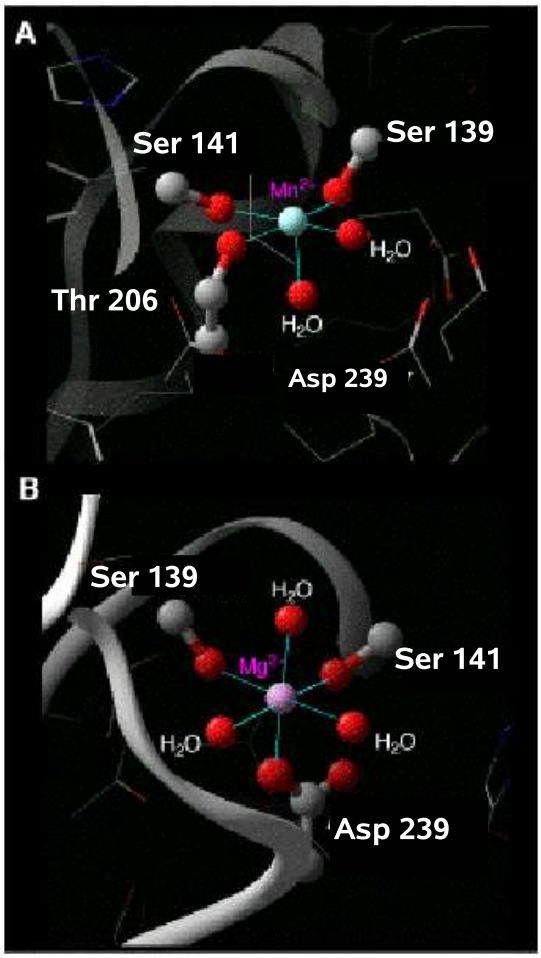
Fig. 2.13: First (top) and second (bottom) family of LFA-1 antagonists designed in this thesis project and used for the present study.

Chapter 3

Quantum Mechanics: A DFT study on the MIDAS motif

The results in this chapter were published in :
E. San Sebastian, J.M. Mercero, R. H. Stote, A. Dejaegere, F. P. Cossio and Xabier Lopez
Journal of the American Chemical Society 2006, vol 128 (11), 3554-3563

Contents


A
Ser 141
Ser 139
Mn2+
H2O
Thr 206
H2O
Asp 239
B
Ser 139
H2O
Mg2+
Ser 141
H2O
H2O
Asp 239

3.1 Introduction

Protein-protein interactions determine the course of many biologically important processes. In particular, interactions between proteins of the Integrin family and Cell Adhesion Molecules (CAMs) are central to a variety of phenomena, including embryonic cell migration, wound healing, immune responses and cancer metastasis [182, 183, 184]. Lymphocyte function-associated antigen-1 (LFA-1,αLβ2) is an integrin mainly confined to leukocytes, which binds to the Intercellular Adhesion Molecule-1 (ICAM-1) located on neighboring endothelial cells via a particular fragment called the Inserted domain (I-domain) (Figure 3.1A), playing a key role in inflammatory diseases and cancer metastasis [185].

Key aspects of the affinity and specificity of this interaction have been addressed by combining the available atomic-resolution structural information for the I-domain of LFA-1 and ICAM-1 with a broad range of biochemical data [186, 187, 60, 78]. A Mg^{2+} ion at the 'top' of the I-domain coordinates to E34 in ICAM-1; this Mg^{2+} ion and the residues to which it is coordinated constitute the metal ion-dependent adhesion site (MIDAS) [41], which has been chosen for study in this work.

From crystallographic analysis, it is known that the MIDAS motif can adopt two alternative conformations, referred to as *open* and *closed*. In both conformations the metal coordinates octahedrically to six polar residues, following the coordination pattern observed for this cation in proteins [188]. In the closed conformation, the metal ion binds to one aspartate (D239) and two serines (S139,S141); three water molecules (Figure 3.1B) complete the coordination sphere. In the *open* MIDAS conformation, the metal ion binds two serines (S139, S141), one threonine (T206) and two water molecules. In the crystallographic environment, a Glu residue from a neighboring I domain in the crystal lattice [41] or a Glu belonging to ICAM-1 [60] (Figure 3.2A) also binds to the MIDAS.

This suggests that when ICAM-1 binds to the I-domain (Figure 3.1A), one of the water molecules of the MIDAS is displaced by E34 in ICAM-1 (Figure 3.1B), and the latter becomes part of the metal ion coordination sphere [60]. The binding of a Glu residue either from ICAM-1 or from a neighboring I-domain appears to be correlated with a change in conformation from the *closed* to the *open* form. These data suggest that the *closed* MIDAS conformation is present in non-activated LFA-1, with low affinity values toward ICAM-1. The *open* MIDAS conformation would be present in activated integrins, which show a higher affinity for the ligand. Thus, this change in the coordination of the MIDAS is believed to be one of the mechanisms that modulate affinity of LFA-1 toward ICAM-1 [189].

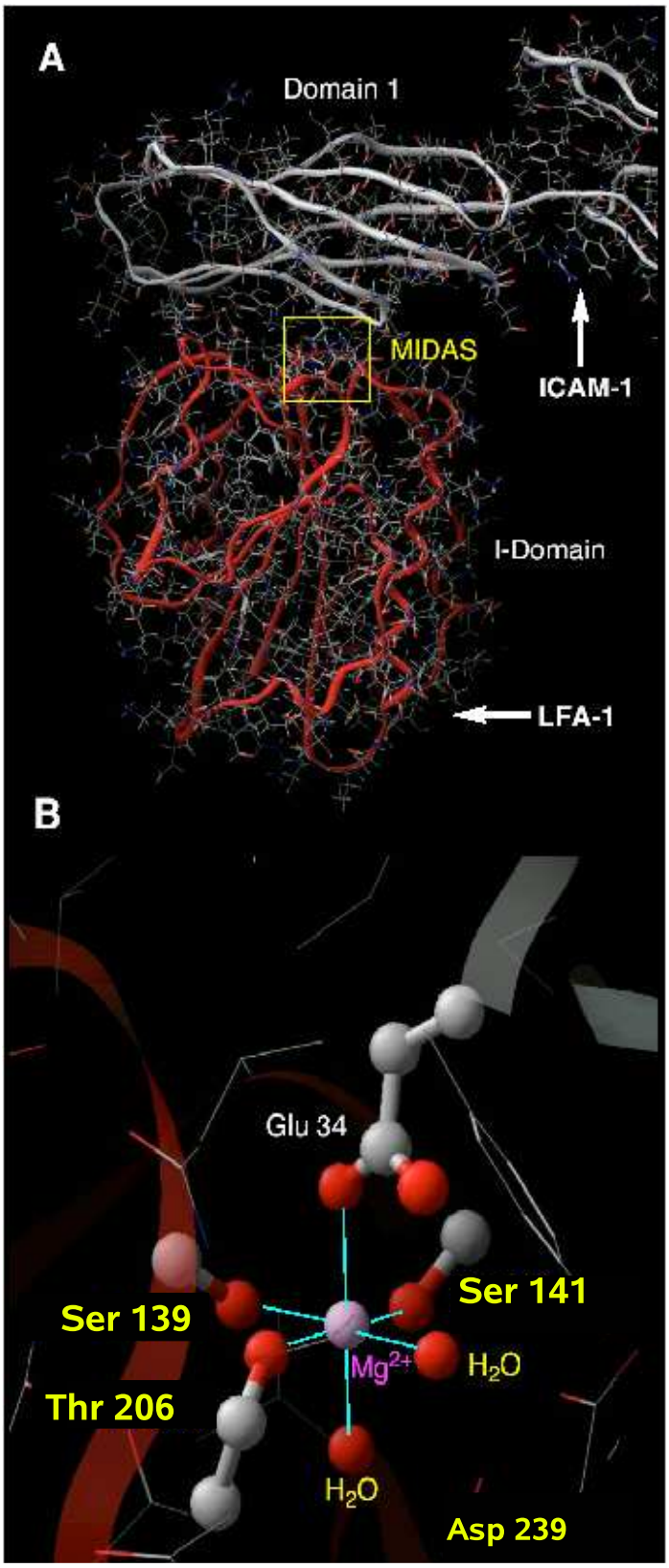


Fig. 3.1: A) Ribbon representation of the X-ray structure of the medium affinity conformation of the I-domain in LFA-1 (red), interacting with ICAM-1 (white) (PDB ID:1MQ8 [60]). B) Highlighted the MIDAS motif (*open* state) in a ball and stick representation, were GLU-34 of ICAM-1 completes the coordination sphere of the metal

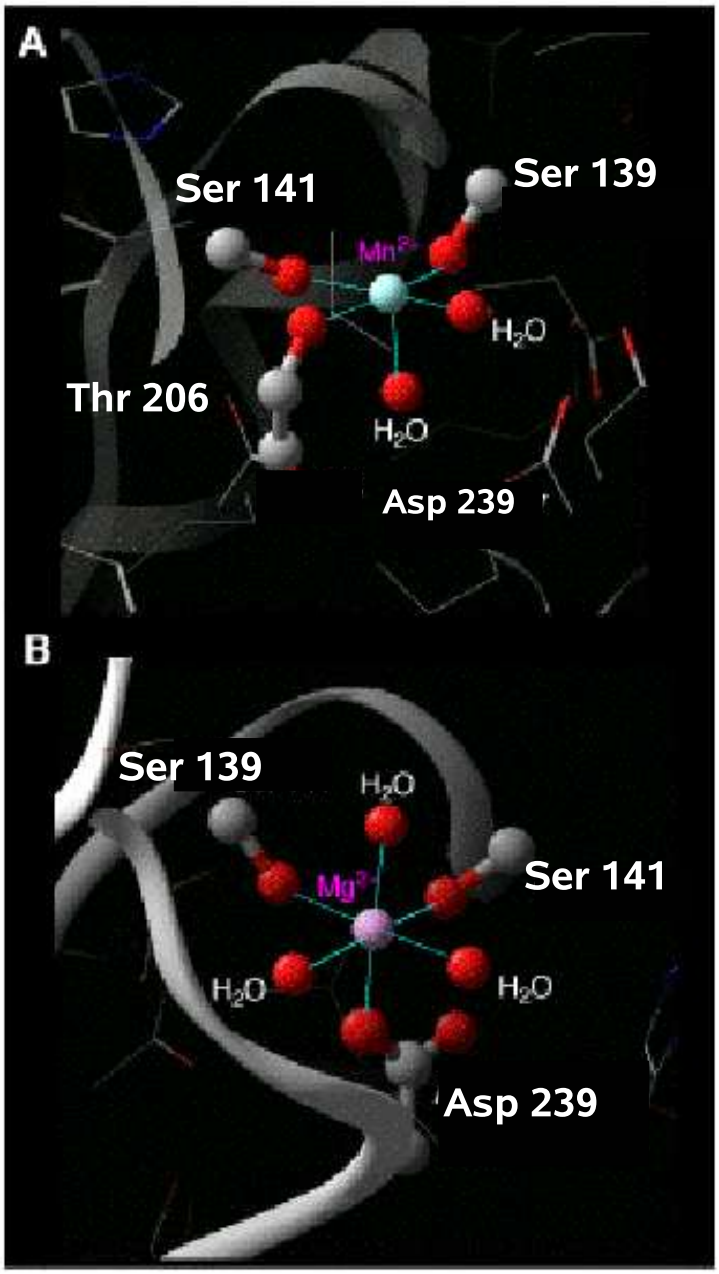


Fig. 3.2: A) Ball and stick representation of the *open* conformation of the MIDAS motif (from PDB ID:1MQ9[60]). The Mn^{2+} cation was replaced by Mg^{2+}. The sixth ligand, which was a Glu from a neighboring I-domain in the crystal lattice was replaced by a water molecule. B) Ball and stick representation of the *closed* conformation of the MIDAS motif (from PDB ID:1MJN [60]).

Besides the role of conformational changes of the MIDAS in affinity modulation, it has also been established that integrin/ligand interactions depend strongly on nature of divalent cations such as Mg^{2+} [63, 45, 61, 187]. Changes in integrin/ligand affinity have been observed as a function of other cations, such as Zn^{+2}, Ca^{+2} and Mn^{+2}, but their binding site(s) and possible role(s) have not been unequivocably established. X-ray structures have confirmed the presence of a Mg^{2+} cation in the *closed* MIDAS motif [45], and a Mg^{2+} or a Mn^{2+} cation in the *open* MIDAS motif [44, 60]. The role of Ca^{2+} cation on integrin activity, with controversial experimental results, has been widely discussed to date. Biochemical assays with isolated I-domains, have suggested that Ca^{2+} is likely to bind differently in closely related I-domains, such as those of LFA-1 and Mac-1; these two I-domains have equivalent MIDAS motifs [64]. On the other hand, the effects of Ca^{2+} on the ligand binding have also been shown to vary as a function of concentration [61, 62, 63, 45, 65]. In addition, at least a high and a low affinity Ca^{2+} binding motifs have been characterized so far, but not unequivocally localized, in intact LFA-1 [61, 64]. All these features of integrins, make it difficult to unequivocally assign a role to Ca^{2+}. On the other hand, to the best of our knowledge, little is known about the effect of Zn^{2+} cation on LFA-1/ICAM-1 binding. According to Griggs and coworkers [64], Zn^{2+} can bind to isolated I-domains of LFA-1 and Mac-1 , which have a single and identical MIDAS motif, whereas Novick et al. [66] assigned an anti-inflammatory role to Zn^{2+}, and proposed that this effect was due to a direct blocking of the interaction between LFA-1 and ICAM-1, but did not establish a direct coordination of Zn^{2+} metal to the MIDAS motif.

The ambiguous interpretation concerning the role of metal ions in integrin/ligand interactions is, in part, due to the difficulty in isolating the direct effect of metal coordination on ligand binding from many other factors [190], such as conformational effects, existence of several metal binding sites or different experimental conditions. In the present work, we used state of the art quantum calculations to establish the intrinsic preference of the metal binding site (MIDAS) in its *closed* and *open* conformations for the cations Mg^{2+}, Mn^{2+}, Zn^{2+} and Ca^{2+} and to quantify the differential ability of these metals to promote ligand binding on the *open* MIDAS motif.

The administration of small inhibitors that bind specifically and effectively to the I-domain, thus blocking the binding of LFA-1 to ICAM-1, constitutes an efficient tool for fighting against the pathologies mentioned earlier. In the interest of designing of novel families of antimetastatic and anti-inflammatory drugs [191], an understanding of the binding modes and affinity regulation at a molecular level is needed. With this purpose in mind, new ligand types were tested for their binding affinities to the MIDAS using quantum mechanical calculations. These ligands could eventually serve as scaffolds for new inhibitory compounds, that could complement those inhibitors already developed based mostly on the geometric and electronic features of the active site of ICAM-1 [192]. In addition, previous studies [188] suggest that local hydrophobic environments can substantially modify the protein-protein interaction equilibria. With this in mind, we analyzed the effect of different dielectric permittivity values on the binding of the ligands to the MIDAS complexes, results which should be applied to the design of inhibitory drugs.

In this work, then, five different aspects of the LFA-1/ICAM-1 interaction were studied by Density Functional Theory (DFT) calculations: i) variation of the affinity energies of the Mg^{2+}-bound MIDAS towards an acetate, as a consequence of changes in the coordination of the metal (*closed* vs *open*), ii) competition/interchange of cations of diverse nature in both the *open* and *closed* MIDAS, iii) variation of the affinity of the *open* MIDAS motif toward the incoming ligand (an acetate molecule), as a function of the metal ion, (Mg^{2+}, Ca^{2+}, Zn^{2+} and Mn^{2+}), iv)

influence of the nature of the incoming ligand on its binding affinity to the Mg^{2+}-bound *open* MIDAS motif and v) effect of the dielectric permittivity of the environment on the binding energies.

3.2 Methods

3.2.1 Preparation of the structures

Initial geometries for the different MIDAS complexes studied in this work were obtained from molecular dynamics (MD) simulations of the following X-ray structures (PDB ID are given) available in the Protein Data Bank (www.rcsb.org/pdb/) [193]: 1MJN for the water-bound *closed* MIDAS, (Figures 3.3 and 3.4, top, left.), 1MQ9 for the *open* MIDAS in the ligand-free I-domain, (Figures 3.3 and 3.4, bottom, left) and 1MQ8 for the *open* MIDAS in the ligand bound I-domain, (Figures 3.3 and 3.4, bottom, right). As there is no entry to date for a crystal structure of a *closed* MIDAS bound to ICAM-1, 1MJN and 1MQ9 were combined to obtain the remaining starting geometries (Figures 3.3 and 3.4, top, right). In 1MJN and 1MQ8 the metal cation in the MIDAS motif is Mg^{2+}, whereas that in 1MQ9 is Mn^{2+}. In the latter we replaced Mn^{2+} by Mg^{2+} prior to running the MD simulations. Molecular dynamics simulations were run by using the CHARMM program [134] with the all atom protein force field [146].

The hydrogen atoms were placed using the HBUILD facility of the CHARMM program [194] and the system was energy minimized by 1000 steps using the Steepest Descent algorithm. The system was solvated with explicit waters, heated to 300K and equilibrated for 25 picoseconds. From the equilibrated structures, first and second shell residues around the Mg^{2+} were extracted for subsequent quantum mechanical calculations. i.e., amino acids and water molecules directly or indirectly coordinated to the metal were identified and isolated from the rest of the protein in the structure files. The details of the molecular dynamics simulations will be presented in more detail in a later publication. From these geometries, two sets of MIDAS model structures were built. In both sets, the minimal number of residues required to adequately reproduce the local features of the environment were used and the remaining valences were saturated with hydrogen atoms, i.e, acetate was used instead of glutamate and aspartate, and methanol was used instead of serine and threonine. In the first set of structures, the MIDAS is represented by the six residues octahedrically coordinated to the cation –which complete the first coordination shell of the Mg^{2+}. The resultant structures for the *closed* and *open* MIDAS conformations are named $1S_{cl}$ (Figure 3.3, top) and $1S_{op}$ (Figure 3.3, bottom), respectively, where the subscripts refer to the *closed* and *open* conformations, and 1S stands for *first shell*.

In the second set of structures, key second shell ligands (amino acids and water molecules) were also included in order to build a more realistic model of the MIDAS as well as to take into better account the environmental changes that occur when passing from the *closed* to the *open* MIDAS conformation. For example, D239 is present in the $1S_{cl}$ structure, but absent in $1S_{op}$, when, in fact, it is displaced to the second shell of the metal in the *open* MIDAS conformation; a water molecule is also present in the second shell of the metal ion and thus absent in the 1S model. Similarly, Thr-206, present in $1S_{op}$ and absent in $1S_{cl}$ is also in the second coordination shell when the MIDAS is *closed*. Therefore, in the second set of structures, these residues were also included. These structures are denoted in this paper by $2S_{cl}$ (Figure 3.4, top) and $2S_{op}$ (Figure 3.4, bottom), respectively, where 2S stands for *second Shell*.

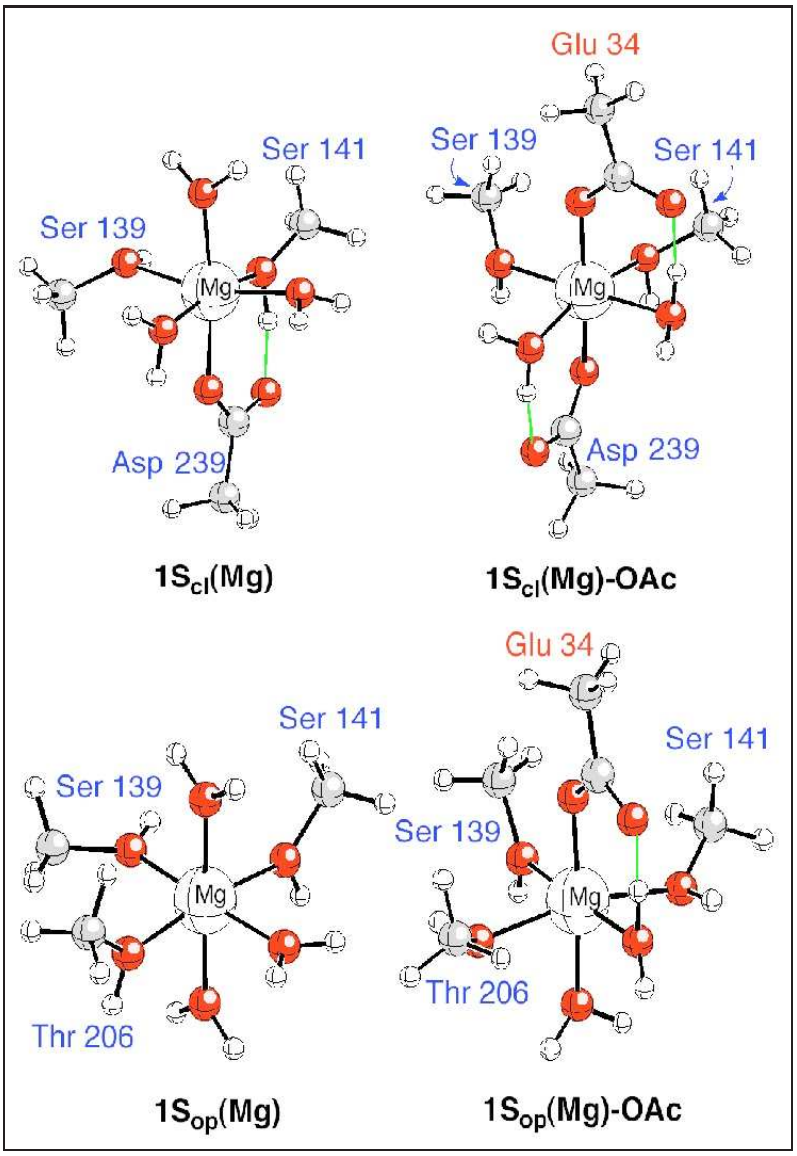


Fig. 3.3: Ball and stick representation of the MIDAS motif with first shell residues. Top: the *closed* conformation in the absence (left) or presence (right) of GLU-34 from ICAM-1. Bottom: *open* conformation in the absence (left) or presence (right) of GLU-34 from ICAM-1.

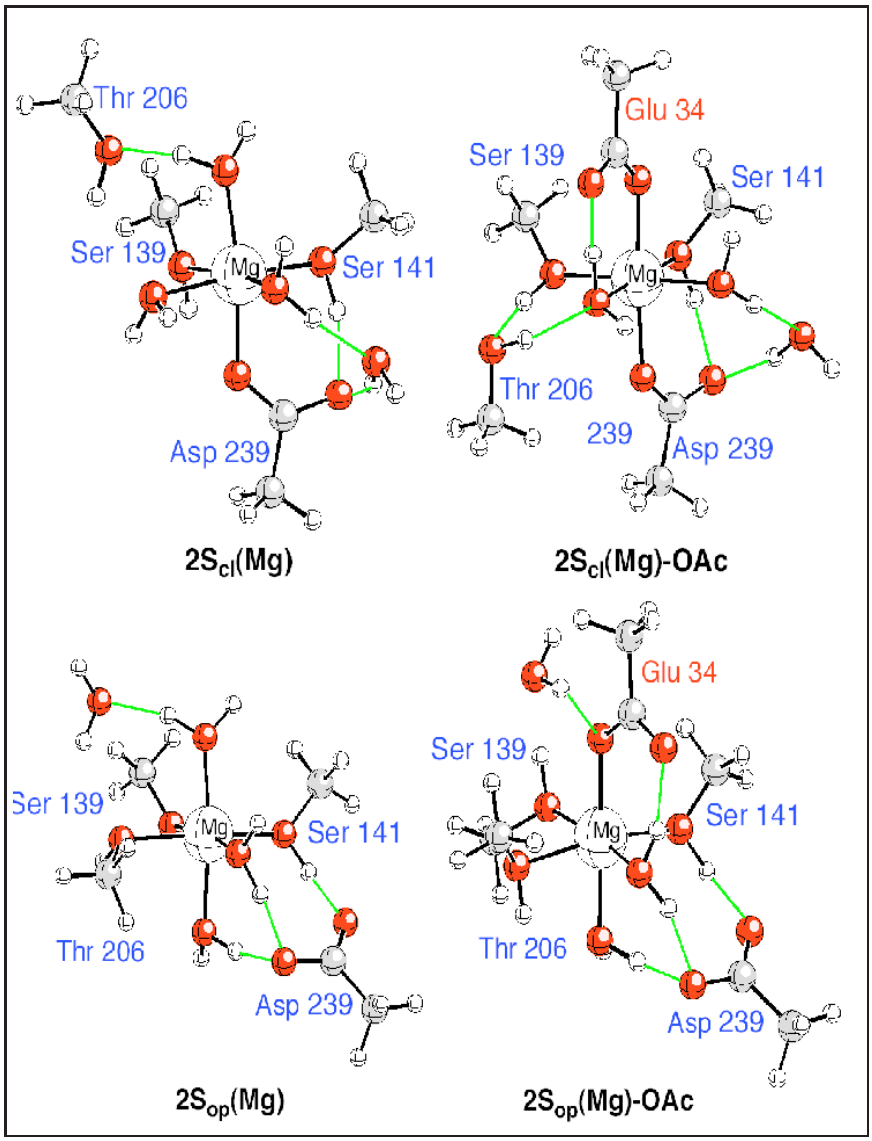


Fig. 3.4: Ball and stick representation of the MIDAS motif with first and second shell residues. Top: *closed* conformation in the absence (left) or presence (right) of GLU-34 from ICAM-1. Bottom: *open* conformation in the absence (left) or presence (right) of GLU-34 from ICAM-1.

3.2.2 Gas-phase results

· To estimate the change in Gibbs free energy, ΔG^s_{aff}, in the following type of equilibrium is considered:

$$(MIDAS)M - W + L \overset{\Delta G^s_{aff}}{\rightarrow} (MIDAS)M - L + W \tag{3.1}$$

$$\begin{array}{r} (MIDAS)Mg - W + M - 6W \\ \downarrow (\Delta G^s_{ex}) \\ (MIDAS)M - W + Mg - 6W \end{array} \tag{3.2}$$

where M stands for metal, L for ligand, W for water, ΔG^s_{aff} for the change in the Gibbs free energy of ligand binding in different solvents and ΔG^s_{ex}, for the change in Gibbs free energy for the cation exchange reaction in different solvents. These two equilibria summarize the reactions studied in this paper and they will be discussed individually later in the text. Here, we note that the (MIDAS)M-W and (MIDAS)M-L terms in equations 1 and 2, represent, respectively, water and ligand bound to the MIDAS motif and where M is one of the following metal ions: Mg^{2+}, Zn^{2+}, Ca^{2+} or Mn^{2+}. The MIDAS motif is represented by the following models: $1S_{cl}$, $1S_{op}$, $2S_{cl}$ or $2S_{op}$. M-6W represents the hexahydrated metals (Ca^{2+}, Zn^{2+} or Mn^{2+}).

Density functional theory (DFT) methods (for a review on quantum methods see [195]) were applied in conjunction with the Becke hybrid [196](B3) exchange functional and the Lee-Yang-Parr (LYP) [197] correlation functional to study and characterize the above mentioned metal-bound complexes, as there is wide agreement as to the adequacy of these method to the study of this type of metal-ligand complex [198, 199]. Full geometry optimization is carried out using the standard all-electron 6-31+G* basis set for the metal and effective core potentials and shared-exponent basis set of Stevens, Krauss, Basch and Jasien (SKBJ) [114] including a polarization and a diffuse function for the ligands. This pseudopotentials/all-electron basis set combination was first found by N. Gresh et al. [200, 201, 202] to represent a well-balanced compromise between accuracy and computational efficiency,and shall be denoted SKBJ+* hereafter. Vibrational frequencies are calculated at this level of theory to determine whether the obtained geometries correspond to a minimum on the potential energy surface (all frequencies real). The energies are then refined by B3LYP/6-311++G(2df,2p) single-point calculations performed on the B3LYP/SKBJ+* geometries.

The changes in the Gibbs free energies of the reactions are calculated according to equation 3.3, where G, H, S and T are the Gibbs free energy, enthalpy, entropy and temperature (T=298 K), respectively. The enthalpy can also be written as in equation 3.4:

$$\begin{aligned} \Delta G &= \Delta H - T\Delta S & (3.3) \\ H &= E + RT = E_0 + E_{vib} + E_{rot} + E_{trans} + RT \\ &= E_e + ZPVE + E_{vib} + E_{rot} + E_{trans} + RT & (3.4) \end{aligned}$$

where E_e, E_{vib}, E_{rot}, E_{trans} and ZPVE are the electronic, thermal vibrational correction, rotational and translational energy components and zero-point vibrational energies, respectively. Zero-point vibrational energies, thermal corrections to the energy (E_{vib}, E_{rot} and E_{trans}) and entropies were calculated at the B3LYP/SKBJ+* level of theory, using standard statistical mechanical methods [203]. This method has been widely used in our group and has shown to be adequate for this type of calculations [204, 205, 206, 207, 208].

3.2.3 Solvation contributions

The environment is known to exert a substantial influence on the metal-ligand interaction energies [188]. Therefore, the effect of the surroundings of the MIDAS motif on its interaction with the ligand, was evaluated doing single point PCM calculations on the gas phase geometries. The solvents used were the following: CCl_4 (ϵ=2.2), ether (ϵ=4.3), acetone (ϵ=20.7) and water (ϵ=78.4), which cover the dielectric environments ranging from completely buried regions in the protein to completely solvent exposed ones [209, 210, 211]. To estimate the affinity energies in solution, the solvation free energy values (for each structure in each one of the solvents) were calculated and combined with the gas-phase affinity energies as shown in Figure 3.5. All calculations were carried out with the GAUSSIAN98 [212] package.

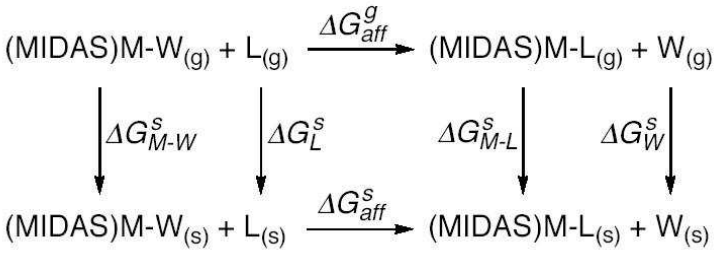


Fig. 3.5: **Scheme 1.** Thermodynamic cycle used to compute the free energy of binding of a ligand to a water-bound MIDAS motif. ΔG^s_{M-W}, ΔG^s_{M-L}, ΔG^s_L and ΔG^s_W are the solvation free energies for the water and acetate-bound MIDAS complexes, for the different ligands and for a water molecule, respectively.

3.3 Results and Discussion

3.3.1 Affinity regulated by *closed*-to-*open* conformational changes in the MIDAS motif.

Affinity free energies of Mg^{2+}-bound $1S_{cl}$, $1S_{op}$, $2S_{cl}$ and $2S_{op}$ models for an acetate molecule were calculated according to equation 3.3. In Table 3.1, the gas-phase results obtained for each model, ΔG^{g}_{aff} (at 298K), are decomposed into its various components: electronic energy (ΔE_e), electronic energy corrected by zero-point vibrational energies (ΔE_0) and enthalpy (ΔH_{298}) and entropy ($T^*\Delta S_{298}$) differences. In Table 3.2, solution phase results (ΔG^{g}_{aff}) are shown for the *closed* and *open* models, together with their relative affinity values, $\Delta\Delta G^{s}_{aff}(1S_{op}/1S_{cl})$ and $\Delta\Delta G^{s}_{aff}(2S_{op}/2S_{cl})$, which were calculated according to equations 3.5 and 3.6, respectively,

$$\Delta\Delta G^{s}_{aff}(1S_{op}/1S_{cl}) = \Delta G^{s}_{aff}(1S_{op}) - \Delta G^{s}_{aff}(1S_{cl}) \quad (3.5)$$

$$\Delta\Delta G^{s}_{aff}(2S_{op}/2S_{cl}) = \Delta G^{s}_{aff}(2S_{op}) - \Delta G^{s}_{aff}(2S_{cl}) \quad (3.6)$$

where $\Delta G^{s}_{aff}(1S_{op})$, $\Delta G^{s}_{aff}(1S_{cl})$, $\Delta G^{s}_{aff}(2S_{op})$ and $\Delta G^{s}_{aff}(2S_{cl})$ are the affinity energies for reaction 1, for the $1S_{op}$, $1S_{cl}$, $2S_{op}$ and $2S_{cl}$ structures, respectively. This is a measure of how the interaction of the MIDAS motif and an acetate molecule is modulated by the *closed*-to-*open* conformational change. Negative $\Delta\Delta G^{s}_{aff}$ values in Table 3.1 indicate that the *open* MIDAS conformation binds an acetate molecule with higher affinity than the *closed* conformation.

Gas phase results

The gas phase results (table I) for $1S_{cl}$ and $1S_{op}$, -107.2 kcal/mol and -196.6 kcal/mol, respectively, denote a high affinity of both conformations towards an acetate anion in the gas-phase, but much higher affinity of the *open* conformation ($\Delta\Delta G^{s}_{aff}$ = -89.4 kcal/mol; Table 3.1). Analyzing the various contributions to the affinity energies in Table 3.2, we can observe that ΔG^{g}_{aff} values are mainly dominated by the electronic energies. Zero-point vibrational energy corrections and the entropic terms in particular, tend, in general, to slightly decrease the affinity energies. The observed enhanced affinity of $1S_{op}$ towards an acetate molecule is likely due to the nature of the residues directly bound to the Mg^{2+} cation. The axial D239 that is present in $1S_{cl}$ but is absent in $1S_{op}$ makes the total charge of the MIDAS to change from +1 in $1S_{cl}$ to +2 in $1S_{op}$, which favors the interaction of the latter with a negatively charged acetate molecule, and, hence, a decrease (more negative) on ΔG^{g}_{aff} is observed.

Table **3.1**: Gas-Phase affinity energies (kcal/mol) of the Mg^{2+}-bound *open* MIDAS motif toward an acetate decomposed in its various components*, for the four different MIDAS models studied.

Midas Conf.	ΔE_e	ΔE_0	ΔH_{298}	$T * \Delta S_{298}$	ΔG^g_{aff}
$1S_{cl}$	-112.5	-110.4	-111.4	-4.2	-107.2
$1S_{op}$	-198.3	-199.2	-199.2	-2.6	-196.6
$2S_{cl}$	-115.2	-115.1	-115.4	-6.1	-109.3
$2S_{op}$	-119.9	-118.1	-119.1	-3.6	-115.5

* E_e=Total electronic energy; E_0=Zero Point corrected total electronic energy; H_{298}=Enthalpy; S_{298}=Entropy; G_{298}= Gibbs free energy at 298 K.

Analysis of the X-ray structures of LFA-1 shows that the activation of the MIDAS from the *closed* to the *open* state is more accurately described as a displacement of D239 from the first coordination shell to the second one, and its substitution by a water molecule coming from the second coordination shell. Similarly, in the equatorial plane there is an exchange between a water molecule and Thr-206. As was done for the 1S models of the MIDAS motif, the affinity energies for an acetate molecule were calculated for the 2S MIDAS models, according to equation 3.3. Data are also collected in Table 3.1 and Table 3.2.

The effect of the second shell ligands is more pronounced on the *open* MIDAS structures than on the *closed* ones. $2S_{cl}$ shows a moderate increase in the affinity for an acetate molecule (of around 2 kcal/mol) as compared to $1S_{cl}$. On the contrary, $2S_{op}$ shows a reduction in affinity of around 75 kcal/mol when compared to $1S_{op}$. The inclusion of second shell ligands makes the net charge of $2S_{op}$ and $2S_{cl}$ the same (+1), provoking the affinity for an acetate molecule to become similar. However, although similar, the gas-phase affinity energies obtained for the $2S_{op}$ conformation are consistently larger (more negative) than those obtained for the *closed* conformations. The relative affinity free energy of $2S_{op}$ with respect to $2S_{cl}$ is -6.2 kcal/mol (Table 3.2). On the other hand, the trends of the different contributions to the gas-phase free energies is kept qualitatively similar (Table 3.1). Zero-point vibrational energy corrections tend to decrease slightly the affinity energy and entropic contributions also lead to lower affinities. In this case, there is a significant contribution of the entropy to the differential affinity, as ca. 40% of the total relative affinity stems from the differential entropic factors between the $2S_{op}$ and $2S_{cl}$ models.

Table **3.2**: Comparison of Δ G values (in kcal/mol) obtained for the *open* and the *closed* MIDAS conformations, in the ligand binding reaction at different dielectric environments.

	$\Delta G^s_{af}(1S_{cl})$	$\Delta G^s_{af}(1S_{op})$	$\Delta\Delta G^s_{af}(1S_{op}/1S_{cl})$	$\Delta G^s_{af}(2S_{cl})$	$\Delta G^s_{af}(2S_{op})$	$\Delta\Delta G^s_{af}(2S_{op}/2S_{cl})$
Gas phase	-107.2	-196.6	-89.4	-109.3	-115.5	-6.2
CCl_4	-49.3	-97.4	-48.1	-60.7	-65.5	-4.8
Ether	-22.1	-52.2	-30.1	-39.0	-43.1	-4.1
Acetone	-5.9	-23.8	-17.9	-26.9	-30.2	-3.3
Water	8.9	-1.2	-10.1	0.8	-1.3	-2.1

Effect of the dielectric environment

In Table 3.2, the effect of different media on the affinity free energy (ΔG^s_{aff}) is shown. Solvation free energies for each of the structures that characterizes the *closed* and *open* states of the MIDAS, $1S_{op}$, $1S_{cl}$, $2S_{op}$ and $2S_{cl}$ in the acetate-bound or acetate-unbound form, as well as for an acetate and a water molecule individually, were calculated. The results of the calculations show that the solvation free energies were modulated by the total charge of the structure, the volume of the structure, and the dielectric permittivity of the environment in the following way: the larger the charge, the smaller the volume and the more polar the environment, the larger was the solvation free energy of the structure (Table 3.8 in Supplementary Material). These energies were combined with the gas-phase affinity free energies to give the total affinity free energy in solution according to Figure 3.5.

In general, more polar environments led to a lower affinity of the MIDAS towards an acetate (i.e., less negative values in Table 3.2), and to smaller affinity difference between *closed* and *open* structures. In the most extreme case, that of $\epsilon = 78.4$ (water), positive values of the affinity energies were found for the *closed* structures: 8.9 kcal/mol for $1S_{cl}$ and 0.8 kcal/mol for $2S_{cl}$, respectively. These effects are due mainly to the partial charge neutralization upon ligand-binding (leading to a lower solvation free energy) combined with the unfavorable desolvation of the acetate.

The relative affinity for an acetate molecule between the *closed* and the *open* conformations ($\Delta\Delta G^s_{aff}$ in Table 3.2), for 1S and 2S MIDAS models, is also highly affected by the polarity of the media. The more polar the environment, the lower the values of $\Delta\Delta G^s_{aff}$. That is, $\Delta\Delta G^s_{aff}$ is reduced (in absolute value) from -89.4 kcal/mol in the gas-phase to -10.1 in water for the 1S MIDAS models, and from -6.2 kcal/mol to -2.1 kcal/mol for the 2S models. It is also interesting to note that as the solvent polarity increases there is a reduced influence of the second coordination shell residues on the *open*/*closed* MIDAS relative affinity energies. For instance, in low-dielectric environments (CCl_4) there is a substantial difference between $\Delta\Delta G^s_{aff}(1S_{op}/1S_{cl})$ and $\Delta\Delta G^s_{aff}(2S_{op}/2S_{cl})$, -48.1 kcal/mol vs -4.8 kcal/mol, whereas in water these differences are reduced (-10.1 kcal/mol vs -2.1 kcal/mol). Nevertheless, regardless the model used for the MIDAS, we find that low dielectric media favor affinity of the MIDAS towards the glutamate-like ligand and also favor the affinity of the *open* states with respect to the *closed* ones. These results quantitatively explain the fact that the MIDAS site in the I-domain of LFA-1 is surrounded by a hydrophobic environment that ensures a low dielectric permittivity and hence more efficient coordination of the metal with the carboxylate of the ligand [78]. Moreover, the large effect exerted by second shell ligands or apolar environments around the MIDAS, have important implications in drug design. These results show that fundamental characteristics of ligand binding by the MIDAS of the I-domain are governed by the changes in conformation around the metal ion, and the hydrophobic environment provided by surrounding residues.

3.3.2 Replacement of cations in the MIDAS

Is the interchange between Zn^{2+}, Ca^{2+} or Mn^{2+}, and Mg^{2+} in the MIDAS motif, thermodynamically favorable? To address this question, reactions depicted in Figure 3.6 were studied with the MIDAS in the two alternative conformations: $1S_{cl}$ and $1S_{op}$. Concerning the preferred coordination of each of the cations under study, an octahedral geometry has been observed for solvated Ca^{2+}, Mg^{2+}, Zn^{2+} and Mn^{2+} and for protein-bound, Mg^{2+} and Mn^{2+} [213, 214, 60]. The coordination number for Ca^{2+} cation ranges from 6 to 8 in proteins [213], whereas for Zn^{2+} cation, the coordination number depends on its role. In enzymes, where the zinc ion is often bound to Glu and His residues, a distorted square pyramidal coordination is usually observed, whereas it adopts a tetrahedric coordination when it plays an structural role [215, 216] and is often found bound to Cys residues. In some proteins, a coordination number of six has also been observed [216, 214]. We emphasize that in every reaction studied in the present work, the incoming metal retains the coordination geometry of the outgoing metal. Regarding the spin multiplicity on Mn^{2+} octahedral complexes, only the high-spin structures will we discussed [217]. Values for ΔG_{ex} can be found in Table 3.3.

When including the effect of the polarity of the environment, we must take into account that the hexahydrated compounds represent fully solvated cations. Therefore, the solvation free energy for these structures was only evaluated in water-like environments. The MIDAS motif can still be exposed to media with different polarities and therefore various possibilities have been considered in Table 3.3.

Our results indicate an interesting dependence of metal exchange energy on the conformation of the MIDAS and nature of the metal. In the case of Zn^{+2}, negative values of ΔG_{ex} are obtained regardless the conformation of the MIDAS, *closed* or *open*, or of the polarity of the environment. The exchange energies range from -1.1 kcal/mol (water-like permittivity) to -2.2 kcal/mol (gas-phase) for the $1S_{cl}$ structure. For the *open* conformation $1S_{op}$, ΔG_{ex} are negative as well, but they are consistently higher in absolute value than for $1S_{cl}$, ranging from -2.4 kcal/mol (water) to -1.6 (gas-phase). In any case, these energies reveal that displacement of Mg^{+2} by Zn^{+2} would be possible in an octahedric *open* or *closed* MIDAS motif.

Table **3.3**: Δ G_{ex} (kcal/mol) calculated for the cations exchange reactions, prior to ligand binding at different dielectric environments.

	$1S_{cl}$			$1S_{op}$		
	Zn^{+2}	Ca^{+2}	Mn^{+2}	Zn^{+2}	Ca^{+2}	Mn^{+2}
Gas Phase	-2.2	9.1	-8.6	-1.6	1.3	-5.3
CCl_4	-2.7	0.7	-9.8	-4.2	-3.5	-4.5
Ether	-2.2	1.5	-9.0	-3.8	-1.9	-3.3
Acetone	-2.3	2.6	-11.0	-3.9	-0.3	-3.3
Water	-1.1	4.1	-3.2	-2.4	7.3	2.3

On the other hand, and according to our results, exchange possibilities for Ca^{2+} cation strongly depend on the conformation of the MIDAS. When the MIDAS is in its *closed* conformation [218] $1S_{cl}(Mg)$, positive values for the ΔG_{ex} energies are obtained, which range from 4.1 kcal/mol in water to 9.1 kcal/mol in the gas phase. These values indicate that the displacement of Mg^{2+} by Ca^{2+} in the *closed* MIDAS is thermodynamically unfavorable. However, in the case of the *open* MIDAS conformation, $1S_{op}(Mg)$, with the metal surrounded by all formally neutral ligands, there are some permittivity values of the media, such as those modeled by the solvents CCl_4 and Ether, for which a negative value of the exchange energy is found, -3.5 kcal/mol and -1.9 kcal/mol respectively. In addition, for acetone, we also get a negative value, -0.7 kcal/mol, but, since this number is lower than 1 kcal/mol in absolute value, we may consider these reaction as thermoneutral. Taking into account these data, it appears that the *closed* MIDAS conformation inhibits displacement of Mg^{+2} by Ca^{2+}, whereas the probability for Mg^{+2}/Ca^{2+} exchange in the *open* MIDAS motif strongly depends on the surroundings. Interestingly, when the optimized metal-MIDAS geometries were compared (see Supplementary Material), one observes that the geometry of the Ca^{2+} bound MIDAS changed more significantly than the others, suggesting that Ca^{2+} may prefer to coordinate to a higher number of ligands; this, in turn, would result in significant structural rearrangement of the surroundings.

On the contrary, exchange between Mg^{+2} and Mn^{+2} is favorable, regardless the conformation of the MIDAS motif, and the dielectric permittivity of the environment. When the MIDAS is in the *closed* conformation, exchange energies range from -8.6 kcal/mol in gas phase to -3.2 kcal/mol in water. The metal exchange in the *open* MIDAS conformation is also favored for all the permittivity values tested, but for that of water. The ΔG_{ex} values range from -5.3 kcal/mol in vacuum to 2.3 kcal/mol in water. Then, it seems that the replacement of Mg^{+2} by Mn^{+2} is favored for both MIDAS conformations, and for either polar or more hydrophobic environments (except water).

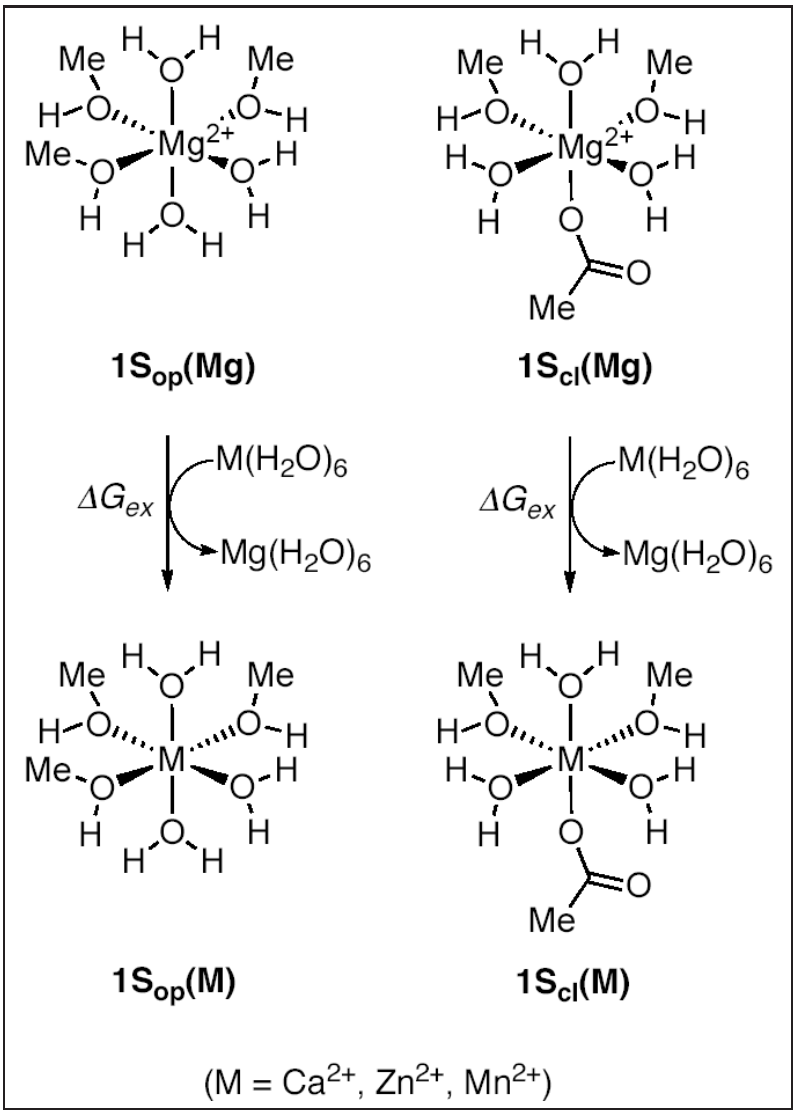


Fig. 3.6: Representation of the reactions of metal exchange for the *closed* (left) and *open* (right) MIDAS motif.

3.3.3 Affinity regulated by the metal cation

The affinity energies towards an acetate ligand were evaluated for the $1S_{op}$ conformation of the MIDAS and with Ca^{2+}, Zn^{2+} and Mn^{2+} as the central metal cation.

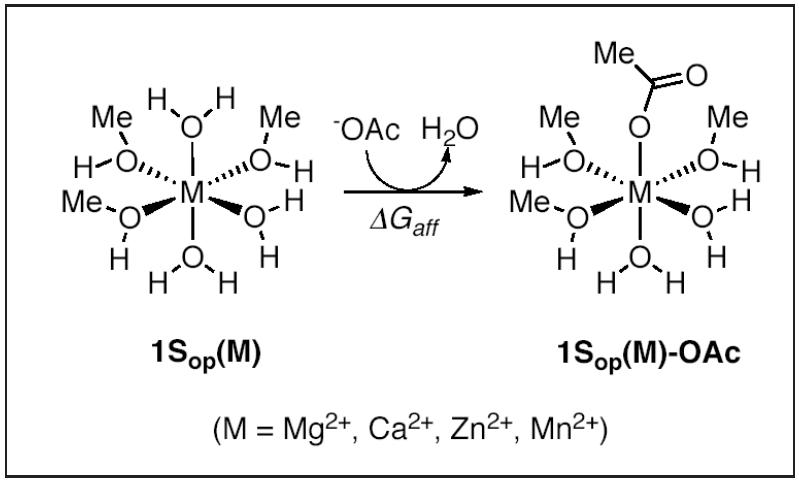


Fig. 3.7: Representation of the ligand binding reaction for the *open* MIDAS motif, with different metals as central cations.

Gas phase results

Gas phase affinity energies, towards an acetate, of the $1S_{op}$ MIDAS structures with different metal cations (Figure 3.7) are collected in Table **3.4**. The ΔG^s_{aff} values obtained with each metal are shown along with the relative $\Delta\Delta G^s_{aff}$ values with respect to Mg^{2+}. The four cations show negative values of ΔG^s_{aff}, i.e., the coordination to an acetate molecule and the displacement of a water molecule is thermodynamically favorable for all metals tested. However, there are sizeable and consistent differences among the affinity results obtained for each cation. The Ca^{2+}-bound *open* MIDAS motif shows lower affinity energies than the Mg^{2+}-bound *open* MIDAS, whereas the presence of Zn^{2+} or Mn^{2+}(IV) in the *open* MIDAS motif, increases the affinity towards an acetate. For the Zn^{2+}-bound *open* MIDAS, the improvement is moderate, whereas Mn^{2+}(IV), provokes a considerable enhancement of the affinity energy, compared to the Mg^{2+}-bound *open* MIDAS.

Effect of the dielectric environment.

Solvent effects are qualitatively similar in all three cases: The more polar the solvent is, the lower individual affinity energies are (data not shown). In general, there is also a decrease in the relative affinities with respect to magnesium with polarity of the media, leading to lower absolute values of $\Delta\Delta G^s_{aff}$, with the only exception of the $\Delta\Delta G^s_{aff}(Zn/Mg)$ in water solvent.

Overall, our results suggest that i) the presence of Zn^{2+} or Mn^{2+} on the MIDAS motif is compatible with an octahedric coordination, whereas ii) the coordination number of the Ca^{2+} cation

may increase in either the *open* or *closed* MIDAS motif; iii) a Ca^{2+} cation bound to the octahedral *open* MIDAS motif would have a decreased affinity for the ligand relative to Mg^{2+}-bound open MIDAS motif, iv) a Zn^{2+} cation would, albeit slightly, increase the affinity of the MIDAS motif for a carboxylic side chain and v) a Mn^{2+} cation would considerably increase the affinity of the MIDAS motif for a carboxylic side chain.

Table **3.4**: Affinity differences (kcal/mol) of the Ca^{2+}- and the Zn^{2+}- bound *open* MIDAS towards an acetate molecule, with respect to the Mg^{2+}-bound *open* MIDAS, and at different dielectric environments.

	ΔG_{aff}				$\Delta\Delta G_{aff}$		
Solvent	Mg^{+2}	Zn^{+2}	Ca^{+2}	Mn^{+2}	$\Delta\Delta G^{Zn}_{Mg}$	$\Delta\Delta G^{Ca}_{Mg}$	$\Delta\Delta G^{Mn}_{Mg}$
Gas phase	-196.6	-198.9	-187.0	-203.5	-2.3	9.6	-6.9
CCl_4	-97.4	-99.4	-92.7	-97.4	-2.0	4.7	-6.6
Ether	-52.2	-53.9	-49.0	-58-5	-1.7	3.2	-6.3
Acetone	-23.8	-25.5	-22.2	-30.5	-1.4	1.6	-6.7
Water	-1.2	-6.1	-0.3	-6.9	-4.6	0.9	-5.7

Computed as $\Delta\Delta G^{Zn}_{Mg} = \Delta G^{s}_{aff}(Zn-MIDAS) - \Delta G^{s}_{aff}(Mg-MIDAS)$
Computed as $\Delta\Delta G^{Ca}_{Mg} = \Delta G^{s}_{aff}(Ca-MIDAS) - \Delta G^{s}_{aff}(Mg-MIDAS)$

In experiments with isolated I-domains of the integrins LFA-1 and MAC-1 [64], it has been shown that Zn^{2+} and Mn^{2+} can bind to both I-domains (with equivalent *closed* and *open* MIDAS motives). In addition, it is well known that Zn^{2+} can compete with Mg^{2+} in other metal binding sites [219, 220, 221, 222]. Our geometry optimizations indicate that the octahedral coordination of the MIDAS site would be maintained in the presence of Zn^{2+}. Moreover our data indicate that a Zn^{2+} coordinating MIDAS site would have a similar affinity for acetate as a Mg^{2+}-bound MIDAS site.

It has also been shown that Mn^{2+} can substitute Mg^{2+} in several metalloproteins, including LFA-1 [64]. In several cases, this substitution is associated with substantial increase in binding affinities. Our calculations on the model system indicate that not only can Mn^{2+} substitute Mg^{2+} as the metal coordinating cation, but that it would substantially increase the affinity of the *open* MIDAS motif for an acetate.

On the other hand, as mentioned above, Ca^{2+} cation has shown controversial results in similar integrins. In the experiments mentioned above with isolated I-domains [64], Ca^{2+} cation seemed to bind to Mac-1 but not to LFA-1. Our data indicate that structural rearrangements of the MIDAS may take place upon Ca^{2+} binding, as the optimized geometry of the Ca^{2+}-bound MIDAS changed the most, compared to those geometries with other metal cations. The different abilities of Ca^{2+} to bind equivalent MIDAS motifs, might be related to protein flexibility around the MIDAS, which may allow conformational rearrangements in LFA-1 but not in Mac-1.

3.3.4 Affinity for different types of ligands

In order to design novel inhibitory drugs, it is necessary to study the affinity energies of the MIDAS towards different ligands or substrates. In this spirit, the affinity energy of the *open* MIDAS structure $1S_{op}$ was evaluated for 10 different ligands. The reactions studied and the tested ligands are shown in Figure 3.8 and the results are found in Table 3.5.

The affinity values obtained for the formate (1a), acetate (1b), propionate (1c) and butyrate (1d) molecules are compared first. These four molecules differ on the length of the aliphatic chain bound to the carboxylate, and they can be taken as probes of the sensitivity of the ligand toward substituents with increased electron σ-donor capability. ΔG^s_{aff} values in gas phase are similar for all of them but lower for the formate (-192.9, -196.6, -195.4 and -195.2 kcal/mol, respectively; for increasing chain length). This similarity is also observed in water (ΔG^s_{aff} =-0.6, -1.2, -0.5 and -1.1 kcal/mol, respectively). An inspection of the relative affinity values with respect to acetate in Table 3.3.4 (in parentheses) shows that formate is not as good a ligand as acetate; the ($\Delta\Delta G^s_{aff}$ ranges from 3.7 in the gas-phase to 0.6 in water, while the values for propionate and butyrate are in general within 1 kcal/mol of the value for acetate. Hence, it can be concluded that beyond the length of an acetate molecule, the increased electron σ-donor character of the aliphatic chain has a minor effect on the affinities. Therefore selectivity towards an aspartate or glutamate amino acid sidechain in the integrin might be determined by complementary shape of the surroundings of the MIDAS with the vicinity of residues of the natural ligand.

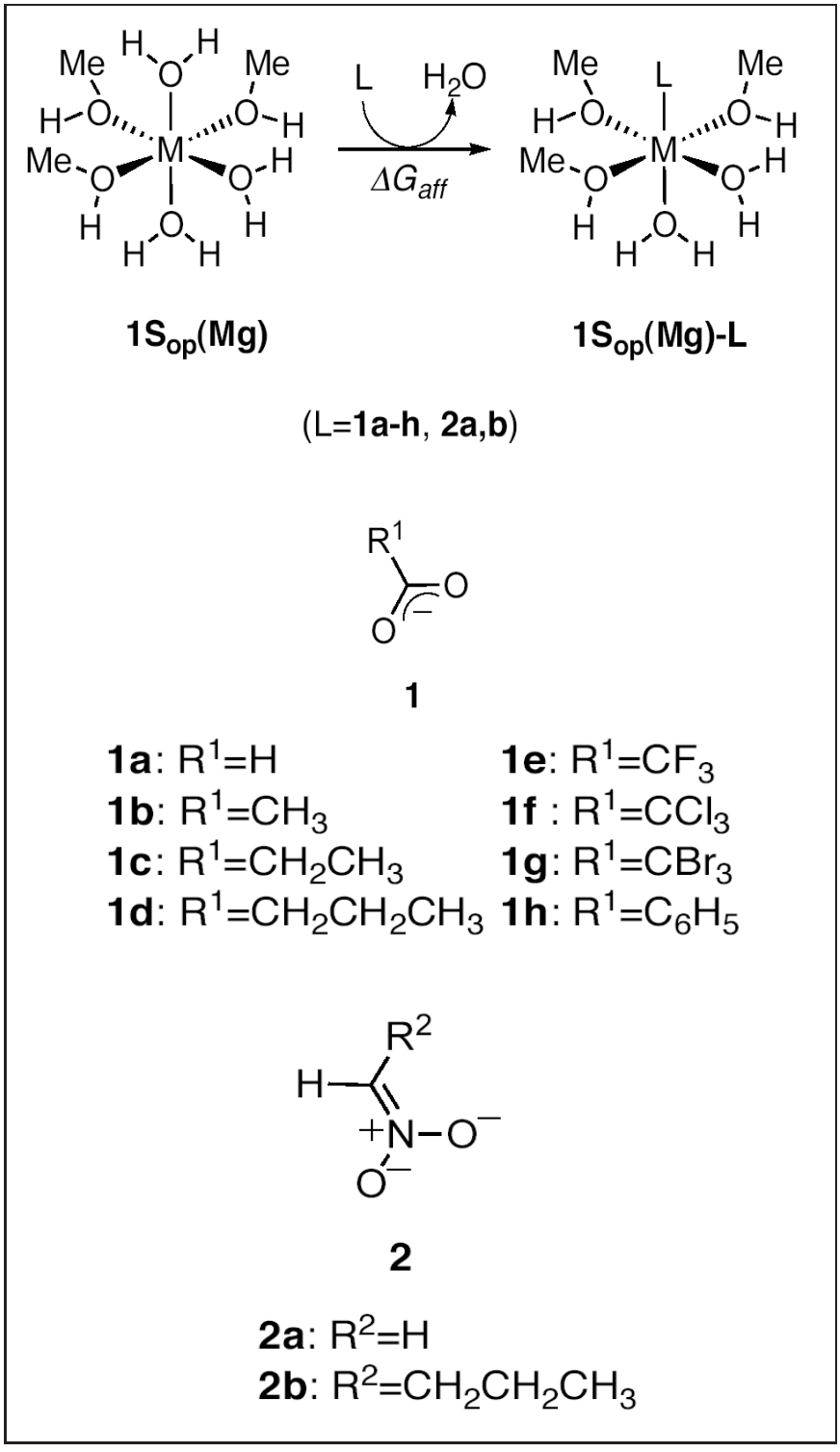


Fig. 3.8: Representation of the reactions studied and the ligands tested:1a=Formate; 1b=Acetate; 1c=Propionate; 1d=Butyrate; 1e=Benzoate; 1f/1g/1h=TriF/TriCl/TriBr-acetate; 2a= Nitromethene; 2b=Nitropropene

Table **3.5**: Comparison of ΔG^s_{aff} values (kcal/mol) obtained for the *open* MIDAS conformations with various ligands and in different model solvents. In parenthesis, the relative affinity ($\Delta\Delta G^s_{aff}$) with respect to the affinity showed towards an acetate, i.e., $\Delta\Delta G^s_{aff}(Ligand) = \Delta G^s_{aff}(ligand) - \Delta G^s_{aff}(Acetate)$, calculated at different dielectric environments.

Solvent	1a	1b	1c	1d	1e	1f	1g	1h	2a	2b
Gas phase	-192.9	-196.6	-195.4	-195.2	-175.2	-173.5	-172.0	-189.6	-191.5	-194.6
	(**+3.7**)	(**0.0**)	(**+1.2**)	(**+1.4**)	(**+21.4**)	(**+23.1**)	(**+24.6**)	(**+7.0**)	(**+5.1**)	(**+2.0**)
CCl_4	-94.2	-97.4	-96.7	-96.9	-83.1	-81.7	-80.6	-92.4	-94.9	-97.6
	(**+3.2**)	(**0.0**)	(**+0.7**)	(**+0.5**)	(**+14.3**)	(**+15.7**)	(**+16.8**)	(**+5.0**)	(**+2.5**)	(**-0.2**)
Ether	-50.4	-52.2	-51.6	-52.1	-43.3	-41.1	-41.8	-47.8	-51.4	-53.8
	(**+1.8**)	(**0.0**)	(**+0.6**)	(**+0.1**)	(**+9**)	(**+11.1**)	(**+10.4**)	(**+4.4**)	(**+0.8**)	(**-1.6**)
Acetone	-22.8	-23.8	-23.8	-24.0	-17.0	-17.9	-14.9	-20.7	-23.9	-26.3
	(**+1.0**)	(**0.0**)	(**0.0**)	(**-0.2**)	(**+6.8**)	(**+5.9**)	(**+8.9**)	(**+3.1**)	(**-0.1**)	(**-2.5**)
Water	-0.6	-1.2	-0.5	-1.1	1.3	4.5	6.3	-1.0	-3.8	-3.9
	(**+0.6**)	(**0.0**)	(**+0.7**)	(**+0.1**)	(**+2.5**)	(**+5.7**)	(**+7.5**)	(**+0.2**)	(**-2.6**)	(**-2.7**)

1a=Formate; 1b=Acetate; 1c=Propionate; 1d=Butyrate; 1e=Benzoate; 1f/1g/1h= TriF/TriCl/TriBr-acetate; 2a= Nitromethene; 2b=Nitropropene

Considered next are ligand substitutions that provoke an electron σ-acceptor inductive effect. For this purpose, the affinity energies for trifluoro-, tricloro- and tribromo- acetate ligands (Figure 3.8: 1e, 1f and 1g, respectively) were calculated. Interestingly, they all show lower affinity energies with respect to acetate. In vacuo, the relative affinity values with respect to acetate are 21.4 kcal/mol, 23.1 kcal/mol and 24.7 kcal/mol, respectively, indicating a substantial decrease in binding capability. When different permittivity values are considered, the above mentioned differences decrease, but, they are all maintained at considerable positive values (2.5, 5.7 and 7.5 kcal/mol, respectively, in water). These results suggest that the presence of substituents with significant σ-acceptor character will lead to a ligand with a lower capacity of electron donation to the metal center, and therefore, with a lower binding affinity.

In line with these results, a benzoate ligand (Figure 3.8: 1h) was tested, which stabilizes the electronic charge of the ligand by a π resonance, and was found to be a ligand that leads to lower affinity energies than the ones showed by the reference acetate ligand. The ΔG^s_{aff} values range from -189.6 kcal/mol in the gas phase to -1.0 kcal/mol in water. From these results we can conclude that the design of inhibitors of the LFA-1/ICAM-1 interaction should include carboxylic acids in their metal binding moiety, but electron withdrawing groups (alfa-halo carboxy donors) or aromatic rings (benzoates or heteroaromatic analogs) should be avoided. On the other hand, these alkyl carboxylate groups must be surrounded by a hydrophobic environment in order to lower the effective ϵ value around Mg^{2+}, thus increasing the binding ability of the MIDAS.

Finally, we also checked the behavior toward negatively charged nitronate ligands (2a and 2b). Affinity values for these ligands were in the same range as for formate and acetate molecules. Furthermore, for nitropropene in a polar media, affinity values were found that were higher than those for acetate. This suggests that nitro functional groups could be a promising substituent of the carboxylate as the protuding group in this sort of ligand without compromising their binding capabilities to the MIDAS.

3.4 Conclusions

The various factors that contribute to the binding energy of integrins and their ligands, such as the nature and coordination of the metal, the activation state of the protein, the nature of the ligand or the experimental conditions, make it difficult to determine experimentally the relative contributions of each individual phenomena; this can be the source of the controversial results in some of these works. In this work, we present the results of quantum calculation on a realistic model of the metal ion dependent adhesion site (MIDAS) of the integrin LFA-1 that significantly help to understand different aspects of the affinity modulating mechanisms of the interaction between the MIDAS motif and key residues in its counter-receptor ICAM-1 (such as E34). This has been accomplished by individually studying the isolated effect on the ligand binding affinity of the different factors mentioned above.

Concerning the two possible coordination modes of the MIDAS motif, we have (i) quantified the change (gain) in ligand affinity related to the *closed* to *open* conformational change in the MIDAS motif, observing that the *open* MIDAS binds E34 of ICAM-1 with a higher affinity than the *closed* MIDAS. A major point in this respect is that the surroundings of the MIDAS motif contribute to the regulation of the affinity towards ligand as much as or to a higher

extent than the conformational change mentioned above. In this respect, we have found that the inclusion of second shell ligands in the MIDAS model dramatically decreases the differences in ligand affinity and that hydrophobic environments (low dielectric value) around the MIDAS and the approaching acetate molecule dramatically increase the strength of this interaction. This behavior helps to understand why the MIDAS motif has only been observed in its *open* state when bound to ICAM-1, and (b) the Mg^{2+} ion in the MIDAS and the anionic E34 residue in ICAM-1 are mainly surrounded by hydrophobic residues in the protein.

Our conclusions related to the effect of the different metals (Mg^{2+}, Zn^{2+}, Ca^{2+} and Mn^{2+}), help to clarify the controversial experimental results obtained for the role of these metals on ligand binding activity of integrins. This controversy may arise from the difficulty to isolate the intrinsic effect of the metals from many other factors that affect ligand binding activity in integrins, as stated by Ajroud et al. [190]. From our results, we observed that Mg^{2+} may be the most abundant but not necessarily the best cation for the purposes of entering the MIDAS and strongly binding to ICAM-1. Other ions, such as Zn^{2+}, and especially Mn^{2+} (IV), could coordinate to the MIDAS motif and substantially favor ICAM-1 binding. On the other hand, too strong bindings are not necessarily desired from a physiological point of view, as there must eventually be a release of ligand at certain point of the biological process taking place. On the contrary, the inhibitory effect of Ca^{2+} cation may be due to the fact that it is not easily accommodated in an octahedral geometry of the MIDAS and its binding may lead to a structural distortion of the MIDAS and the surrounding environment.

Concerning the synthesis of new inhibitors of the LFA-1/ICAM-1 interaction, we concluded that carboxylic acids included in the metal binding moiety would improve the affinity for the LFA-1 I-domain, whereas electron withdrawing groups (alfa-halo carboxy donors) or aromatic rings (benzoates or heteroaromatic analogs) should be avoided because they lead to lower binding affinities. On the contrary, nitronates surrounded by hydrophobic groups, may yield binding affinities as good as or even better than the carboxylic acids. Interestingly, we observed that there is no direct correlation between the capacity of the ligands to donate charge to the metal and the increase in their affinity towards the MIDAS (paper in preparation).

3.5 Supporting Information

ΔG^{g}_{aff} values for the acetate binding reaction of the Mg^{2+}-, Ca^{2+}- and Zn^{2+}-bound $1S_{op}$ MIDAS models were calculated at three different levels of theory to calibrate our method, and data collected in the supplementary material. Computed and experimental metal-oxygen distances can also be found as supplementary material.

3.5.1 Calibration of the method

In previous studies on Mg^{2+}, Ca^{2+} and Zn^{2+} complexes by Peschke et. al. [223] and Dudev et. al [224, 225, 226], it is suggested that calculations at the more expensive B3LYP/6-311++G(d,p) or B3LYP/6-31++G(2d,2p) level, respectively, are needed to converge to results in good agreement with experimental values. To assess whether the use of pseudopotentials, B3LYP/6-311++G(2df,2p)//B3LYP/SKBJ*, leads to reasonable results for the cases of complexes with Zn^{2+} and Ca^{2+} dications, the affinity free energy of the $1S_{op}$ MIDAS model was evaluated at the three different levels mentioned before, using Mg^{2+}, Ca^{2+} or Zn^{2+} as central cations and with acetate as the incoming ligand (Figure 3.9).

Table **3.6**: ΔG^{g}_{aff} values (kcal/mol) in the gas phase for Mg^{2+}-, Ca^{2+}- and Zn^{2+}-bound *open* MIDAS in the ligand binding reaction. In parenthesis, the relative affinity free energies with respect to the Mg^{2+}-bound MIDAS, namely, $\Delta\Delta G^{g}_{aff} = \Delta G^{g}_{aff}(M-MIDAS) - \Delta G^{g}_{aff}(Mg-MIDAS)$ with M=Ca^{2+},Zn^{2+}.

Metal	6-311++G(d,p) [223]	6-31+G(2d,2p) [226]	This work*
Mg^{2+}	-193.3	-196.2	-196.6
	(0.0)	(0.0)	(0.0)
Ca^{2+}	-184.6	-187.2	-187.0
	(+8.7)	(+9.0)	(+9.6)
Zn^{2+}	-195.6	-198.3	-198.8
	(-2.3)	(-2.1)	(-2.2)

* B3LYP/6-311++G(2df,2p)//B3LYP/SKBJ+*

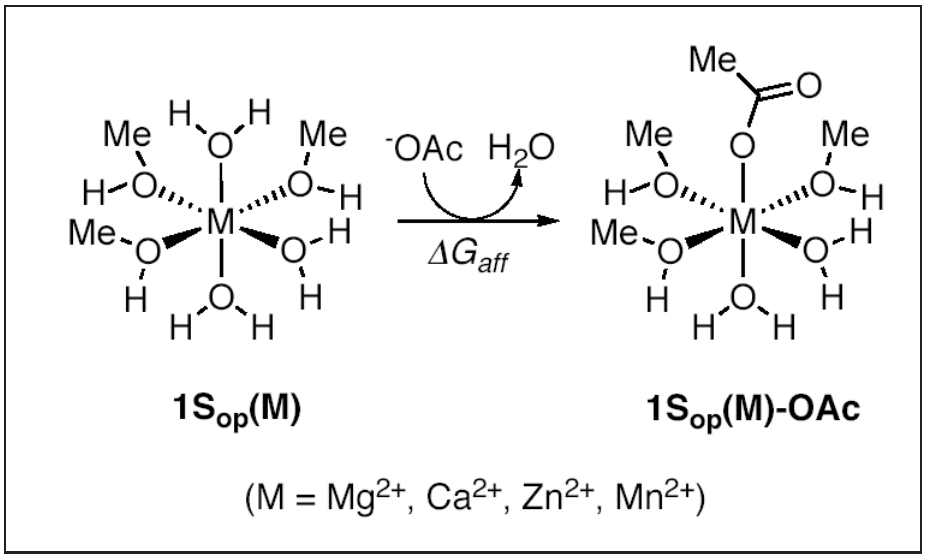


Fig. 3.9: Representation of the acetate-binding reaction for a Mg^{2+}-, Ca^{2+}- and Zn^{2+}-bound $1S_{op}$ MIDAS model.

Comparison of $\Delta\Delta G^{g}_{aff}$ values. Table 3.6 reports the ΔG^{g}_{aff} values obtained with the three methods mentioned before. An excellent agreement is found, but our method compares best with that of Dudev *et al.* [224-226]. Both methods show a consistent difference of ca. 3 kcal/mol with respect to the level of theory used by Peschke et. al. [223]. Nevertheless, the overall trends in affinities are equally reproduced by the three methods, and the agreement is very good if one compares the relative affinities ($\Delta\Delta G^{g}_{aff}$ values in parenthesis in Table 3.6) for each metal with respect to the affinity displayed by the Mg^{2+}-bound MIDAS. We also observed (data not shown) that the affinity energies are not sensitive to the use of unscaled or scaled B3LYP frequencies when evaluating the entropic contributions. In summary, these data suggest that the use of the B3LYP/6-311++G(3df,2p)//B3LYP/SKBJ+* method is appropriate and reliable, especially in the evaluation of trends in relative affinity energies ($\Delta\Delta G_{aff}$).

Comparison of the metal-oxygen distances. To see if this method reproduces the experimentally observed geometries, metal-oxygen bond distances in our optimized geometries were compared with the corresponding metal-oxygen bond distances in 1MJN, 1MQ9 and 1MQ8. In Figure 3.10 and Table 3.7 , shown are the metal-oxygen distance values in our optimized geometries, below which and in parentheses, shown are the corresponding distances in the X-ray structures. For $1S_{op}(Mn^{2+})$ d1 value is missing, as L1 corresponded to a residue belonging to a nearby crystal, whose coordinates were not included in the PDB file in the Protein Data Bank. It can be seen, in any case, that the agreement between both experimental and computational data is very good, expecially metal-oxygen distances involving water molecules.

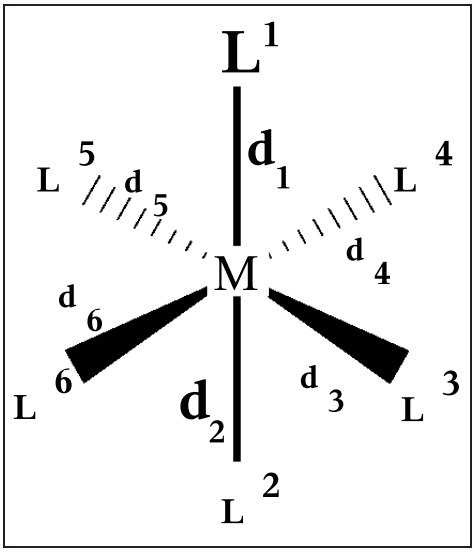


Fig. 3.10: Representation of the Metal-Ligand distances referred to in Table 3.7.

These results confirm that both the computational method and the selected MIDAS motif models provide plausible results.

Table **3.7**: Metal-oxygen distances on the optimized MIDAS complexes obtained in this work. In parentheses, when available, the corresponding distances on the X-ray structures (in Å).

	Structure							
Distance	$1S_{cl}(Mg^{2+})$ (1MJN)	$1S_{op}(Mg^{2+})$-OAc (1MQ8)	$1S_{op}(Mg^{2+})$	$1S_{op}(Mn^{2+})$ (1MQ9)	$1S_{op}(Ca^{2+})$ (-)	$1S_{op}(Ca^{2+})$-OAc (-)	$1S_{op}(Zn^{2+})$ (-)	$1S_{op}(Zn^{2+})$-OAc (-)
d1	2.120^{w}	2.002^{a}	2.110^{w}	1.971^{a}	2.458^{w}	2.313^{a}	2.130^{w}	1.983^{a}
	(2.072)	(2.235)	(-)	(-)	(-)	(-)	(-)	(-)
d2	2.010^{a}	2.160^{w}	2.148^{w}	2.138^{w}	2.462^{w}	2.459^{w}	2.191^{w}	2.222^{w}
	(2.065)	(2.073)	(-)	(2.043)	(-)	(-)	(-)	(-)
d3	2.153^{w}	2.055^{w}	2.123^{w}	2.007^{w}	2.456^{w}	2.388^{w}	2.150^{w}	2.082^{w}
	(2.080)	(1.976)	(-)	(2.044)	(-)	(-)	(-)	(-)
d4	2.045^{m}	2.145^{m}	2.099^{m}	2.410^{m}	2.430^{m}	2.498^{w}	2.090^{w}	2.193^{w}
	(2.072)	(2.076)	(-)	(2.088)	(-)	(-)	(-)	(-)
d5	2.136^{m}	2.120^{m}	2.092^{m}	2.040^{m}	2.426^{m}	2.432^{w}	2.089^{w}	2.162^{w}
	(2.212)	(2.192)	(-)	(2.078)	(-)	(-)	(-)	(-)
d6	2.133^{w}	2.139^{m}	2.113^{m}	2.208^{m}	2.437^{m}	2.446^{w}	2.114^{w}	2.174^{w}
	(2.043)	(2.049)	(-)	(2.080)	(-)	(-)	(-)	(-)

For clarity, the type of ligand is indicated with a superscript. Water:w; Acetate: a ; Methanol: m.

3.5.2 Solvation free energies

We calculated the solvation free energies for each of the structures that characterizes the *closed* and *open* states of the MIDAS: $1S_{op}$, $1S_{cl}$, $2S_{op}$ and $2S_{cl}$ in the acetate-bound or acetate-unbound form, as well as for an acetate and a water molecule (Table 3.8). These energies were combined with the gas-phase affinity free energies to give the total affinity free energy in solution according to Figure 3.5. As expected, the solvation free energies are modulated by the total charge of the structure, the volume of the structure, and the dielectric permittivity of the environment in the following way: the larger the charge, the smaller the volume and the more polar the environment, the larger the solvation free energy of the structure (Table 3.5.2).

Table **3.8**: Solvation free energies (in kcal/mol) for acetate and water molecules, and for the different Mg^{2+}-bound MIDAS models

Structure	CCl4	ether	acetone	water
water	-1.5	-3.6	-4.0	-7.4
OAc^-	-29.3	-44.9	-52.7	-78.2
$1S_{cl}$-W	-18.2	-32.9	-40.2	-53.1
$1S_{cl}$-OAc^-	5.1	0.1	-1.8	-8.1
$1S_{op}$-W	-80.6	-122.9	-148.6	-173.4
$1S_{op}$-OAc^-	-15.9	-30.4	-37.7	-49.2
$2S_{cl}$-W	-11.7	-24.7	-32.1	-41.6
$2S_{cl}$-OAc^-	9.1	4.3	1.6	-2.2
$2S_{op}$-W	-15.1	-29.9	-38.1	-49.9
$2S_{op}$-OAc^-	7.0	1.2	-1.4	-6.5

Chapter 4

DFT vs QMC: Combined study on Mn^{2+}

The results in this chapter were published in :
E. San Sebastian, J.M Matxain, R. H. Stote, A. Dejaegere, F. P. Cossio and Xabier Lopez
Journal of Chemical Physics B 2007 Jul 3; 17608410

Contents

4.1 Introduction

Manganese is highly abundant and accessible in biological systems, and a critical trace element [227]. Many proteins contain a manganese dication as an intrinsic paramagnetic center that can play a catalytic or structural role. Alternatively, it is transported by metalloproteins to tissues where it is needed [228]. In a prototypic biological environment, the first coordination shell of the Mn^{2+} cation often contains a variety of amino acid sidechains and water molecules octahedrically coordinated [40, 229, 230, 231, 232, 233, 234].

The Mn^{2+} cation has an electron configuration of [Ar] $3d^5$ (Figures 4.1 and 4.2). An octahedral ligand-field would provoke the splitting of the five 3d-orbitals of the Mn^{2+} cation into two energy levels known as t_{2g} and e_g; the three low-energy orbitals coming from $3d_{xy}$, $3d_{xz}$, and $3d_{yz}$, and the two high energy ones from $3d_{x2-y2}$ and $3d_{z2}$. The coordinating ligands in a protein environment (amino acids and water molecules) lead, in principle, to only weak ligand fields; it is therefore assumed that in this metal binding site, the energy required for electron-pairing in the metal is not balanced by a stronger interaction with the ligand-field. As a consequence, it has been assumed that in biological/protein environments the ground state of Mn^{2+} is the high (sextuplet) spin state and, therefore, this spin state has been employed in quantum mechanical studies [235, 236, 237, 119].

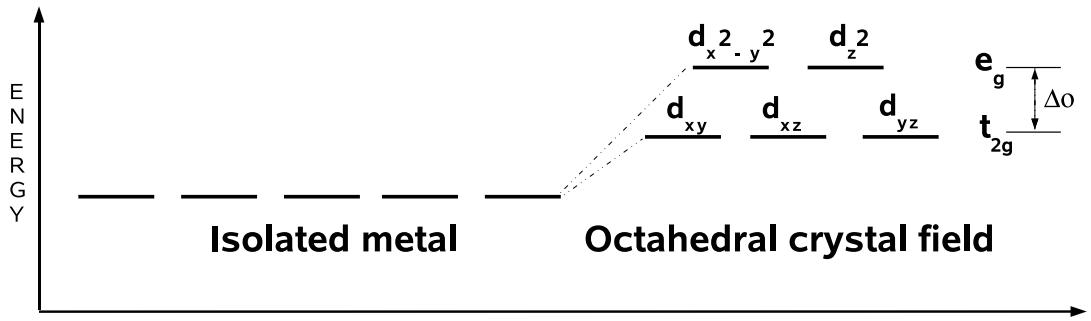


Fig. 4.1: Splitting of the five 3d orbitals of Mn2+ cation in the presence of a ligand field.

The correct prediction of the relative energies between different spin states in transition metal complexes requires the use of accurate ab initio methods, but is, even then, a difficult task. Attempts to predict the high spin state of different Mn^{2+} complexes have failed to do so [119, 120], since a variety of effects need to be accurately taken into account, as will be explained below. In the context of DFT, reliable exchange and correlation potentials are needed. In addition, Mn complexes can show important near-degeneracies, also known as nondynamical electron correlation, which can be problematic in the context of DFT. In fact, pure DFT functionals have been observed to yield incorrect ordering of spin states [120]. On the other hand, the hybrid B3LYP functional, which includes a fraction of the exact exchange, is better able to describe this feature [121]. Therefore, although DFT is widely applied in organometallic systems, caution should be exercised when applying these methods to the study of metals, like Mn+2, where different spin states and near-degenerate orbitals are present. Alternatively, quantum Monte Carlo methods are known to recover 95% of the electron correlation in chemical systems [238].

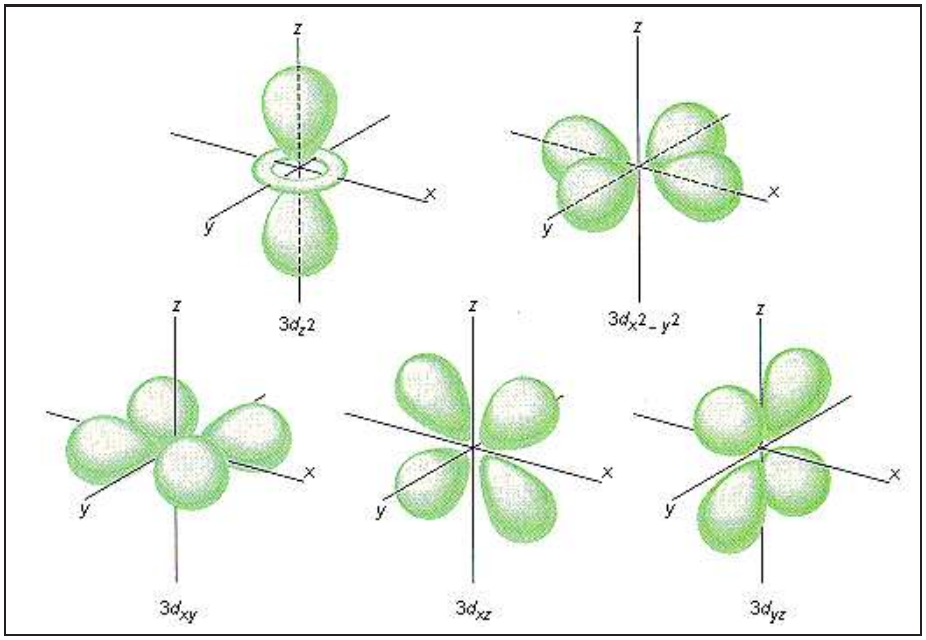


Fig. 4.2: Size and shape of the 5 3d orbitals of the Mn2+ cation.

In particular, the diffusion Monte Carlo method (DMC) together with the variational Monte Carlo method (VMC) have become the method of choice for treating these types of difficult problems [126]. Briefly, in VMC [238] expectation values are calculated using an approximate trial wave function, the integrals being performed by a Monte Carlo method. In DMC [238, 128] the imaginary-time Schrödinger equation is used to evolve an ensemble of electronic configurations toward the ground state. The fermionic symmetry is maintained by the fixed-node approximation [129] in which the nodal surface of the wavefunction is constrained to be equal to that of a trial wave function. Compared to DFT, DMC provides a more exact picture of the interaction under study, but at larger computational effort. We present here a combined B3LYP and DMC study of the influence of prototypical and biologically relevant coordination shells of the Mn^{2+} cation on the stability of low, intermediate and high spin states of the complexes. Three types of coordination shells have been chosen: (i) the hexahydrated Mn^{2+} complex, (ii) the coordination shell found in the metal-ion dependent adhesion site (MIDAS) of the integrin LFA-1 bound to six neutral ligands, known as the open conformation of the MIDAS in LFA-1 (Figure 4.3, left), and (iii) MIDAS bound to five neutral ligands and one negatively charged acetate, representing a typical MIDAS coordination shell when protein-protein interactions are present, known as the open conformation of the MIDAS bound to its natural protein-ligand, ICAM-1 (Figure 4.3, right). The goals of this work are: (i) to check using different high level theoretical methods whether the experimental assumption that Mn-protein complexes are high spin is correct, (ii) to compare the behavior of the B3LYP method in comparison with the more sophisticated DMC approach, (iii) to check to what extent the nature of the coordinated ligands influence the relative stability of high-spin complexes.

4.2 Methods

All geometry optimizations were carried out at the B3LYP level of theory [196, 197]. For Mn, the standard all-electron 6-31+G* basis set was used, while for the remaining atoms the effective core potentials and shared-exponent basis set of Stevens, Krauss, Basch, and Jasien [239] were used, including a set of polarization and diffuse functions for the ligands. This approach is hereafter named B3LYP/SKBJ+*. Vibrational frequency calculations were carried out at the same level of theory, to calculate the enthalpic and entropic contributions to the Gibbs free energy.

The electronic energies were refined performing single-point calculations at the B3LYP/SKBJ+* geometries using two different levels of theory. On one hand, the B3LYP/6-311++G(2df,2p) level, and on the other, the DMC level of theory. For the DMC calculations[1], the trial wavefunctions are written as a product of a Slater determinant and a Jastrow factor, developed recently [124]. The determinantal part was calculated at the UHF level of theory, combined with the relativistic Stuttgart pseudopotentials for Mn, C, and O, and, the 6-311++G(2df,2p) basis set [240] for H, motivated on by their earlier successful performance in DMC calculations on the Fe atom and FeO + H_2 reactions [241, 242]. The Jastrow factor, containing up to 113 parameters, was optimized using variance minimization techniques.

Preliminar calculations showed small influence of the spin state in the bulk solvation free energies of each structure in high dielectric media such as water; therefore, bulk solvent effects have not been further considered in this work, expecting their contribution to be small with respect to the differences in the relative energies among the three spin states. All B3LYP and UHF calculations were carried out with the GAUSSIAN98 [212] and GAUSSIAN 03 [243] program packages. The CASINO v1.7 program [244] was used for DMC calculations.

4.3 Results and Discussion

In this section the obtained results are introduced and discussed. First, the main geometrical features of the Mn-complexes are shown. Then, the influence of the ligand field on the relative energies of the three spin states (high, S = 5/2; medium, S = 3/2; low, S = 1/2) of the Mn-complexes is described and interpreted. Finally, some biological consequences for the particular case of protein-protein recognition are discussed.

4.3.1 Optimized Structures

As described in the Methods section, the geometries are optimized at the B3LYP/SKBJ+* level of theory. Experimental structural data are available for the hexahydrated Mn^{2+} complex3 [215] and for the pseudo-liganded open MIDAS motif (crystallographic structure; PDB ID = 1MQ9, [60]). Comparison between experimental and theoretical metal-ion-oxygen distances are given in Table **4.1**.

[1]We are grateful to Dr. Jon Mattin Matxain for performing the DMC calculations

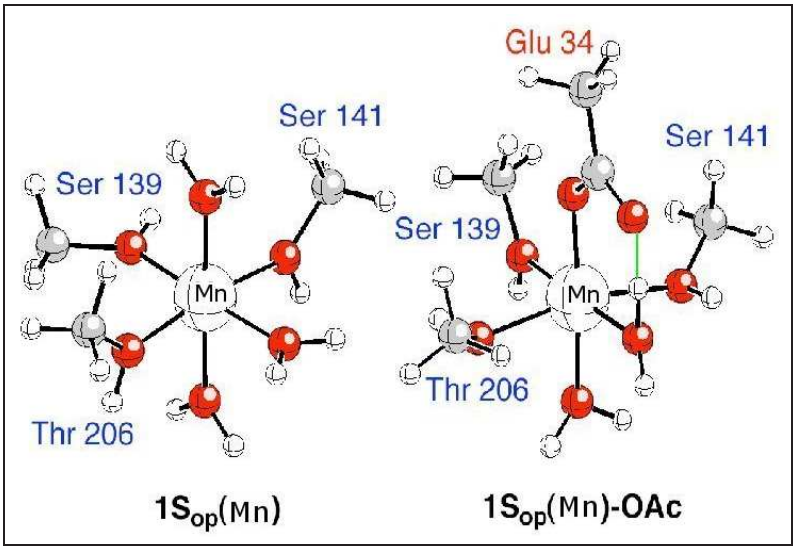


Fig. 4.3: Ball and stick representation of the Mn^{2+}-bound MIDAS motif with reduced models of first shell residues. Left: the open conformation in the absence of GLU-34 from ICAM-1. Right: open conformation in the presence of GLU-34 from ICAM-1. First coordination shells of Mn in the MIDAS motif.

In the case of the hexahydrated Mn^{2+}-complex, the best agreement between theoretical and experimental MnO distances is obtained for the high-spin state, 2.21 Åversus 2.19 Å, respectively. The agreement with experimental data decreases as we increase electron pairing, as the intermediate quadruplet spin-state shows an average MnO distance of 2.13 Å, and the doublet low-spin state an average value of 2.10 Å. These results point to the high-spin state of the complex as the one found in aqueous solution.

In the case of the optimized geometry of the acetate-bound open MIDAS model (1SopAc), it is less clear upon comparison with the experimental X-ray distances whether the values correlate best with low, medium or high-spin states. A closer analysis is, therefore, needed. MnWater distances correlate best with the high-spin states and worse with the low-spin state. However, the low spin state shows the best agreement for Mn^{2+}-methanol distances. The shortest Mn^{2+}-O distance is found for the negatively charged acetate irrespective of the spin-state of the complex: 2.05 Å(doublet), 1.97 Å(quadruplet), and 1.98 Å(sextuplet). Interestingly, MnWater distances are shorter in the MIDAS than in the hexahydrated complex for the sextuplet and quadruplet spin states, but not for the doublet state. For the sextuplet, the MnW distance in the hexahydrated complex is 2.21 Å, whereas in the MIDAS complex, distances of 2.11 Åand 2.04 Åare found for Mn-W$_{ax}$ and Mn-W$_{eq}$, respectively. For the quadruplet state of the (1SopAc) complex, the Mn-W$_{ax}$ distance is longer than for the equatorial Mn-W$_{eq}$ distance, 2.14 Åv.s. 2.01 Å, respectively. However, for the doublet state there is a lengthening of both MnW distances, axial (2.29 Å) and equatorial (2.15 Å), with respect to its value in the doublet hexahydrated complex, 2.10 Å. The experimental trend is to shorten the MnW distances when passing from the hexahydrated complex to the MIDAS, from 2.19 Åto 2.09 Åfor the axial and 1.99 Åfor the equatorial position, in agreement with the behavior of the high-spin states.

Consideration of Mn methanol distances show a better comparison with the low spin state as

Table **4.1**: Experimental versus theoretical MnO Distances (Å) in the Mn^{2+}-complexes

	Exp.	Theory		
Distance		S=1/2	S=3/2	S=5/2
	$Mn(H_2O)_6^{+2}$			
Mn-W_{ax}	2.19[a]	2.10	2.26	2.21
Mn-W_{eq}	2.19[a]	2.10	2.11/2.03	2.21
	$1S_{op} - Ac$			
Mn-Ac_{ax}	-	2.05	1.97	1.98
Mn-W_{ax}	2.09	2.29	2.14	2.11
Mn-W_{eq}	1.99	2.15	2.01	2.04
Mn-M_{eq}	2.20	2.27	2.05	2.07
Mn-M_{eq}	2.23	2.23	2.04	2.12
Mn-M_{eq}	2.33	2.24	2.34	2.08
	$1S_{op} - W$			
Mn-W_{ax}	-	2.22	2.20	2.12
Mn-W_{ax}	-	2.23	2.19	2.12
Mn-W_{eq}	-	2.25	2.07	2.12
Mn-M_{eq}	-	2.18	2.21	2.07
Mn-M_{eq}	-	2.18	2.04	2.09
Mn-M_{eq}	-	2.18	2.21	2.07

a) average experimental Mn-O distance. Ac= acetate; M=methanol; W=water; ax. and eq. stand for axial and equatorial positions, respectively, with complexes oriented as in Figure 2.

the bond distance are longer by 0.1-0.25 Åthan the ones found for the sextuplet state and in good agreement with experiment. The origin of this discrepancy may be due to our use of a reduced model for the ligand. Recall that we use methanol as model for the sidechains of the amino acids Serine and Threonine of the protein structure. Also, as shown in previous quantum calculations of the MIDAS [67] specific second-shell amino acids are likely to have an effect on the geometries, as could the effect of the dielectric environment.

4.3.2 Relative Energies of different spin states

Relative Energies of Different Spin States. The relative energies of each electronic state with respect to the sextuplet high-spin state at B3LYP and DMC levels of theory are given in Table **4.2**, for the Mn^{2+}-complexes $Mn(H_2O)_6^{2+}$, $1S_{op} - W$ and $1S_{op} - Ac$. Results for the Mn^{2+} ion are also provided for comparison. First, we analyze the bare Mn^{2+} ion. Experimentally [245, 246] it has been claimed that the sextuplet is the lowest-energy state for this ion, with the quadruplet and doublet states lying 76.7 kcal/mol and 112.0 kcal/mol higher in energy, respectively. Both B3LYP and DMC are able to reproduce this trend. However, B3LYP calculations show their limitation in that both of these states are overstabilized with respect to the sextuplet by 26.1 kcal/mol for the former and 41.5 kcal/mol for the latter. Employing DMC, the agreement with experiment improves significantly and we obtain ΔE values of 88.8 kcal/mol for the quadruplet and 101.9 kcal/mol for the doublet, in semiquantitative agreement with experimental values as the differences with the experimental results are reduced to around 10-12 kcal/mol.

In the case of the hexahydrated Mn^{2+} complex, there has been a significant stabilization of both low-spin and intermediate-spin states with respect to the ground-state sextuplet. Results for both B3LYP and DMC calculations show the same trend; the relative energies at the DMC level of theory are 73.6 kcal/mol and 89.9 kcal/mol for the quadruplet and doublet states, respectively, and 47.6 kcal/mol and 54.6 kcal/mol, respectively, at the B3LYP level. A similar stabilizing effect of low/intermediate spin states with respect to the isolated Mn^{2+} is observed when the ion is in the MIDAS environment, that being, when Mn+2 is surrounded by six neutral ligands ($1S_{op}$-W). Here again, the B3LYP values are smaller than the DMC ones, but both methods show the same qualitative trend when compared to the hexahydrated Mn^{2+} compounds. Taking $Mn(H_2O)_6^{2+}$ as reference, we observe that the doublet state has increased its relative energy (by approximately 6 kcal/mol at the DMC level), and the quaduplet state has decreased (by around 9 kcal/mol). At the DMC level of theory, the energy values are 95.9 kcal/mol for the doublet state and 66.9 kcal/mol for the quadruplet state. Introduction of the negatively charged acetate into the coordination shell of the ion has significant effect on the relative energies. Both the B3LYP and DMC calculations yield similar values for the relative energies, namely 46.5 kcal/mol (DMC) and 46.5 kcal/mol (B3LYP) for the low-spin state, and 26.7 kcal/mol (DMC) and 33.4 kcal/mol (B3LYP) for the quadruplet spin state. Note that both methods, and especially DMC, show that there is a very high spin-pairing stabilization upon complexation with negatively charged ligands. In biological systems, negatively charged ligands are often acetate-type side chains (Glutamate and Aspartate). With respect to the isolated ion, the ligand-field provided by the 1SopAc environment has reduced the energy differences with respect to the sextuplet state to a third in the case of the quadruplet (from 88.8 kcal/mol to 26.9 kcal/mol) and by half in the case of the doublet state (from 101.9 kcal/mol to 46.5 kcal/mol). This stabilization can be understood by a simple $n_O \rightarrow \mathrm{d}_{Mn}$ dative bond model, in which electron pairing in the d-shell of Mn^{2+} allows for an enhancement of the dative character of the bond between Mn and the lone-pair of the oxygens of the ligands, and therefore, enhanced stabilization of lower spin states upon complexation with negatively charged ligands as these have a higher electron donation capability.

Table **4.2**: Relative Energies of the Different Spin States of the Studied Mn Complexes, Calculated at the B3LYP and DMC Levels of Theory

Spin State	Mn^{2+}		$Mn(H_2O)_6^{2+}$	
	ΔE(B3LYP)	ΔE(DMC)	$\Delta E/\Delta G$(B3LYP)	$\Delta E/\Delta G$ (DMC)
doublet	70.5	101.9±0.4	54.6 / 57.2	89.9±0.9 / 92.5±0.9
quadruplet	50.6	88.8±0.5	47.6 / 47.6	73.6±0.9 / 76.3±0.9
sextuplet	0.0	0.0	0.0 / 0.0	0.0 / 0.0
	$1S_{op}$-Ac		$1S_{op}$-W	
	$\Delta E/\Delta G$(B3LYP)	$\Delta E/\Delta G$ (DMC)	$\Delta E/\Delta G$(B3LYP)	$\Delta E/\Delta G$ (DMC)
doublet	46.5 / 53.1	46.5±1.2 / 53.1±1.2	59.5 / 62.5	95.9±1.0 / 98.9±1.0
quadruplet	33.4 / 35.4	26.9±1.3 / 28.9±1.3	41.5 / 42.7	66.9±1.1 / 68.7±1.1
sextuplet	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0

Table **4.3**: ΔG of the ligand (acetate) binding reaction to the open MIDAS motif (kcal/mol)

Spin State	$\Delta E/\Delta G$ (B3LYP)	$\Delta E/\Delta G$ (DMC)
doublet	-211.5 / -205.6	-222.6±1.5 / -216.7±1.5
quadruplet	-206.9 / -203.5	-213.3±1.6 / -209.9±1.6
sextuplet	-203.4 / -196.2	-173.2±1.5 / -166.0±1.5

In summary, both B3LYP and DMC methods predict a sextuplet ground-state for the Mn^{2+}-complexes studied in this work and with the quadruplet state having lower energy than the doublet spin-state. While indicative of trends, interpretation of the quantitative values of the relative energies of these spin-states must be done with caution as they are highly sensitive to (i) the method used and (ii) the specific coordination shell around the ion, especially the presence of negatively charged ligands. Are the different spin states of Mn^{2+} of biological significance? We try to answer this question in the next section.

4.3.3 Natural-Ligand Affinity and Spin State of the MIDAS

In order to study whether the spin state of the Mn^{2+} could influence the biological function of the MIDAS, we have calculated the affinity free energy of binding to an acetate type molecule. The MIDAS-bound metal in the ligand binding domain of integrin LFA-1 has the primary purpose of recognizing and binding a glutamate side chain present in the ICAM-1 protein, known to be the natural ligand. Recently, we have demostrated that the main relative trends in natural ligand binding can be reproduced by the use of reduced models of the MIDAS like the ones shown in this work, where the glutamate side chain is represented as an acetate [67]. Thus, as a probe of the affinity of the MIDAS motif toward natural ligands can be evaluated by calculating the free energy change associated to the equilibrium below (equation 4.1), where where Mn^{2+}_{MIDAS}-W and Mn^{2+}_{MIDAS}-L are the water or acetate-bound open MIDAS motives, respectively, and W and L represent water and ligand, respectively.The relative free energy ΔG^s_{aff} was calculated for each spin state and the results are summarized in Table **4.3** .

$$Mn^{2+}_{MIDAS} - W + L \quad \stackrel{\Delta G^s_{aff}}{\longrightarrow} \quad Mn^{2+}_{MIDAS} - L + W \tag{4.1}$$

We observe that both the DMC and B3LYP method give that the higher the spin state, the lower the affinity for the ligand. Both methods predict a high affinity for the ligand, with ΔG's of 205.6 kcal/mol at the B3LYP level and 216.7 kcal/mol at the DMC level of theory for the doublet state. The intermediate-spin complex shows a slightly lower affinity, with ΔG's of 203.5 kcal/mol and 209.9 cal/mol at the B3LYP and DMC levels, respectively. Finally, both methods predict the high-spin state of Mn^{2+} to lead to the lowest binding affinities, less than 200 kcal/mol, although with important differences in the values between the methods.

In summary, the different spin states show significant differences in the affinity toward an acetate ligand. The affinity increases with increasing the spin pairing; the difference between low and high spin states is around 50 kcal/mol at the DMC level. Comparing with the relative energies between different spin states presented in the previous section, the results suggest that the gain in affinity obtained from a better interaction with the negatively charged ligand in the low spin state does not counterbalance the energy penalty to be paid for electron pairing in this particular system. Thus we can conclude that ligand binding occurs in the high spin state of Mn^{2+}.

4.4 Conclusions

We have applied DFT/B3LYP and quantum diffusion Monte Carlo methods to predict the spin state of the Mn^{2+} cation in prototypical biological environments. Our calculations confirm that the sextuplet high-spin state of this ion is the lowest-energy spin state for the complexes studied, in agreement with the standard asumption of high-spin ground states for Mn complexes in proteins. However, the relative energies of the various spin-states (sextuplet, quadruplet, and doublet), are highly sensible to the ligand field provided by amino acid side chains. In particular, there is a significant stabilization of doublet and quadruplet spin states with respect to the sextuplet upon complexation with a negatively charged acetate ligand, which can represent the glutamate or aspartate sidechains. From a methodology perspective, B3LYP is able to yield the right ordering of the relative energies of the spin states and their trends upon ligand-field stabilization, even though the B3LYP values for the bare ion and for the complex with 6six neutral ligands are too low with respect to the experimental and DMC results. However, one has to take into consideration that for transition metal systems [120] with large exchange and correlation effects, HF methods or pure density functional theory (DFT) methods often yield incorrect ordering of spin states. In this sense, it is remarkable that B3LYP with a proper basis set is able to reproduce the qualitative trend of energies among the various spin states as found by other authors [121] without the need to include second-shell ligands or counterions.

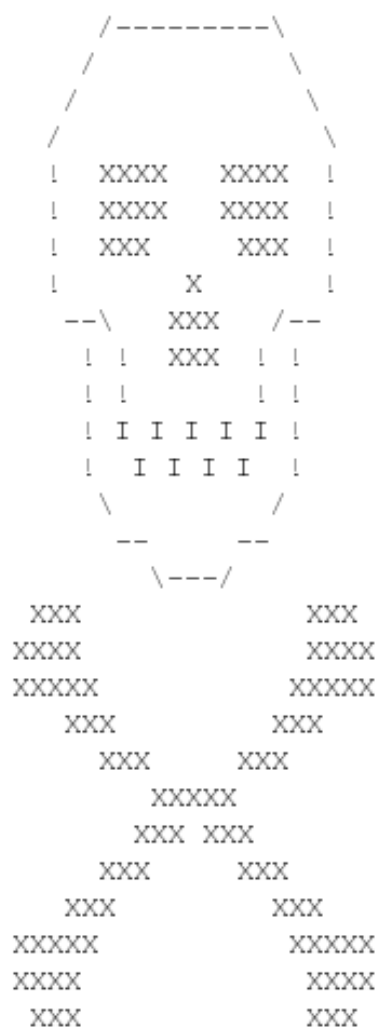
Finally, we comment on the dependence of natural-ligand binding affinity on the spin state. From the calculations presented here, we find lower binding affinities for the higher spin states. However, it is unlikely that Mn^{2+} undergoes a spin change to optimize the ligand affinity, as our results suggest that the energy cost of electron-pairing is higher than the gain by an enhanced ligand-binding affinity.

Chapter 5

Locating Hotspots: LFA-1/ICAM-1 and LFA-1/Inhibitor

Contents

***** EOF during read
****
BOMLEV (-1) IS REACHED - TERMINATING. WRNLEV IS 4



Execution terminated due to the detection of a fatal error.

Brief summary of Chapter 5

The analysis of the X-ray structure of the αL I-domain in complex with ICAM-1 published in the Brookhaven Protein Data Bank (PDB hereafter) provides us with the distances and geometries of interacting amino acid residues, the details of which depend on the resolution of the structure. X-ray and NMR structures are a key piece of information that can help us to understand the affinity modulating mechanism of the integrin/ligand interactions. Unfortunately, X-ray structures by themselves represent but a single snapshot of a dynamic protein in solution. They say little about the underlying binding energies and the eventual location of "hotspots" in the LFA-1/ICAM-1 interface. As interacting hotspots may be responsible for a major part of complex formation, it is important to identify any hotspots in the LFA-1/ICAM-1 interface that may therefore be important for the regulation of the affinity or the strength of the interaction between ICAM-1 and the αL I-domain in the three known activation states. The goal of this theoretical part of the present thesis work was to establish a relationship between specific residues or hotspots in the interaction of the low, intermediate and high affinity structures of the αL I-domain with ICAM-1 and the experimentally measured affinity constants [60], i.e., their increasing affinity for the natural ligand ICAM-1 (Kd=1500μM, 3μM and 0.15μM, respectively).

With this goal in mind, a deeper sampling/analysis was performed on the structures of the αL I-domain and its complexes with ICAM-1 by means of molecular dynamics (MD hereafter) simulations. In addition, to locate the interaction hotspots, MD simulations were combined with a computationally tractable yet reliable Free Energy Decomposition (FED hereafter) protocol based on the Molecular Mechanics/Poisson Boltzmann Solvent Accessible model (MM/PBSA hereafter) [171, 172, 173]. Binding free energies were calculated and decomposed on a per-residue basis, systematically analyzing the magnitude and the nature of the contribution of all the amino acids in the interaction between the low, medium and high affinity conformations of the I-domain of integrin LFA-1 with its counterpart ICAM-1.

In the following paragraphs, the reader will come across the expressions "low affinity complex" and "high affinity complex". Since no low affinity or high affinity complexes between the αL I-domain and ICAM-1 have been solved to date, it is important to notice that these expressions refer to ICAM-1 binding to the low affinity and high affinity αL I-domains, respectively.

All the work presented in this chapter was carried out in collaboration with the Laboratoire de Biophysicochimie Moléculaire, Institut de Chimie at Université Louis Pasteur (Strasbourg, France) under the supervision of Professor Roland Stote, as well as in collaboration with the Biocomputing Group of Structural Biology and Genomics Department at the Ècole Superieure de Biotechnologie de Strasbourg-IGBMC (Illkirch, France) under the supervision of Professor Annick Dejaegere. All the Free Energy Decomposition calculations were submitted and managed by the PhD student Elyette Martin.

5.1 Protein preparation and MD details

To analyze the protein interface and to identify possible hotspots, an analysis based on the MM/PBSA method [171, 172, 173] was performed. Multiple conformations were extracted from molecular dynamics simulations (described below) and used in this analysis. The results were averaged over all conformations extracted, as explained in the Methods section. The advantages of using the MM/PBSA method are that it is relatively fast and reasonably accurate. It is fast because one avoids the sampling of the solvent configurations (the configurations of the solute are sampled in the MD simulations) and accurate, because long-range electrostatic interactions, ionic-strength and some aspects of polarization are included. As explained in Chapter 2, a thermodynamic cycle is considered where the ligand-free proteins (the free I-domain in three possible conformations, and the free ICAM-1 molecules) interact in the gas phase to form the low affinity (L.A. hereafter), medium affinity (M.A. hereafter) and high affinity (H.A. hereafter) complexes, and the binding free energy of the interaction in the gas phase is obtained, ΔG^0_{bind}. This value is added to $\Delta\Delta G^{solv}_{bind}$, i.e, the free energy cost of desolvating the system, to obtain the total binding free energy in solution, ΔG^s_{bind}. The free and bound proteins required certain modifications prior to being subject of the MM/PBSA study. The changes performed are summarized in the lines below.

o Protein Preparation

Engineered disulfide bonds: The CYS residues that were mutationally introduced to stabilize the intermediate (PDB ID=1MQ8) and open conformations (PDB ID=1MQ9) (L161C/F299C and K287C/K294C, respectively) of the αL I-domain were reverted in this study to the wild-type sequence, and the sidechain orientations were rebuilt using the Scrlw2.5 program, as explained in the Methods section (http://www.fccc.edu/research/labs/dunbrack/sidechain.html ; see [247, 248, 249]). The chosen low affinity αL I-domain structure (PDB ID=1ZOP) contained no mutationally introduced disulfide bonds.

Determination of protonation states: The pKa of titratable groups were determined for both isolated proteins and complexes, using continuum electrostatics, as described by Schaefer and coworkers [165], and as summarized in Chapter 2. The UHBD program [159] was used to calculate the free energy of different protonation states and a series of cshell scripts were used [165] to obtain the titration curves (data not shown). The protonation states of selected residues were determined at physiological pH of 7.4. The most remarkable result is that obtained for residue HIS264, which is located in the complex interface, that was predicted to consist of a neutral histidine residue (See Figure 2.5 in Chapter 2.)

Building Hydrogens: Hydrogen atoms were added to the structures mentioned above using the HBUILD module of the CHARMM [134] program version 29a1. The all atom protein force field [146] was used. Each structure was subjected to an energy minimization of 2000 steps using the steepest descent and adopted basis Newton Raphson algorithms (See Methods). Harmonic constraints added initially to prevent any large perturbation of structure were gradually reduced over the course of the minimization. The minimization was stopped when the gradient of the energy was less than 10^{-4} kcal.mol^{-1}.$\AA^{-1}$. A non-bonded cutoff of 13 $\AA$ and a distance dependent dielectric (4r, where r is the distance in $\AA$ between two atoms) were used during the energy minimization processes.

Free and Bound Proteins: The coordinates of the ligand-free low affinity I-domain (PDB ID: 1ZOP), pseudoliganded high affinity I-domain (PDB ID: 1MQ9) and the intermediate I-domain bound to Domain 1 of ICAM-1 were obtained from the Protein Data Bank. ICAM-1 present in the medium affinity complex (PDB ID=1MQ8) was isolated and docked in this study to the L.A and H.A ligand-free structures, to form the L.A and H.A complexes, respectively (Figure 5.5). With respect to the docking procedure, and unless otherwise indicated, a rigid body procedure was followed, were the L.A ligand-free I-domain and the H.A pseudoliganded I-domain crystal structures were superimposed onto the I-domain of the intermediate affinity complex structure. The CHARMM software package was used to minimize the root mean square deviation (RMSD hereafter) values between the backbone atoms of each I-domain.

Regarding possible conformational changes on ICAM-1 upon binding to the low or high affinity I-domains, RMSD values of the backbone atoms of each residue in the crystal structures were analyzed of both the free form (PDB ID=1IC1) and the I-domain-bound form of ICAM-1 (ICAM-1 in 1MQ8). These values were found to be small (<2 $\mathring{A}$), specially for the interface residues (data not shown), which indicates that there may be very small conformational changes on the free ICAM-1 upon its binding to the intermediate affinity I-domain. It was assumed that this change would also be small upon ICAM-1 binding to the low and high affinity conformers of the I-domain, and a rigid body docking was thus performed.

o Details of the MD simulation

Accurate application of the MM/PBSA methodology requires that averages must be obtained over a number of representative conformations [250]. For this purpose, MD simulations were carried out on all the complexes under study using the CHARMM software package [134] and the all atom protein force field [146]. The energy minimized protein-protein complexes were solvated in a 74 $\mathring{A}$ x 74 $\mathring{A}$ x 74 $\mathring{A}$ box of TIP3P water molecules. Bond lengths involving bonds between heavy atoms and hydrogen atoms were constrained using the SHAKE algorithm [251] during the whole process. Periodic boundary conditions were set and with the protein-protein complex fixed, the water molecules were equilibrated for 20 ps at 300 K. The constraints were removed and the entire system was again equilibrated for 20 ps. Van der Waals interactions were truncated at a cutoff distance of 12.5 $\mathring{A}$ using the switch function. Electrostatic interactions were truncated at 12.5 $\mathring{A}$ using the *fshift* function.

5.2 ICAM-1 binding to the αL I-domain

LFA-1 binding to its ligands is highly regulated by conformational changes induced by inside-out [14, 18, 19] signaling events, whereas conformational changes provoked due to outside-in [16, 17] signaling are not completely evident yet. These inside-out signals provoke conformational changes in the integrin (Figure 5.1-top, A and B), including its ligand binding domain or I-domain (Figure 5.1-bottom). The I-domain adopts the so-called dinucleotide-binding or Rossmann fold [41, 51] with α helices surrounding a central β-sheet (Figure 5.1, bottom).

In many of the crystal structures of the isolated αL I-domain disulfide bonds were introduced by site-directed mutagenesis[1] in order to stabilize the I-domain in the so called *open* conformation [60], the *intermediate* conformation [60], or the *closed* conformation [42, 43, 44, 45, 60], respectively, which showed decreasing affinities for ICAM-1 in biochemical experiments [60]. As explained in the paragraph below, the main structural differences of the three reported affinity conformers of the αL I-domain concern i) the identity of some of the metal cation coordinating residues [252], ii) the arrangement of surrounding loops and α helixes [252] and iii) the occupancy of the so-called "ratchet" pocket by a residue preceding the α7 helix.

The metal ion and the residues to which it is coordinated constitute the well-known MIDAS motif [41], which has been observed in the crystal structures to adopt two alternative coordination modes, referred to as *closed* (inactive) and *open* (active). In addition, in the *open* form, there is a movement of the ion of about 2 $\AA$ toward the solvent. Concerning more general structural differences among the three different affinity conformers, there are conformational changes of the loops that bear the coordinating residues (β1-α1, α3-α4 and β4-α5; see Figure 5.1-bottom) [252], changes in the hydrophobic core of the central β sheet (known as the ratchet pocket) and in the α helices distant from the ligand binding site, such as the C-terminal helix, α7, that moves 10 $\AA$ down the side of the domain. The occupancy of the ratchet pocket by residues preceding the C-terminal α7 helix (L295, F292 and L289, in the low, intermediate and high affinity structures, respectively) is characteristic of the different affinity states of the I-domain. The latter, however, is beyond the goal of the analysis performed in this thesis book.

[1] Site-directed mutagenesis is a molecular biology technique in which a mutation is created at a defined site in a DNA molecule, usually a circular molecule known as a plasmid. Following the mutagenesis, the DNA molecule is usually inserted in bacterial cells for protein expression and purification.

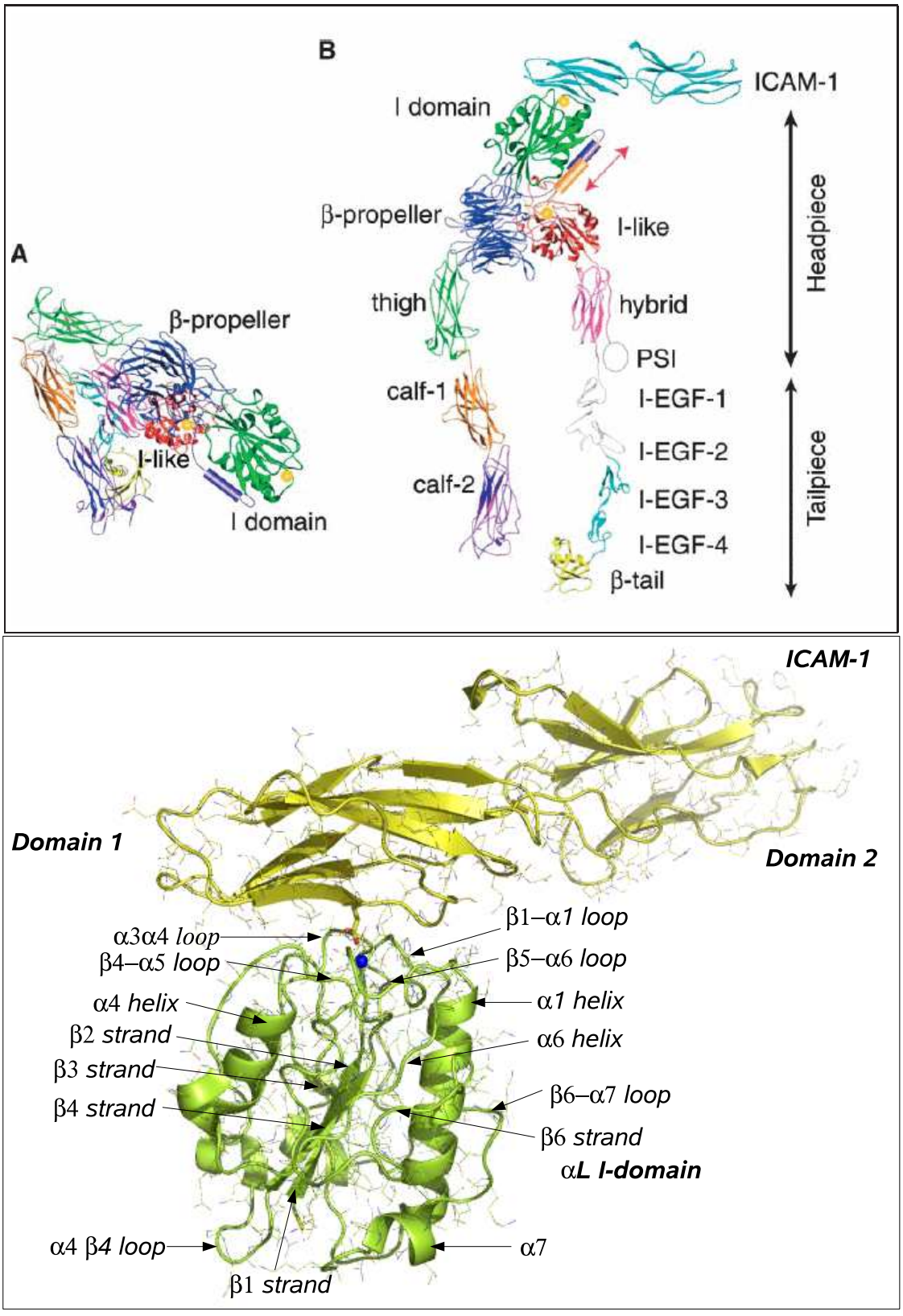


Fig. 5.1: Top: Cartoon representation of the extracellular domain of the LFA-1 integrin; A) Inactive bent conformation and B) active extended conformation. Bottom: Cartoon representation of the crystal structure of the αL I-domain in the medium affinity conformation in complex with ICAM-1 (Domain 1 and Domain 2), PDB ID= 1MQ8 [60].

5.2.1 Structural Characterization

The Medium Affinity Complex

o *Structural rearrangements on the Medium Affinity complex*

In order to analyze structural rearrangements in the complex during the MD simulation, RMSD time-series were calculated and plotted in Figure 5.2. The latter shows the RMSD time-series of backbone atoms of ICAM-1 (crosses) and the M.A. αL I-domain (solid line) separately, together with the RMSD values of all the backbone atoms in the system (dots), which tell about the presence of a global reorientation of ICAM-1 with respect to the I-domain. The MD simulation was 2 ns long and the RMSD with respect to the minimized X-ray structure stabilized early in the dynamics. As observed, the relative orientation of ICAM-1 with respect to the intermediate affinity I-domain does not vary significantly in the course of the simulation, since the RMSD values of αL I-domain and ICAM-1 account for most of the RMSD of all the backbone atoms in the system. Similarly, the RMSD value of ICAM-1 or the αL I-domain separately with respect to the minimized crystal structure of the medium affinity complex stabilized at around 1.2 $\mathring{A}$ and 2 $\mathring{A}$, respectively, which indicates that the overall conformation of each counterpart does not change significantly during the MD simulation. These results support the idea that the structure of the M.A. complex as it is published in the PDB (PDB ID=1MQ8) represents a stable complex between the I-domain and ICAM- 1. A per-residue decomposition of the RMSD values to locate dynamic parts of the system is beyond the scope of this chapter, but is being carried out and will be included in a future report.

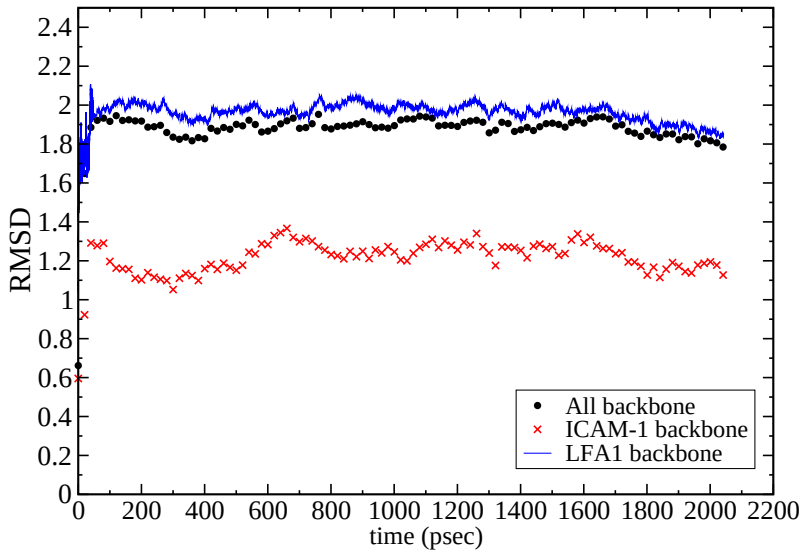


Fig. 5.2: Time series of the RMSD values for all the backbone atoms of the M.A. complex (dots), for all the backbone atoms of ICAM-1 (crosses) or the medium affinity I-domain (line).

The binding free energy: As observed in Table **5.1**, the total free energy for the medium affinity complex formation was estimated to be -53.02 kcal/mol, which further suggests that the medium affinity complex as it is in the PDB may represent an stable complex.

Table **5.1**: Contribution of ICAM-1 and the αL I-domain to the free energy of binding. Values correspond to energy values estimated over a single geometries of each complex (the docked structure prior to the MD simulation). Values are given in kcal/mol.

Protein	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
			Low Affinity				
ICAM-1*	-1.71	6.36	4.64	-5.43	-1.60	-7.03	-2.39
I-domain*	-3.44	7.82	4.38	-5.43	-1.49	-6.92	-2.54
Complex*	-5.16	14.18	9.02	-10.86	-3.09	-13.95	-4.93
ICAM-1**	-54.11	47.98	-6.13	-23.19	-4.14	-27.33	-33.46
I-domain**	-50.08	42.46	-7.62	-23.21	-4.23	-27.44	-35.06
Complex**	-104.20	90.45	-13.75	-46.40	-8.38	-54.78	-68.52
			Medium Affinity				
ICAM-1	-36.19	31.60	-4.58	-15.25	-3.83	-19.08	-23.66
I-domain	-53.96	43.51	-10.44	-15.26	-3.65	-18.91	-29.35
Complex	-90.14	75.12	-15.02	-30.51	-7.48	-37.99	-53.02
			High Affinity				
ICAM-1	-24.78	18.97	-5.81	-6.66	-2.76	-9.42	-15.23
I-domain	-38.83	33.72	-5.10	-6.66	-2.56	-9.22	-14.32
Complex	-63.61	52.69	-10.92	-13.31	-5.32	-18.64	-29.56

* Complex generated using a rigid body docking procedure; * * Complex generated using a Monte Carlo docking procedure; Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar (sum of vdW and SAS components); TOTAL = Total binding free energy (Total Elec. + Total NP)

The Low Affinity Complex

o *Quality of the starting structure*

As mentioned in the Methods section, ICAM-1 was docked to the low affinity and the high affinity forms of the αL I-domain in order to structurally characterize a possible low affinity and a high affinity complex between the αL I-domain and ICAM-1, which have not been cocrystallized to date. Prior to running MD simulations with both complexes, a first analysis was performed on the quality of the structures that resulted from the docking protocol followed (see Methods section). With this goal in mind, the interfaces thus generated were analyzed via submission of the complexes to a per-residue decomposition of the free energy of binding following the MM/PBSA method, which permitted the localization of possible steric clashes present in the complexes.

No steric clashes: A per-residue decomposition of the free energy of binding of the low affinity complex showed a complete absence of steric clashes (data not shown) and, therefore, the latter was used as starting geometry for a MD simulation. The L.A. complex is shown in Figure 5.5, together with the M.A. and H.A. complexes. A significant difference among the structures is the distance between residue E34 and the Mg^{2+} cation, which, in the L.A. complex (*closed* MIDAS motif) is known to be less exposed to the solvent. The latter may directly influence the strength of the complex generated. In fact, it is observed in Table **5.1** that the total binding free energy of ICAM-1 to 1ZOP to form a low affinity complex is -4.93 kcal/mol, which is relatively small but still favorable.

o *Structural rearrangements on the LA complex (I)*

The low affinity complex is not stable: Figure 5.3 shows the RMSD time-series of the L.A. I-domain (solid line) and ICAM-1 (crosses) separately and the RMSD time-series of all the backbone atoms in the system (dots). The MD simulation was 1.3 ns long and the RMSD with respect to the initial geometry stabilized after approximately 100 ps and remained stable for 500 ps. Later in the MD simulation, the RMSD values of all the backbone atoms increased with respect to the I-domain or ICAM-1 alone, i.e., the relative orientation of ICAM-1 and the low affinity I-domain changed significantly. Analyzing a snapshot of the complex at the end of the simulation (not shown), it was observed that ICAM-1 had parted away, thus dissociating the initial L.A. complex. The results summarized above can be regarded as an indication that the low affinity I-domain as it is in the crystal structure (PDB ID=1ZOP) needs to undergo structural rearrangements in order to form a complex with ICAM-1. Therefore, additional structural changes are required prior to ICAM-1 binding. It can also be inferred from the results that ICAM-1 is not able to provoke these structural rearrangements in the isolated I-domain. This result may represent an additional support for the theory of inside-out signaling in integrins.

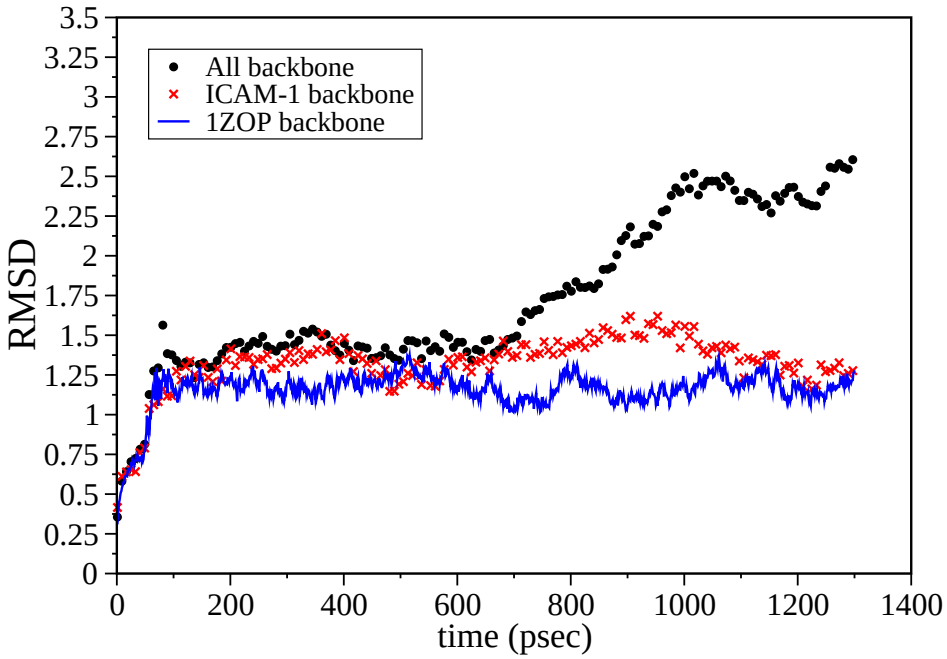


Fig. 5.3: Time series of the RMS deviation values of all the backbone atoms of ICAM-1 (crosses) and the L.A. I-domain (solid line) separately, and of all the backbone atoms (dots). 1ZOP is the PDB ID of one of the crystallized low affinity I -domain of LFA-1 (see text).

o *Generation of a new Low Affinity Complex*

In order to discard the possibility that a poor starting geometry of the L.A. complex is responsible for the instability of the MD simulation, a new L.A. complex was built following a different docking protocol based on the Monte Carlo procedure (see Chapter 2). The procedure and the results are explained below.

A virtual axis was defined between the Mg^{2+} cation in the low affinity I-domain and the carboxylic carbon atom (CD atom hereafter) in E34 in ICAM-1 (Figure 5.4). This axis guided the movements of ICAM-1 towards the I-domain during the Monte Carlo procedure. ICAM-1 was drawn toward the I-domain while the latter was free to rearrange in order to adapt to the approaching ICAM-1. The initial distance between the mentioned atoms along this axis was set to 7 $\mathring{A}$, and ICAM-1 (referenced in atom E34 CD) was moved 1 $\mathring{A}$ towards the Mg^{2+} cation after each one of the 5 loops that were run. Each loop was 2000 steps long, during which ICAM-1 could perform a translation motion of a maximum of 0.5 $\mathring{A}$ and a rotation of a maximum of 10 degrees in each step. In addition, the maximum change allowed to happen in torsion angles was of 20 degrees.

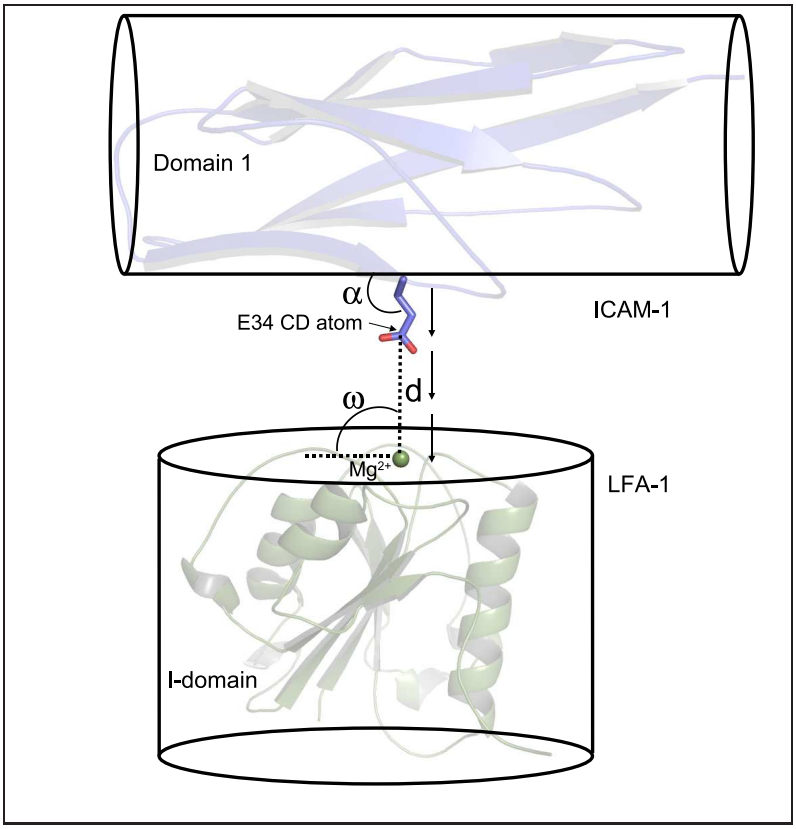


Fig. 5.4: Scheme used in the alternative docking between the domain 1 of ICAM-1 and the low affinity I-domain of LFA-1. The distance d was decreased along the docking process, but fixed in each loop. The relative ω and α angles, as well as other internal coordinates, were allowed to vary within certain limits.

After every rotation, translation or torsion, the system was minimized during 20, 20 and 10 steps, respectively, using the steepest descent method. Simultaneously, all the atoms in the I-domain located at a maximum distance of 15 $\AA$ from the E34 CD atom were allowed to move and adapt to the new position of ICAM-1. After the fifth loop, the atoms in the system were released and a short MD simulation was run for 2000 steps. The structure generated was used as starting geometry for a MD simulation of the low affinity complex. During the MD simulation, the initial geometry was released to rearrange as necessary in order to minimize the energy of the complex, to break the complex or to reinforce it.

A low affinity complex with a new interface: During the Monte Carlo docking, ICAM-1 moderately rotated over the imaginary axis between the E34 CD atom and the Mg^{2+} cation (Figure 5.4), generating an alternative interface for the low affinity complex. With respect to the changes in the I-domain after the flexible docking procedure, the $\alpha 7$ C-terminal helix, the $\alpha 1$ helix and the $\alpha 1$-$\beta 1$ loop (Figure 5.1-bottom) underwent some minor rearrangements (not shown).

Dramatic increase in the binding free energy: As observed in Table **5.1**, the initial total free energy of binding for the new low affinity complex was estimated to be -68.52 kcal/mol, which represents a dramatic increase on the strength of the complex compared to the previous structure (-4.93 kcal/mol). This result suggests that the present low affinity complex may be more stable than the one generated with the rigid body procedure.

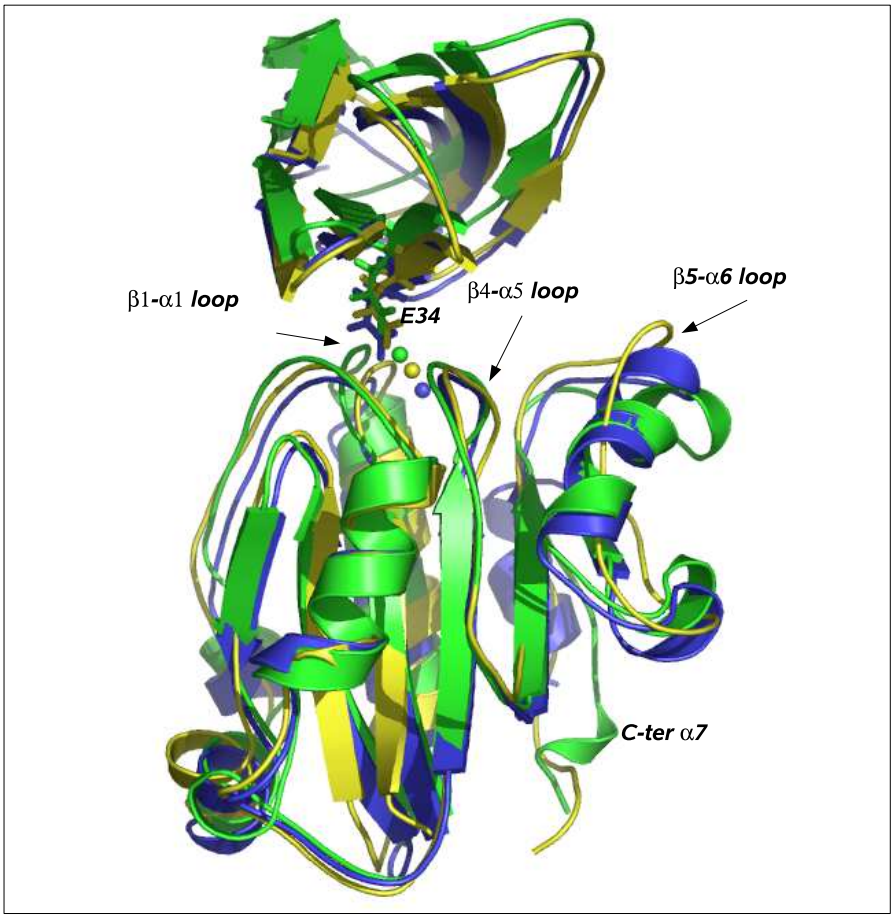


Fig. 5.5: Cartoon representation of the low affinity (blue), medium affinity (yellow) and high affinity (green) complex structures used as starting geometries for the separate MD simulations.

o *Structural rearrangements on the low affinity complex (II)*

From the MD simulations of ICAM-1 bound to LFA-1 in the low affinity conformation, the overall structural changes were characterized. First, in Figure 5.6, shown is the RMSD time-series of the low affinity αL I-domain and ICAM-1 separately, as well as the total RMSD (ICAM-1 + I-domain), which can be used to identify the presence of a global reorientation of one protein with respect to the other. The MD simulation was only 300 ps long by the time this thesis book was written but the RMSD with respect to the initial geometry stabilized early in the dynamics. Again, the relative orientation of ICAM-1 with respect to the low affinity I-domain seems to have varied significantly. A visual inspection of the complex after the short MD simulation reveals that ICAM-1 has slightly moved away from the low affinity I-domain (figure not shown).

The total RMSD of the system (dots) seems to have reached a maximum value at the end of the simulation (around 2.25 $\mathring{A}$), but longer simulation times are still required to avoid ambiguous results. On the other hand, the RMSD value of ICAM-1 or LFA-1, are stabilized at around 1.25 $\mathring{A}$, which indicates that the overall conformation of each counterpart has only varied moderately. Despite the fact that the present MD simulation is short, later in this chapter an analysis is performed on the results of a free energy decomposition analysis on the obtained snapshots, since the total binding free energy indicates that this L.A. complex is significantly more stable than the previous one, and therefore, the assumption is made that the MD simulation would have revealed that the present L.A. complex is stable along longer periods of simulation.

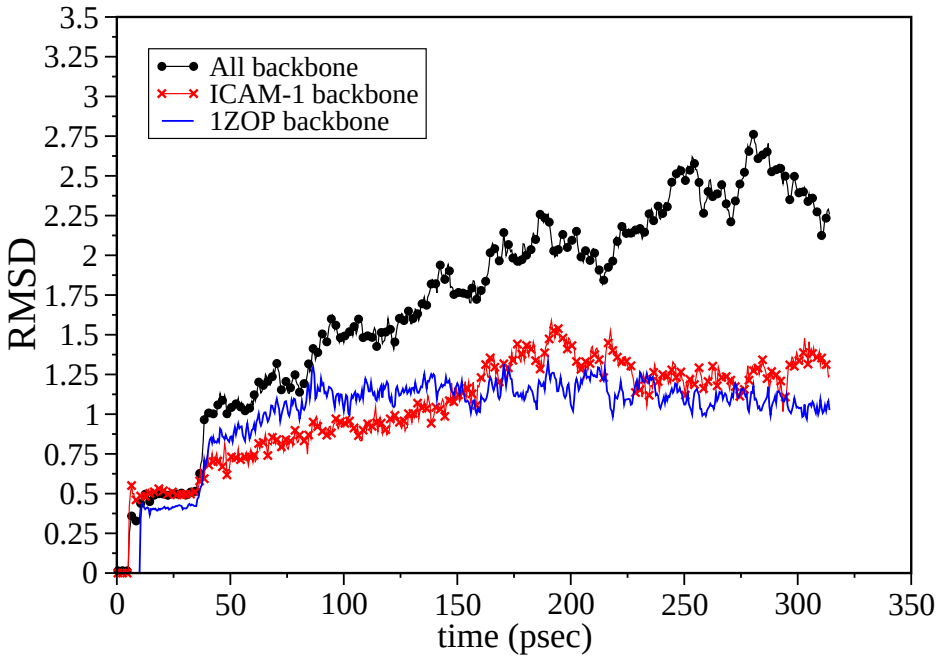


Fig. 5.6: Time series of the RMSD values for all the backbone atoms of the L.A. complex (dots), for all the backbone atoms of ICAM-1 (crosses) or the low affinity I-domain, PDB ID = 1ZOP (line).

The High Affinity Complex

o *Quality of the starting structure*

As with the L.A. complex, the starting H.A. structure was obtained by using the CHARMM programme to overlay the high affinity I-domain (1MQ9) onto the medium affinity I-domain (1MQ8), so that the RMSD values of the backbone atoms of both conformers was as small as possible. The geometry obtained was submitted to a free energy decomposition analysis using the MM/PBSA method, so that possible steric clashes present in the complex were identified.

A new interface in the H.A. complex: Contrary to what happened to the L.A. complex, the H.A. complex generated with the rigid body docking methodology showed important steric clashes (data not shown). Residues E34, K39, M64, Y66, N68, Q73 and T75 in ICAM-1 and residues S139, M140, S141, L205, T243, the Mg^{2+} cation and one of the water molecules coordinated to the metal disrupted the interaction with important vdW clashes. Interestingly, all these residues are the ones known to interact with each other in the M.A. complex [60]. These data suggest that the mentioned residues may need to reorient with respect to the others in order to form a high affinity complex. Since these steric clashes would not have yield an stable MD simulation, an alternative H.A. complex was generated, as explained below.

o *Generation of a new high affinity complex*

In order to modify the first H.A. complex generated, ICAM-1 was systematically pulled apart a distance of 0.5 $\mathring{A}$ from the I-domain using the CHARMM programme and taking the distance between E34 CD atom in ICAM-1 and the Mg^{2+} cation in the H.A. I-domain as reference. Each new H.A. complex generated was subjected to the MM/PBSA protocol and decomposed in a per-residue basis; according to the latter results, ICAM-1 was pulled apart until no more steric clashes were detected (data not shown). It was necessary to displace ICAM-1 about 2 $\mathring{A}$ away from the I-domain, as can be observed in the final structure shown in Figure 5.7. As listed in Table **5.1**, the free energy of formation of this alternative H.A. complex was estimated to be -22.23 kcal/mol. This energy value suggests that the H.A. complex is less stable than the low affinity or the medium affinity complexes and, therefore, in order to be in line with experimental results, the system should undergo conformational rearrangements during the MD simulation in order to display a binding free energy value which is at least qualitatively correct.

Comparison to the crystal structure of the H.A. αL I-domain in complex with ICAM-3: In order to further verify the good quality of this second high affinity complex, the latter was overlaid onto the high affinity structure of the αL I-domain with ICAM-3 (PDB ID=1TOP). It was observed that both complexes are closely related, being ICAM-1 and ICAM-3 are equally oriented with respect to the high affinity αL I-domain, despite the fact that it is possible to observe that E34 in ICAM-1 is further from the Mg^{2+} cation than E37 in ICAM-3 (data not shown). This difference is probably due to the displacement of ICAM-1 in order to avoid the steric clashes mentioned above.

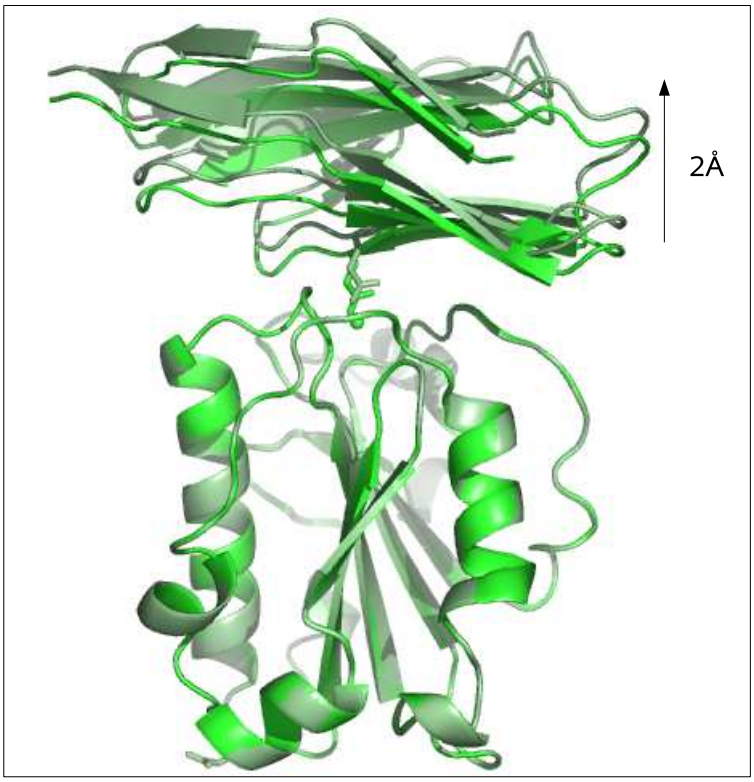


Fig. 5.7: Cartoon representation of the first (green) and second (pale green) high affinity complexes generated, before and after the Monte Carlo docking procedure, respectively.

o *Structural rearrangements on the H. A. complex*

From the MD simulations of ICAM-1 bound to the H.A. αL I-domain, structural changes were characterized. First, in Figure 5.8, shown is the time-series of the RMSD values of the I-domain and ICAM-1 separately, and the RMSD for all the backbone atoms in the high affinity complex. The MD simulation was 850 ps long. The RMSD value of the backbone atoms in ICAM-1 and LFA-1 was stable for the first 500ps. From then on, the RMSD values started to increase significantly and regularly, which means that conformational changes may be occurring on both counterparts. As observed, the RMSD value of all the backbone atoms of the system increases at the same rate than the RMSD of ICAM-1 or 1MQ9. From the

latter it is inferred that the relative orientation of ICAM-1 with respect to the high affinity I-domain may not vary significantly during the MD simulation, but the individual proteins may have undergone structural rearrangements. The latter interpretation is verified looking at the snapshots generated (data not shown).

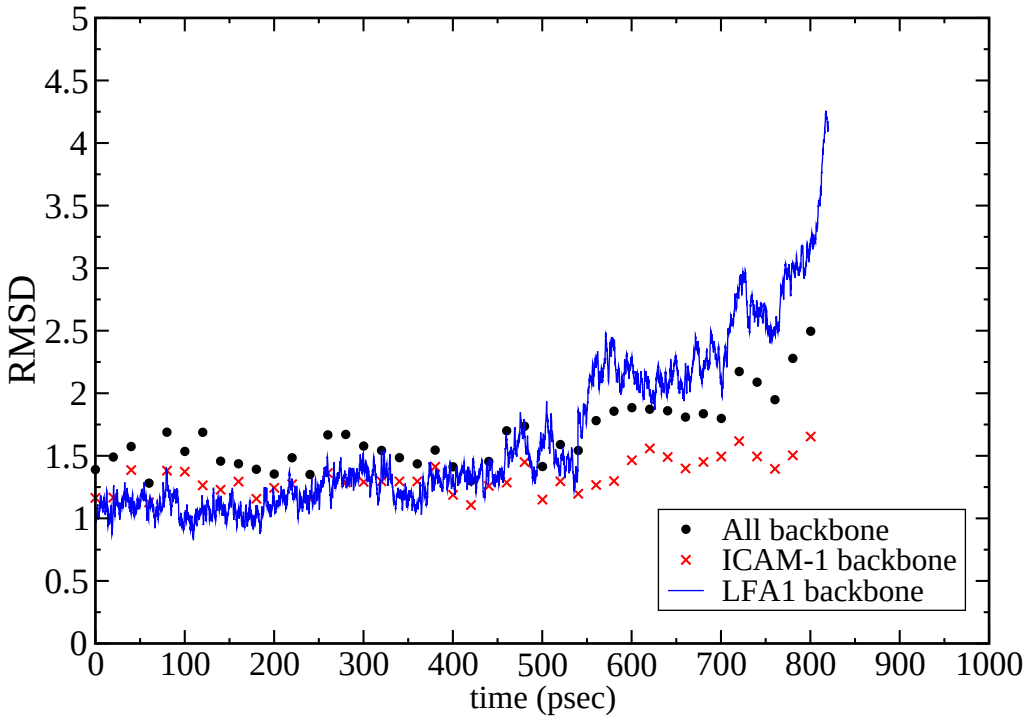


Fig. 5.8: Time series of the RMSD of all the backbone atoms in ICAM-1, LFA-1 or the whole complex.

Big rearrangements on the C terminal $\alpha 7$ helix: Despite the fact that the per-residue analysis of the conformational rearrangements occurring during the MD simulations is beyond the goals of this thesis book, it is interesting to explain the conformational changes mentioned in the paragraph above. As observed in Figure 5.9, in the presence of ICAM-1 the $\alpha 7$ helix significantly modifies its position with respect to the minimized crystal structure of 1MQ9, which was the initial geometry for the I-domain in the H.A. complex. This rearrangement consists of the helix flipping away from the central core of the protein. This was also predicted by a normal mode analysis study by Gaillard and coworkers [253].

ICAM-1 binding to the H.A. αL I-domain provokes significant conformational changes in a reduced region of the domain, the C-terminal α7 helix. This behavior has not been observed in MD simulations run on the disulfide-free H.A. αL I-domain in the absence of ICAM-1. For the sake of brevity, the latter results were not included in this thesis book. In Figure 5.9, shown are structures obtained of the H.A. αL I-domain before and after having run the MD simulation in the presence of ICAM-1. Figure 5.9 shows that, together with less significant changes, a marked swing motion is observed in the C-terminal α7 helix. hese results suggest that ICAM-1 may initiate a signaling process at the top of the domain that reaches the C-terminal α7 helix, which may be related to the outside-in signaling referred to in the literature [16, 17].

Binding free energy of the initial H.A. structure: As observed in Table **5.1**, prior to the MD simulation mentioned above, the total free energy of ICAM-1 binding to the high affinity complex was found to be -29.56 kcal/mol. The L.A. complex showed therefore the best binding free energy, followed by the M.A. complex and the the H.A. complex. At the end of the MD simulation run with each complex, the structural rearrangements described in the pages before should have provoked a change on the total binding free energy of the systems as to modify this affinity ranking and convert the H.A. complex into the highest affinity complex. The latter outcome is analyzed in following sections.

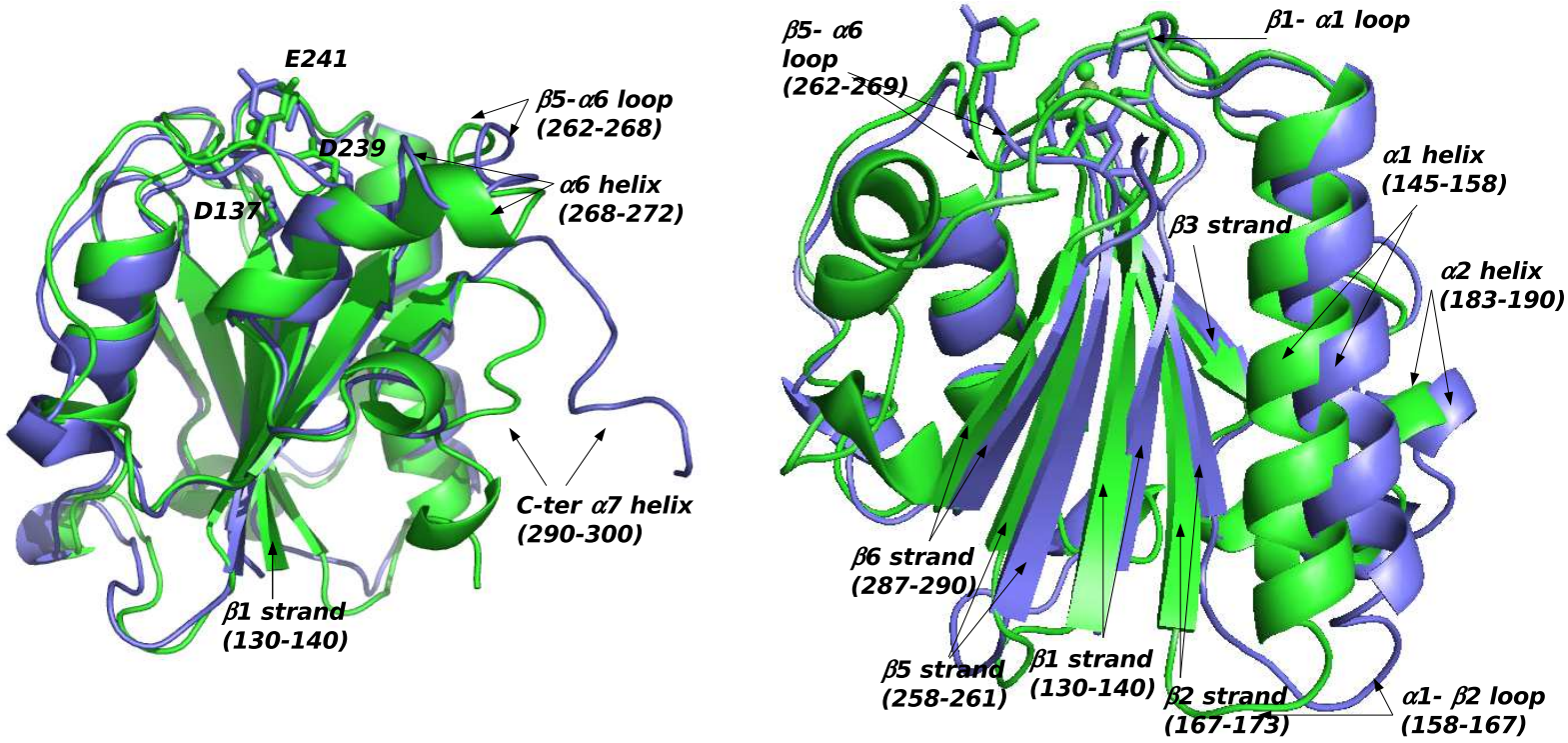


Fig. 5.9: Cartoon representation of the crystal structure of the high affinity αL I-domain in the absence of ICAM-1 (grey) and in the presence of ICAM-1 after the simulation (green). The most relevant residues are highlighted with arrows. The most significant conformational change is the swing motion performed by the C-terminal α7 helix.

5.2.2 Free Energy Decomposition Analysis

For each snapshot generated in the MD simulation, gas-phase interaction energies (E_{gas}) were calculated by summing contributions from internal energies (including bond, angle and torsional angle energies), electrostatic energies, and vdW energies, using the CHARMM programme. The snapshots were grouped into 10 families, according to the calculated gas-phase interaction energies, and the structure in each of the ten groups with the closest E_{gas} to the average E_{gas} of the group, was chosen for the subsequent free energy decomposition analysis (10 structures in total). This subset of snapshots was subjected to a free energy decomposition analysis in order to analyze the effect of the structural rearrangements on the total binding free energy. Later in this section the total biding free energy values will be decomposed, analyzed and discussed on a per-residue basis, in order to locate the binding hotspots in each complex.

Total binding free energy

In Table **5.2**, shown are the the total free energy values of ICAM-1 binding to the low, medium and high affinity I-domains, values which are the average over a number of hotspots extracted from the MD simulations. As observed, the total free energy of ICAM-1 binding to the low affinity αL I-domain was found to be -28.46 kcal/mol, approximately 36 kcal/mol smaller than the total free energy of ICAM-1 binding to the M.A. I-domain (64.0 kcal/mol). Similarly, the total free energy of ICAM-1 binding to the high affinity αL I-domain was found to be -67.78 kcal/mol, in the order of 4 kcal/mol higher than for its binding to the medium affinity αL I-domain. These results were compared with the same values calculated using as input, for the FE analysis, the starting geometries in each MD simulation (Table **5.1**). According to the latter, the structural rearrangements or protein-protein rearrangements analyzed in the preceding sections have converted the complex of ICAM-1 with the high affinity αL I-domain into the highest affinity complex, and the complex of ICAM-1 with the low affinity αL I-domain into the lowest affinity complex.

The docking protocols, the chosen geometries, the MD simulation protocols and the free energy decomposition protocols followed in this study, gave rise to qualitatively reasonable results.

In the following paragraphs, Table **5.2** will be studied in terms of the different type of contributions to the total binding free energy of each complex, which will shed light onto the nature of the complexes generated.

Table **5.2**: Contribution of ICAM-1 and the αL I-domain to the free energy of binding, in the L.A., M.A and H.A. complexes (kcal/mol). Results are weighed over ten coordinate sets selected from multiple snapshots grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

Protein	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
			Low Affinity				
ICAM-1	-155.11 (2.95)	174.10 (0.99)	18.99(2.77)	-23.66(0.82)	-3.84 (0.03)	-27.50	-8.51(2.66)
I-domain	-136.38 (9.81)	144.18 (12.40)	7.80 (2.66)	-23.67 (0.82)	-4.08 (0.02)	-27.75	-19.95 (2.45)
Complex	-291.50 (12.38)	318.29(12.65)	26.79 (2.01)	-47.33(1.63)	-7.93 (0.03)	-55.26	-28.46 (0.96)
			Medium Affinity				
ICAM-1	-101.24 (8.31)	+83.73 (10.19)	-17.50 (3.71)	-15.76 (1.63)	-3.39 (0.07)	-19.15	-36.65 (3.58)
I-domain	-170.13 (11.82)	+161.89 (12.78)	-8.24 (3.40)	-15.76 (1.63)	-3.36 (0.07)	-19.12	-27.36 (3.38)
Complex	-271.36 (19.56)	+245.62 (22.12)	-25.74 (5.78)	-31.52 (3.26)	-6.75 (0.13)	-38.27	-64.01 (5.59)
			High Affinity				
ICAM-1	-104.09 (2.87)	+80.75 (2.82)	-23.35 (3.13)	-13.97 (1.21)	-2.99 (0.08)	-16.96	-40.31 (3.63)
I-domain	-167.80 (5.45)	+157.22 (6.97)	-10.58 (3.49)	-13.97 (1.22)	-2.93 (0.05)	-16.90	-27.48 (3.08)
Complex	-271.89 (5.83)	+237.96 (8.54)	-33.92 (5.18)	-27.94 (2.43)	-5.92 (0.11)	-33.86	-67.78 (8.54)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar (sum of vdW and SAS components); TOTAL = Total binding free energy (Total Elec. + Total NP)

The low affinity complex is stabilized by van der Waals contacts: As observed in Table **5.2**, the vdW contacts in the L.A. complex (-47.33 kcal/mol) are crucial for its stabilization. Despite the fact that the favorable electrostatic interactions (-291.50 kcal/mol) are much larger than the favorable vdW interactions, the former are "minimized" by a high electrostatic desolvation penalty (+318.29 kcal/mol). This makes the total electrostatic component unfavorable (+26.79 kcal/mol). It can be inferred, therefore, that the low affinity complex is mainly a vdW complex.

The medium affinity complex has more vdW than electrostatic interactions: Similar to the L.A. complex, the vdW interactions (-31.52 kcal/mol) are less important than the electrostatic interactions (-271.36 kcal/mol), but the high electrostatic desolvation penalty paid by the complex (+245.62 kcal/mol) decreases the efficiency of the electrostatic contacts. As a whole, the M.A. complex is thus stabilized mainly by vdW contacts (-31.52 kcal/mol), being total electrostatics (-25.74 kcal/mol), nevertheless, much more important than in the L.A. complex (+26,8 kcal/mol), and quantitatively significant.

In addition, in the M.A. complex the I-domain shows a higher contribution to electrostatic contacts (-170.13 kcal/mol) than ICAM-1 (-101.24 kcal/mol), which implies also a higher desolvation penalty to be paid (+161.89 kcal/mol vs 83.73 kcal/mol, respectively).

The high affinity complex, shows an equilibrium between vdW and electrostatic interactions: The H.A. complex is stabilized by important electrostatic contacts (-271.89 kcal/mol) which, in spite of being hidden behind a high electrostatic desolvation penalty (+237.96 kcal/mol), are still favorable (-33.92 kcal/mol). This value is moderately higher than, but still comparable to, the vdW interactions that stabilize the complex, which sum up -27.94 kcal/mol.

Summarizing the results above it can be said that when going from the low affinity to the medium affinity and to the high affinity complex of the αL I-domain with ICAM-1, the weight of favorable vdW interactions to the total binding free energy decreases with respect to the weight of the total electrostatic interactions (interactions + desolvation). In this respect, in the low affinity complex, the total non-electrostatic component represents the 200 % of the total binding free energy. In the medium affinity complex, the total electrostatic and total non-electrostatic components represent, the 40.2 % and 59.8 %, respectively, of the total binding free energy. In the high affinity complex, these values sum 50 % each, having reached an equilibrium between the non-electrostatic and electrostatic components.

The increased affinity of the H.A. αL I-domain for ICAM-1, results from a decrease in the electrostatic desolvation penalty: The electrostatic desolvation penalty paid by the H.A. I-domain to bind ICAM-1 (+157.22 kcal/mol) is 4.67 kcal/mol smaller than the one paid by the M.A. αL I-domain (161.89 kcal/mol). On the other hand, the desolvation of hydrophobic surfaces is equivalent (-2.93 kcal/mol and -3.36 kcal/mol, respectively). This means that, whereas the surface area of desolvated non-polar residues exposed to the interface seems to be equivalent, there seems to be less polar/charged residues exposed by the H.A. αL I-domain to the binding interface than by the M.A αL I-domain. The electrostatic desolvation penalty paid by ICAM-1 is moderately lower (-2.98 kcal/mol) in the high affinity complex formation (80.75 kcal/mol) than in the M.A. complex formation (83.73 kcal/mol).

As will be explained later in the text, the above mentioned changes in the H.A. complex, are related to a decrease in the number of interactions established, and an increase in their efficiency.

The M.A. and the H.A. I-domain, are highly similar structures: Comparing values in Table **5.2** for the separate contribution of ICAM-1 or the αL I-domain, to the M.A. and the H.A. complex stabilization, it is observed that the total contribution of the H.A. αL I-domain (1MQ9) to the formation of the complex with ICAM-1 (-27.48 kcal/mol) is equivalent to the contribution of the M.A αL I-domain (-27.36 kcal/mol). The RMS difference of the backbone atoms of both conformers is small (data not shown), and the total ICAM-1 binding free energies are also similar. The same is true for ICAM-1 in both complexes, whose total contribution is just -3.66 kcal/mol higher when binding to the H.A αL I-domain (-40.31 kcal/mol) than to the M.A. αL I-domain (-36.65 kcal/mol). I.e., the small difference in the free energy of binding of ICAM-1 to the medium affinity and the high affinity αL I-domain, does not arise from significant structural differences in the αL I-domain, such as the position of the C-terminal α7 helix, but from the development of improved interactions between residues in the interface (see later).

Per-residue decomposition of the total free energy of binding

In the analysis of the results reported below, only those residues with a total contribution to the total binding free energy > 0.5 kcal/mol (|DG|, absolute value) were considered as being important for complex formation or complex stabilization, all other residues being excluded from any direct interpretation. In Table **5.3** found is a summary of the number and nature of hotspots found in each partner and in each complex. In addition, in Tables **5.4**, **5.5**, **5.6**, **5.7**, **5.8** and **5.9** a list of residues is found, those which in ICAM-1 and the αL I-domain were observed to make a significant contribution to the free energy of binding. The data and the conclusions are analyzed in the paragraphs below.

The relative number of hotspots is not directly related to the relative binding free energies: The number and type of hotspots found in the L.A., M.A. and H.A. complexes (Tables **5.4**, **5.5**, **5.6**, **5.7**, **5.8** and **5.9**), are summarized in Table **5.3**. As observed, 40 hotspots in the L.A. complex, 28 in the M.A. complex and 23 hotspots in the H.A. complex showed significant contribution to the binding free energy. Among them, the favorable hotspots sum up 25 residues in the L.A. complex, 18 residues in the M.A. complex and 17 residues in the H.A. complex. These results suggest that a higher number of favorable hotspots does not imply a higher (more negative) binding free energy. In the context of ICAM-1, it is observed that the latter utilizes 21 (12+9) residues to bind to the L.A. I-domain, 13 (9+4) residues to bind to the M.A. I-domain and 8 (7+1) residues to bind to the H.A. I-domain. The L.A. I-domain utilizes 19 residues to bind to ICAM-1, whereas 14 residues and 15 residues, respectively, are utilized by the M.A. and the H.A. I-domain, respectively.

Table **5.3**: Number and nature of the contribution of the hotspots found in the L.A., M.A. and H.A. complexes. In parentheses, the contribution of the hotspots to the total binding free energy in kcal/mol.

		Favorable hotspots		Unfavorable hotspots	
Complex	# Hotspots	ICAM-1	I-domain	ICAM-1	I-domain
Low Affinity	40 (-28.46)	12 (-25.69)	13 (-32.59)	9 (+20.76)	6 (+15.86)
Medium Affinity	28 (-64.01)	9(-40.01)	9(-32.80)	4(+4.60)	5(+7.77)
High Affinity	23 (-67.78)	7(-41.42)	10(-31.51)	1(+2.39)	5(+5.65)

The increase in affinity is accompanied by a decrease in the number of favorable hotspots: Comparing the number of residues, in the different I-domains, that established favorable interactions with ICAM-1, it is observed that the low affinity I-domain established 13 of these, whereas the intermediate affinity and the high affinity I-domain established 9 and 10 residues each, respectively. The number of residues in the I-domain that favor binding to ICAM-1 is the highest in the case of the L.A. I-domain, being a higher number of favorable hotspots, therefore, not related to an increase in affinity for ICAM-1. The same analysis was done to ICAM-1 residues. The number of favorable hotspots in ICAM-1 decreases from 12 (when interacting with the L.A. I-domain) to 9 (when binding to the M.A. I-domain) and 7 (to the H.A. I-domain). Interestingly, again, the number of favorable contributions decreases significantly.

The increase in affinity is accompanied by a decrease in the number of unfavorable hotspots in ICAM-1: The number of residues in ICAM-1 contributing unfavorably to I-domain binding decreases as in the case of favorable hotspots; from 9 residues when binding to the L.A. I-domain, to 4 and 1 residues, respectively, when binding to the M.A. and H.A. I-domains. There is a substantial decrease in the number of hotspots, both favorable and unfavorable. Alternatively, the number of residues in the I-domain contributing unfavorably to the free energy of binding decreases from 6 in the low affinity I-domain, to 5 in the intermediate and high affinity I-domains; i.e, the change in the number of hotspots does not vary significantly as to be considered responsible for the increase in affinity for ICAM-1. In summary, the total number of unfavorable hotspots decreases from 15, to 9 and to 6, when going from the L.A. to the M.A. and H.A. complexes, respectively, and may be responsible for the increase in their affinity for ICAM-1.

o The low affinity complex: Identifying hotspots

The low affinity complex will be now described in a per-residue basis; the description is based on the results listed in Tables **5.4** and **5.5**.

Hotspots in ICAM-1 (Table **5.4***):* The residues that contribute favorably to the free energy of binding are: L30 (-3.47 kcal/mol), I33 (-0.83 kcal/mol), E34 (-1.39 kcal/mol), P36 (-1.58 kcal/mol), L37 (-1.37 kcal/mol), P38 (-0.78 kcal/mol), K50 (-4.40 kcal/mol), M64 (-1.28 kcal/mol), Y66 (-6.63 kcal/mol), N68 (-3.50 kcal/mol), P70 (-1.57 kcal/mol) and D71 (-0.56 kcal/mol) (Table **5.4**-top). In the same way, the residues that contribute unfavorably to the free energy of binding of ICAM-1 to the low affinity I-domain are K29 (1.26 kcal/mol), G32 (0.81 kcal/mol), K39 (3.91 kcal/mol), K40 (1.72 kcal/mol), E41 (7.94 kcal/mol), L44 (0.69 kcal/mol), Y52 (2.17 kcal/mol), T75 (0.73 kcal/mol) and K77 (1.33 kcal/mol).

Hotspots in the L.A. I-domain (Table **5.5***):* The same analysis as the one above was done for the low affinity I-domain. It was observed that D137 (-1.76 kcal/mol), M140 (-0.69 kcal/mol), S141 (-0.86 kcal/mol), L205 (-1.96 kcal/mol), D230 (-0.73 kcal/mol), T243 (-5.0 kcal/mol), D244 (-3.04 kcal/mol), S245 (-4.34 kcal/mol), K263 (-3.75 kcal/mol), H264 (-6.44 kcal/mol), Q266 (-1.02 kcal/mol), T267 (-1.46 kcal/mol) and water 303 (-2.54 kcal/mol) are favorable residues in the interaction with ICAM-1. The residues that contribute unfavorably are E146 (0.73 kcal/mol), E241 (3.21 kcal/mol), K268 (0.62 kcal/mol), E269 (0.90 kcal/mol), S270 (0.59 kcal/mol) and Mg 301 (7.81 kcal/mol).

The interacting hotspots listed in the paragraphs above do not correlate with the hotspots found in the M.A. and the H.A. complexes, as will be described below; the latter suggests that the interface residues that stabilize the L.A. complex is different from that in the M.A. and the H.A. complex. As mentioned before, the rigid body docking of ICAM-1 to the L.A. I-domain resulted in a complex that showed to be unstable after running a MD simulation. As a consequence, an alternative L.A. complex was generated via a Monte Carlo procedure, after which the L.A. αL I-domain underwent some structural rearrangements in order to adapt to the presence of ICAM-1. When this complex was subjected to a MD simulation, both proteins evolved in such a way that the interface generated was different from that of the M.A. complex. Only a severe modification of the interface generated gave rise to a stable low affinity complex, with a reasonable relative binding free energy, but a completely new interface.

Table **5.4**: Low Affinity Complex: Residues in ICAM-1, with a significant contribution to the free energy of binding. The total contribution is split into its components. Results are weighed over ten coordinate sets selected from 3002 structures grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

ICAM-1	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
LYS29	-27.57 (0.63)	29.84 (0.54)	2.27 (0.70)	-0.76 (0.12)	-0.25 (0.01)	-1.01	1.26 (0.82)
LEU30	-0.50 (0.08)	0.12 (0.01)	-0.38 (0.08)	-2.68 (0.19)	-0.41 (0.01)	-3.09	-3.47 (0.28)
GLY32	0.68 (0.10)	0.27 (0.03)	0.95 (0.09)	-0.14 (0.01)	0.00 (0.00)	-0.14	0.81 (0.10)
ILE33	-0.75 (0.10)	0.23 (0.06)	-0.52 (0.05)	-0.31 (0.03)	0.00 (0.00)	-0.31	-0.83 (0.08)
GLU34	-35.71 (1.92)	36.81 (1.36)	1.10 (0.96)	-2.08 (0.26)	-0.41 (0.01)	-2.49	-1.39 (1.23)
PRO36	-0.07 (0.05)	0.41 (0.02)	0.34 (0.04)	-1.64 (0.13)	-0.28 (0.01)	-1.92	-1.58 (0.19)
LEU37	-2.31 (0.34)	1.50 (0.12)	-0.80 (0.23)	-0.55 (0.04)	-0.02 (0.01)	-0.57	-1.37 (0.27)
PRO38	-0.46 (0.08)	0.38 (0.03)	-0.07 (0.10)	-0.60 (0.07)	-0.11 (0.01)	-0.71	-0.78 (0.18)
LYS39	-30.29 (1.04)	36.51 (0.67)	6.22 (0.74)	-1.93 (0.41)	-0.38 (0.00)	-2.31	3.91 (1.15)
LYS40	-1.02 (0.17)	3.86 (0.54)	2.84 (0.70)	-1.02 (0.08)	-0.11 (0.02)	-1.12	1.72 (0.79)
GLU41	-10.72 (0.54)	19.93 (0.69)	9.20 (0.44)	-1.06 (0.06)	-0.20 (0.01)	-1.26	7.94 (0.51)
LEU44	0.28 (0.07)	0.66 (0.07)	0.93 (0.02)	-0.21 (0.01)	-0.03 (0.00)	-0.24	0.69 (0.14)
LYS50	-24.25 (0.88)	18.92 (0.74)	-5.34 (1.11)	1.01 (0.37)	-0.07 (0.00)	0.94	-4.40 (1.48)
TYR52	0.16 (0.21)	2.62 (0.20)	2.78 (0.29)	-0.57 (0.05)	-0.04 (0.01)	-0.61	2.17 (0.35)
MET64	-0.14 (0.03)	0.13 (0.02)	0.00 (0.04)	-1.05 (0.20)	-0.23 (0.02)	-1.27	-1.28 (0.26)
TYR66	-9.87 (0.64)	3.69 (0.19)	-6.18 (0.49)	-0.34 (0.56)	-0.11 (0.01)	-0.45	-6.63 (1.05)
ASN68	-6.40 (0.48)	4.77 (0.18)	-1.64 (0.57)	-1.69 (0.18)	-0.17 (0.01)	-1.86	-3.50 (0.75)
PRO70	-2.26 (0.21)	0.82 (0.06)	-1.44 (0.19)	-0.02 (0.32)	-0.10 (0.01)	-0.13	-1.57 (0.51)
ASP71	-1.13 (0.25)	1.62 (0.11)	0.49 (0.18)	-0.98 (0.05)	-0.08 (0.00)	-1.05	-0.56 (0.23)
THR75	0.70 (0.16)	0.50 (0.09)	1.20 (0.24)	-0.36 (0.02)	-0.10 (0.01)	-0.47	0.73 (0.28)
LYS77	-0.19 (0.02)	1.80 (0.45)	1.60 (0.47)	-0.19 (0.03)	-0.08 (0.01)	-0.27	1.33 (0.51)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar (sum of vdW and SAS components); TOTAL = Total binding free energy (Total Elec. + Total NP)

Table **5.5**: Low Affinity Complex: Residues in the L.A. I-domain (bottom), with a significant contribution to the free energy of binding. The total contribution is split into its components. Results are weighed over ten coordinate sets selected from 3002 structures grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

LFA-1	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
ASP137	-3.84 (1.22)	2.09 (0.55)	-1.74 (0.68)	-0.01 (0.00)	0.00 (0.00)	-0.01	-1.76 (0.68)
MET140	0.25 (0.10)	0.39 (0.05)	0.63 (0.14)	-1.03 (0.16)	-0.29 (0.03)	-1.32	-0.69 (0.33)
SER141	-0.19 (0.09)	0.43 (0.05)	0.25 (0.06)	-0.96 (0.06)	-0.14 (0.00)	-1.10	-0.86 (0.13)
GLU146	-0.95 (0.21)	1.87 (0.28)	0.92 (0.13)	-0.14 (0.01)	-0.05 (0.00)	-0.19	0.73 (0.14)
LEU205	-0.77 (0.24)	0.62 (0.03)	-0.15 (0.26)	-1.54 (0.18)	-0.26 (0.01)	-1.81	-1.96 (0.45)
ASP239	-4.35 (1.42)	3.95 (0.63)	-0.40 (0.82)	-0.32 (0.01)	0.00 (0.00)	-0.32	-0.73 (0.84)
GLU241	-30.09 (1.57)	34.72 (1.30)	4.63 (0.39)	-1.00 (0.46)	-0.42 (0.01)	-1.42	3.21 (0.86)
THR243	-6.35 (0.46)	4.14 (0.08)	-2.21 (0.47)	-2.44 (0.20)	-0.34 (0.01)	-2.79	-5.00 (0.68)
ASP244	-3.50 (0.31)	1.86 (0.08)	-1.63 (0.30)	-1.30 (0.13)	-0.10 (0.00)	-1.40	-3.04 (0.44)
SER245	-8.69 (0.54)	5.02 (0.11)	-3.67 (0.54)	-0.38 (0.40)	-0.29 (0.00)	-0.67	-4.34 (0.94)
LYS263	-4.34(0.33)	4.39 (0.17)	0.05 (0.48)	-3.46 (0.20)	-0.34 (0.01)	-3.80	-3.75 (0.69)
HSE264	-12.07 (0.45)	6.90 (0.08)	-5.17 (0.44)	-1.01 (0.44)	-0.26 (0.01)	-1.27	-6.44 (0.88)
GLN266	-2.19 (0.44)	4.47 (0.16)	2.28 (0.35)	-2.89 (0.09)	-0.41 (0.01)	-3.30	-1.02 (0.45)
THR267	-2.11 (0.38)	3.33 (0.28)	1.22 (0.51)	-2.25 (0.07)	-0.43 (0.01)	-2.68	-1.46 (0.59)
LYS268	0.11 (0.08)	0.91 (0.28)	1.02 (0.35)	-0.35 (0.01)	-0.05 (0.01)	-0.40	0.62 (0.36)
GLU269	-21.63 (0.87)	24.12 (0.58)	2.49 (0.57)	-1.11 (0.37)	-0.48 (0.01)	-1.59	0.90 (0.95)
SER270	0.89 (0.14)	0.63 (0.07)	1.52 (0.16)	-0.90 (0.02)	-0.03 (0.00)	-0.93	0.59 (0.19)
Mg	-28.95 (6.58)	36.77 (9.65)	7.83 (3.09)	-0.01 (0.00)	0.00 (0.00)	-0.01	7.81 (3.09)
TIP02	-6.89 (0.30)	3.15 (0.09)	-3.74 (0.30)	1.26 (0.25)	-0.05 (0.00)	1.20	-2.54 (0.56)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar (sum of vdW and SAS components); TOTAL = Total binding free energy (Total Elec. + Total NP)

Inside-out or outside-in signaling?: The results in the preceeding paragraphs do not imply that the low affinity complex between the αL I-domain and ICAM-1 is only stabilized by a severe modification of the interface, but suggest that the low affinity I-domain as it is published in the PDB (PDB ID=1ZOP), in the presence of ICAM-1, is not able to evolve to a medium affinity or high affinity conformation in the absence of other elements. The wild type I-domain in the context of the intact integrin may undergo these conformational changes upon binding to ICAM-1.

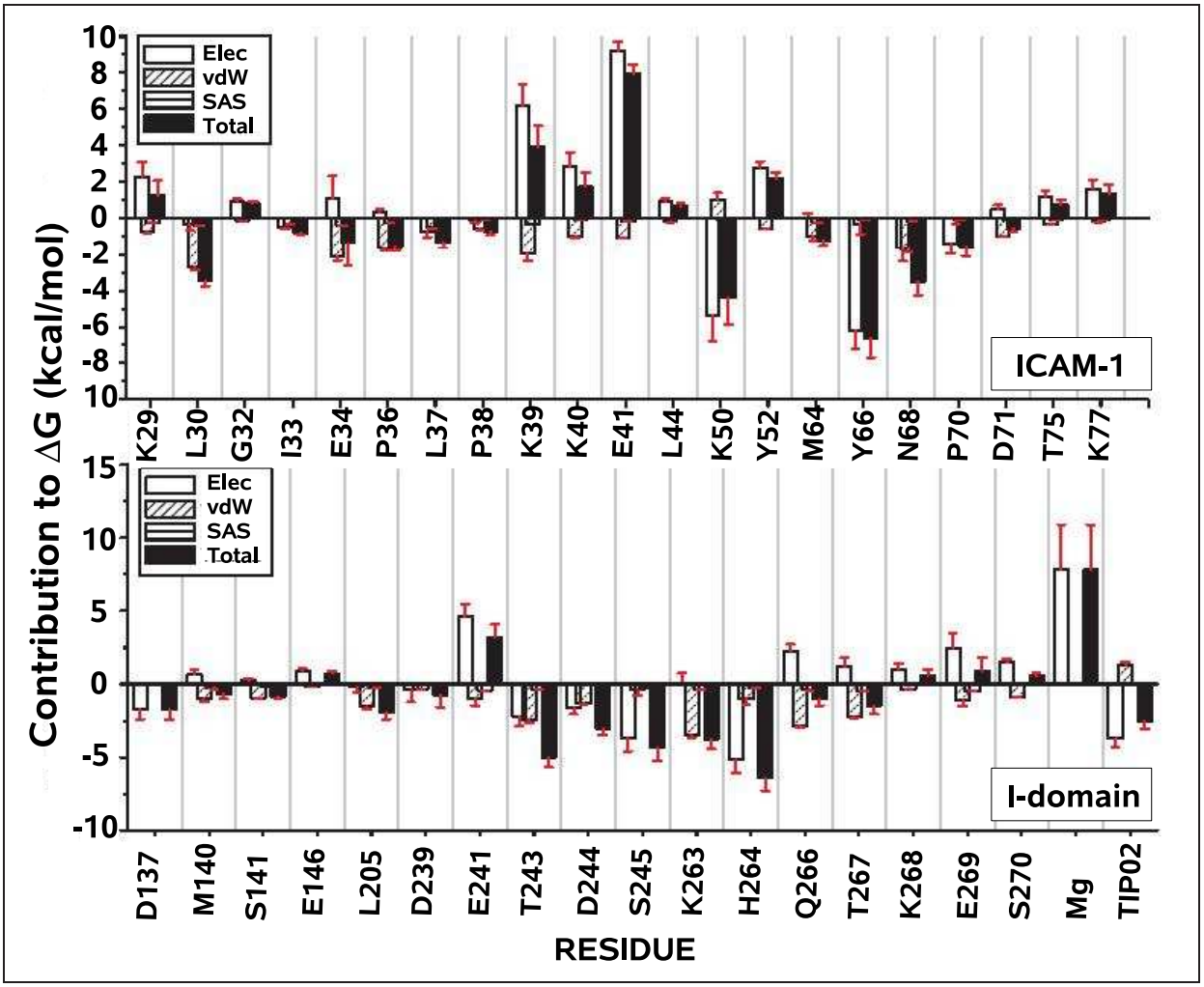


Fig. 5.10: Contribution of each amino acid in ICAM-1 (top) and the I-domain (bottom) to Low Affinity complex formation with LFA-1. Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; TOTAL = Total binding free energy (Total Elec. + Total NP)

o The medium affinity complex: Identifying hotspots

In the following paragraphs the medium affinity complex will be described in a per-residue basis; the description is based on the results listed in Tables 5.6 and 5.7.

Hotspots in ICAM-1 (Table **5.6**): The residues that contribute favorably to the free energy of binding are: L30 (-1.37 kcal/mol), E34 (-28.41 kcal/mol), T35 (-0.63 kcal/mol), P36 (-3.79 kcal/mol), L37 (-0.51 kcal/mol), P38 (-1.15 kcal/mol), K39 (-0.76 kcal/mol), M64 (-2.78 kcal/mol) and Q73 (-0.61 kcal/mol) (Table5.6-top). In the same way, the residues that contribute unfavorably to the free energy of binding of ICAM-1 to the medium affinity I-domain are K29 (1.22 kcal/mol), Q62 (0.61 kcal/mol), Y66 (0.53 kcal/mol) and K77 (2.24 kcal/mol).

Hotspots in the I-domain (Table **5.7**): The residues in the medium affinity I-domain that contribute favorably to the free energy of binding are: M140 (-4.05 kcal/mol), L204 (-0.58 kcal/mol), L205 (-4.62 kcal/mol), T206 (-0.53 kcal/mol), N207 (-1.95 kcal/mol), E241 (-2.30 kcal/mol), T243 (-2.97 kcal/mol), H264 (-0.54 kcal/mol) and the Mg^{2+} (-17.56 kcal/mol) cation. In the same way, the residues in the medium affinity I-domain that contribute unfavorably to the free energy of binding to ICAM-1 are: D137 (1.93 kcal/mol), S139 (0.58 kcal/mol), E146 (1.2 kcal/mol), D239 (1.40 kcal/mol) and the two water molecules coordinated to the metal (2.05 kcal/mol for TIP01 and 0.61 kcal/mol for TIP02, respectively).

Among these residues, E34 in ICAM-1 (-28.41 kcal/mol) and the Mg^{2+} cation in the M.A. I-domain (-17.56 kcal/mol) make the most important interaction.

The residues that form the MIDAS motif, destabilize the binding to ICAM-1: Among the 5 residues of the αL I-domain directly coordinated to the metal (S139, S141, T206 and the two water molecules), three residues display an unfavorable contribution to the binding free energy. These residues are S139 and the two water molecules, which moderately unfavor binding to ICAM-1 (0.58 kcal/mol, 2.05 kcal/mol and 0.61 kcal/mol, respectively). As observed in Table 5.7, the first two have a relatively high electrostatic desolvation penalty value which, together with the slightly unfavorable electrostatic term, destabilize the complex. Residue TIP02, despite being less exposed to the solvent than TIP01 (see the smaller SAS in Table 5.7), still makes significant unfavorable electrostatic contacts. In summary, most of the residues in the M.A. LFA-1 I-domain coordinated to the metal unfavor binding. Still, their contribution is relatively small, suggesting that the natural selection of residues to coordinate the metal in the MIDAS motif had a high affinity for the metal was high, with a minimum disruption of the binding to ICAM-1.

The electrostatic desolvation penalty is the main reason for the unfavorable contributions: For the residues listed in Table 5.7, it is always the electrostatic desolvation penalty the term responsible for the unfavorable total contribution, except for TIP02, with repulsive electrostatic interactions (probably, with E34). The electrostatic contacts found in the M.A. complex may not be efficient enough to counterbalance the high electrostatic desolvation penalty.

Table **5.6**: Medium Affinity Complex: Residues in ICAM-1 with a significant contribution to the free energy of binding. The total contribution is split into its components. Results are weighed over ten coordinate sets selected from 3002 structures grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

ICAM-1	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
LYS 29	-0.18 (0.16)	1.9 (1.15)	1.74(1.15)	-0.44(0.09)	-0.09(0.01)	-0.53	1.22(1.25)
LEU 30	-0.17(0.05)	0.1 (0.05)	-0.12(0.05)	-1.04(0.12)	-0.22(0.01)	-1.26	-1.37(0.18)
GLU 34	-77.50(6.26)	46.8 (2.68)	-30.72(2.68)	2.94(1.60)	-0.64(0.02)	2.30	-28.41(4.30)
THR 35	-0.88(0.4)	1.9 (0.55)	1.01(0.55)	-1.54(0.19)	-0.10(0.01)	-1.64	-0.63(0.76)
PRO 36	-1.38(0.77)	1.3 (0.48)	-0.031(0.48)	-3.24(0.64)	-0.52(0.03)	-3.76	-3.79(1.15)
LEU 37	-0.30(0.33)	0.8 (0.30)	0.47(0.30)	-0.91(0.19)	-0.07(0.02)	-0.98	-0.51(0.51)
PRO 38	-0.35(0.1)	0.2 (0.10)	-0.14(0.10)	-0.88(0.20)	-0.13(0.02)	-1.01	-1.15(0.32)
LYS 39	-13.26(1.98)	12.1 (2.30)	-1.14(2.30)	0.62(1.09)	-0.25(0.02)	0.37	-0.76(4.10)
GLN 62	-0.53 (0.35)	2.3 (0.3798)	1.73(0.38)	-1.00(0.22)	-0.11(0.01)	-1.11	0.61(0.61)
MET 64	-0.17(0.22)	0.6 (0.1752)	0.45(0.17)	-2.93(0.29)	-0.30(0.03)	-3.23	-2.78(0.50)
TYR 66	-0.30(0.08)	3.5 (0.5534)	3.20(0.55)	-2.41(0.40)	-0.25(0.01)	-2.66	0.53(0.96)
GLN 73	-3.17(0.03)	5.0 (1.0973)	1.88(1.10)	-2.05(0.48)	-0.43(0.02)	-2.48	-0.61(1.60)
LYS 77	-1.11(0.06)	3.8 (1.2307)	2.67(1.23)	-0.37(0.03)	-0.06(0.02)	-0.43	2.24(1.28)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar ; sum of vdW and SAS components; TOTAL = Total binding free energy (Total Elec. + Total NP)

Table **5.7**: Medium Affinity Complex: Residues in the M.A. I-domain (bottom), with a significant contribution to the free energy of binding. The total contribution is split into its components. Results are weighed over ten coordinate sets selected from 3002 structures grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

LFA-1	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
ASP 137	0.11 (1.87)	1.87 (0.23)	1.98(0.32)	-0.05(0.00)	0.00(0.00)	-0.05	1.93(0.32)
SER 139	-0.19 (0.01)	0.91 (0.11)	0.72(0.39)	-0.14(0.49)	-0.00(0.00)	-0.14	0.58(0.89)
MET 140	-1.14 (0.01)	2.30 (0.36)	1.16(0.17)	-4.60(0.30)	-0.62(0.02)	-5.22	-4.05(0.50)
GLU 146	-0.63 (0.00)	1.93 (1.85)	1.30(1.20)	-0.07(0.01)	-0.00(0.00)	-0.07	1.2(1.21)
LEU 204	-1.09 (0.02)	2.12 (0.10)	1.03(0.30)	-1.47(0.20)	-0.14(0.03)	-1.61	-0.58(0.53)
LEU 205	-1.22 (0.02)	0.84 (0.08)	-0.38(0.35)	-3.70(0.69)	-0.54(0.03)	-5.24	-4.62(1.07)
THR 206	-3.18 (0.03)	1.59 (0.24)	-1.59(0.89)	1.06(1.10)	-0.01(0.01)	1.05	-0.53(1.98)
ASN 207	-4.13 (0.00)	2.72 (0.14)	-1.41(0.90)	-0.47(0.22)	-0.07(0.00)	-1.54	-1.95(1.11)
ASP 239	-1.67 (8.41)	3.20 (2.74)	1.53(0.38)	-0.14(0.02)	-0.00(0.00)	-0.14	1.40(0.40)
GLU 241	-5.90 (0.00)	8.43 (1.20)	-2.53(1.37)	0.51(1.04)	-0.27(0.01)	0.24	-2.30(2.43)
THR 243	-0.47 (0.01)	2.76 (0.33)	2.29(0.55)	-2.60(0.33)	-0.37(0.01)	-2.97	-0.67(0.89)
HSE 264	-5.90 (0.45)	3.82 (0.59)	2.08(0.70)	-2.25(0.35)	-0.37(0.04)	-2.62	-0.54(1.09)
MG	-141.56 (0.01)	119.38 (10.94)	-22.19(5.21)	4.63(0.90)	-0.00(0.00)	4.63	-17.56(6.10)
TIP01	0.48 (0.13)	1.31 (0.73)	1.80(0.86)	0.30(0.24)	-0.03(0.01)	0.27	2.05(1.10)
TIP02	0.59 (0.00)	0.14 (0.02)	0.73(0.13)	-0.12(0.01)	0.00(0.00)	-0.12	0.61(0.14)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar ; sum of vdW and SAS components; TOTAL = Total binding free energy (Total Elec. + Total NP)

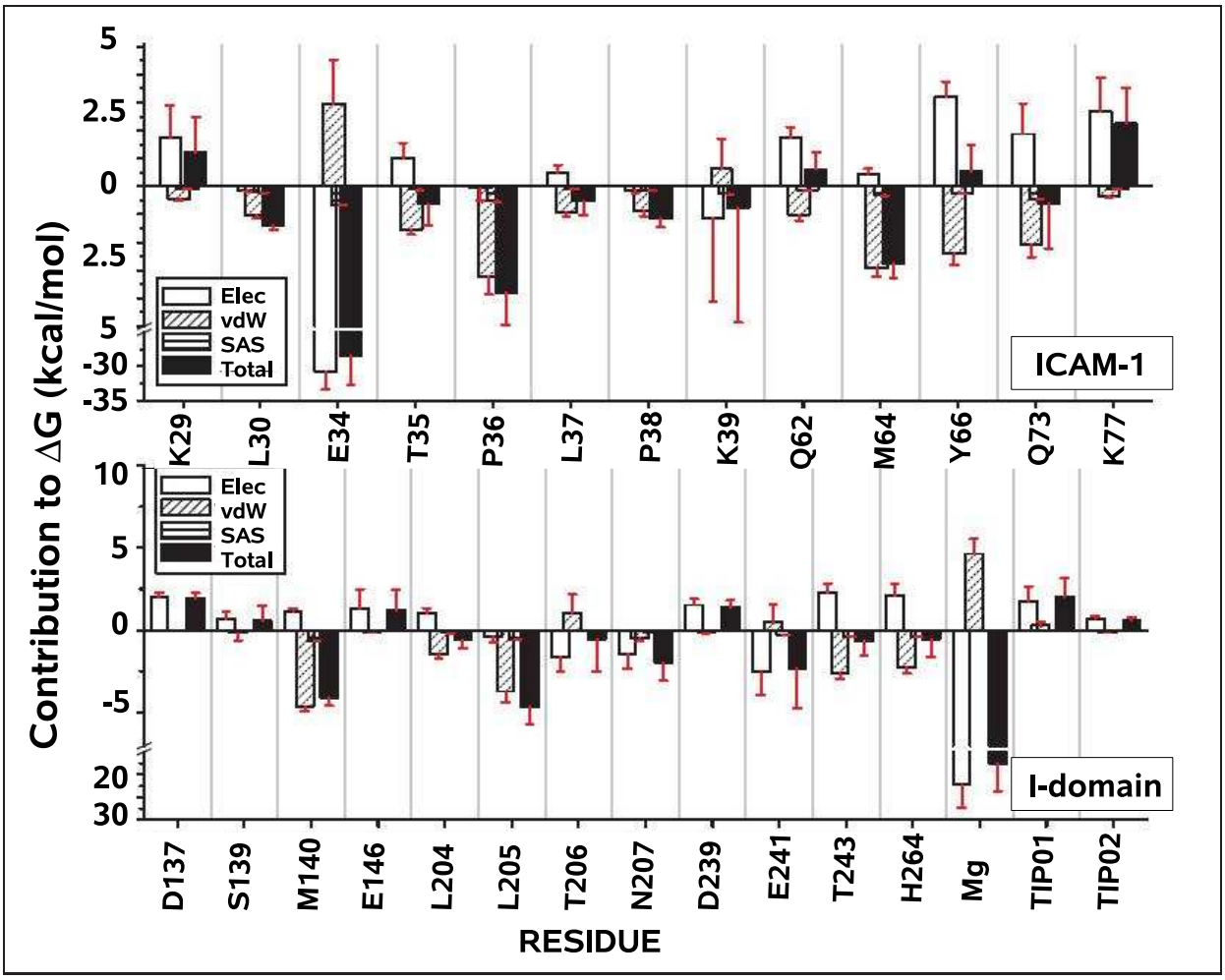


Fig. 5.11: Contribution of each amino acid in ICAM-1 to Medium Affinity complex formation with LFA-1. Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; TOTAL = Total binding free energy (Total Elec. + Total NP)

High number of effective vdW interactions: Among the residues that favor binding, only E34 and K39 in ICAM-1 are electrostatic in nature. The former, with -28.41 kcal/mol is the most effective residue in ICAM-1 followed, from far behind, by K39, with a total contribution of -0.76 kcal/mol, which is mainly due to favorable electrostatic contacts. The rest of the residues in ICAM-1 favor binding via effective vdW interactions with the I-domain, especially P36, with a total contribution of -3.79 kcal/mol, coming mainly from vdW interactions. In the M.A. I-domain, T206, N207, E241 and the Mg^{2+} cation favor binding by interactions that are mainly electrostatic in nature, and M140, L204, L205, T243 and H264, with interactions of vdW nature. Though the magnitude of the electrostatic interactions (without desolvation) is much larger than the magnitude of the vdW interactions, the large number of the latter makes it possible that the sum of the vdW interactions in the M.A. complex (-38.27 kcal/mol) be more important than the total (interaction+desolvation) electrostatic interactions (-25.74 kcal/mol).

o The high affinity complex: Identifying hotspots

In the high affinity complex formation, there are seven residues in ICAM-1 and ten in the H.A. I-domain that favor the binding: L30, E34, T35, P36, M64, N68, and Q73 in ICAM-1 and M140, S141, Q143, L204, L205, T206, N207, E241, T243 and the Mg^{2+} cation in the H.A. I-domain. The sum of their contribution is as high as -41.42 and -31.35 kcal/mol, respectively. On the other hand, a single residue in ICAM-1 (K77) and 5 residues in the H.A. I-domain (D137, S139, D239, S245 and TIP 309) contribute unfavorably to the total free energy of binding, which sum up 2.39 kcal/mol and 5.65 kcal/mol, respectively. Among these residues, E34 in ICAM-1 (-31.31kcal/mol) and the Mg^{2+} cation in the H.A. I-domain, (-10.76 kcal/mol), form the most important interaction.

The residues that form the MIDAS motif disrupt binding: It is interesting to note again that two residues in the M.A. I-domain directly coordinated to the metal showed an unfavorable contribution to the binding free energy. These residues are S139 and, more importantly, TIP02, which disfavor binding to ICAM-1 with up to 0.55 and 2.11 kcal/mol, respectively. Both residues show a relatively high electrostatic desolvation penalty value and, in the case of TIP02, unfavorable vdW interactions. In summary, some of the residues that constitute the MIDAS motif disrupt binding to ICAM-1, but yet, less than in the case of the M.A. complex.

The electrostatic desolvation penalty is the main responsible for the unfavorable contributions: In the H.A. complex, among the unfavorable residues listed in Table **5.8**, it is always the electrostatic desolvation penalty that they pay which is responsible for their unfavorable total contribution to the free energy of binding. The most clear example is the case of K77, that makes both significant electrostatic (-2.28 kcal/mol) and significant vdW contacts (-0.77 kcal/mol), but the high electrostatic desolvation penalty paid (5.49 kcal/mol), makes its total contribution unfavorable (2.39 kcal/mol).

Table **5.8**: High Affinity Complex: Residues in ICAM-1 with a significant contribution to the free energy of binding. The total contribution is split into its components. Results are weighed over ten coordinate sets selected from 240 structures grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

ICAM-1	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
LEU 30	-0.10 (0.11)	0.06 (0.02)	-0.04 (0.10)	-0.84 (0.29)	-0.17 (0.04)	-1.01	-1.05 (0.44)
GLU 34	-79.31 (0.21)	45.76(0.12)	-33.55 (0.13)	2.82 (0.03)	-0.58 (0.00)	2.23	-31.31 (3.58)
THR 35	-0.74 (2.70)	1.02 (1.42)	0.28 (2.31)	-1.04 (1.27)	-0.08 (0.01)	-1.13	-0.84 (0.22)
PRO 36	-0.43 (0.24)	1.17 (0.28)	0.74 (0.14)	-2.60 (0.06)	-0.43 (0.01)	-3.03	-2.30 (1.07)
MET 64	-0.47 (0.12)	0.41 (0.03)	-0.06 (0.11)	-3.14 (0.15)	-0.27 (0.02)	-3.41	-3.47 (0.28)
ASN 68	-2.18 (0.46)	2.17 (0.27)	0.00 (0.35)	-0.64 (0.32)	-0.09 (0.01)	-0.73	-0.74 (0.68)
GLN 73	-3.84 (0.70)	4.57 (0.37)	0.73 (0.43)	-2.02 (0.36)	-0.41 (0.03)	-2.43	-1.71 (0.82)
LYS 77	-2.28 (0.40)	5.49 (0.75)	3.22 (0.78)	-0.77 (0.07)	-0.06 (0.01)	-0.83	2.39 (0.86)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar (sum of vdW and SAS components); TOTAL = Total binding free energy (Total Elec. + Total NP)

Table **5.9**: High Affinity Complex: Residues in the H.A. with a significant contribution to the free energy of binding. The total contribution is split into its components. Results are weighed over ten coordinate sets selected from 240 structures grouped in ten families according to their gas phase energy. In parentheses, the standard error. Values are given in kcal/mol.

LFA-1	Elec. Int.	Desolv.	Total Elec.	vdW.	SAS	Total NP	TOTAL
ASP 137	-0.23 (0.57)	1.95(0.12)	1.71 (0.60)	-0.05 (0.00)	0.00 (0.00)	-0.05	1.66 (0.61)
SER 139	-0.10 (0.30)	0.77(0.11)	0.67 (0.25)	-0.11 (0.24)	0.00 (0.00)	-0.11	0.56 (0.49)
MET 140	-0.82 (0.34)	2.10(0.13)	1.28 (0.37)	-5.22 (0.51)	-0.62 (0.02)	-5.84	-4.55 (0.90)
SER 141	-1.43 (0.50)	2.48(0.12)	1.05 (0.60)	-1.48 (0.40)	-0.21 (0.02)	-1.69	-0.64(1.01)
GLN 143	-0.21 (0.41)	0.59(0.42)	0.37 (0.19)	-0.81 (0.21)	-0.17 (0.02)	-0.98	-0.61(0.43)
LEU 204	-1.23 (0.27)	1.91(0.35)	0.68 (0.43)	-1.70 (0.17)	-0.16 (0.03)	-1.86	-1.18(0.63)
LEU 205	-1.21 (0.45)	0.76(0.07)	-0.45(0.42)	-3.88 (0.82)	-0.49 (0.03)	-4.37	-4.82(1.27)
THR 206	-3.61 (0.74)	1.55(0.17)	-2.06(0.63)	0.87 (0.62)	-0.01 (0.00)	0.87	-1.19(1.25)
ASN 207	-4.00 (0.52)	2.87(0.21)	-1.12(0.39)	-0.33 (0.31)	-0.07 (0.01)	-0.41	-1.53(0.70)
ASP 239	-0.70 (0.50)	1.75(0.40)	1.04 (0.51)	-0.21 (0.05)	-0.01 (0.01)	-0.23	0.82 (0.57)
GLU 241	-11.83 (0.71)	7.44(0.60)	-4.40(0.71)	1.09 (1.01)	-0.28 (0.01)	0.80	-3.59(1.73)
THR 243	-3.00 (0.61)	3.21(0.13)	0.21 (0.66)	-2.50 (0.36)	-0.35 (0.02)	-2.85	-2.64(1.04)
SER 245	-0.05 (0.26)	1.51(0.47)	1.46 (0.37)	-0.71 (0.11)	-0.25 (0.02)	-0.96	0.50 (0.50)
MG	-138.22 (5.82)	123.03(7.31)	-15.19(5.65)	4.43 (0.72)	0.00 (0.00)	4.43	-10.76(6.37)
TIP02	0.32 (0.46)	1.43(0.21)	1.75 (0.44)	0.41 (0.19)	-0.05 (0.01)	0.36	2.11 (0.64)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar (sum of vdW and SAS components); TOTAL = Total binding free energy (Total Elec. + Total NP)

High number of effective vdW interactions: Among the residues that significantly favor binding, only E34 (-31.31 kcal/mol) in ICAM-1 is electrostatic in nature. The rest of the residues in ICAM-1 (L30, T35, P36, M64, N68 and Q73), favor binding via effective vdW interactions with the I-domain, specially M64, with a total contribution of -3.47 kcal/mol, which mainly comes from vdW interactions (-3.14 kcal/mol). In the H.A. I-domain, T206, N207, E241 and the Mg^{2+} cation favor binding by interactions of electrostatic nature, and M140, S141, Q143, L204, L205 and T243, with interactions of vdW nature. Though the magnitude of the individual electrostatic interactions is much larger than the magnitude of the vdW interactions, the large number of the latter, makes of the vdW interactions in the H.A. complex (-33.92 kcal/mol) the most important component, being the electrostatic interactions of lower value (-33.86 kcal/mol). The complex is highly electrostatic in nature, but the high electrostatic desolvation penalty to pay makes the vdW interactions nearly as important as the electrostatic contacts.

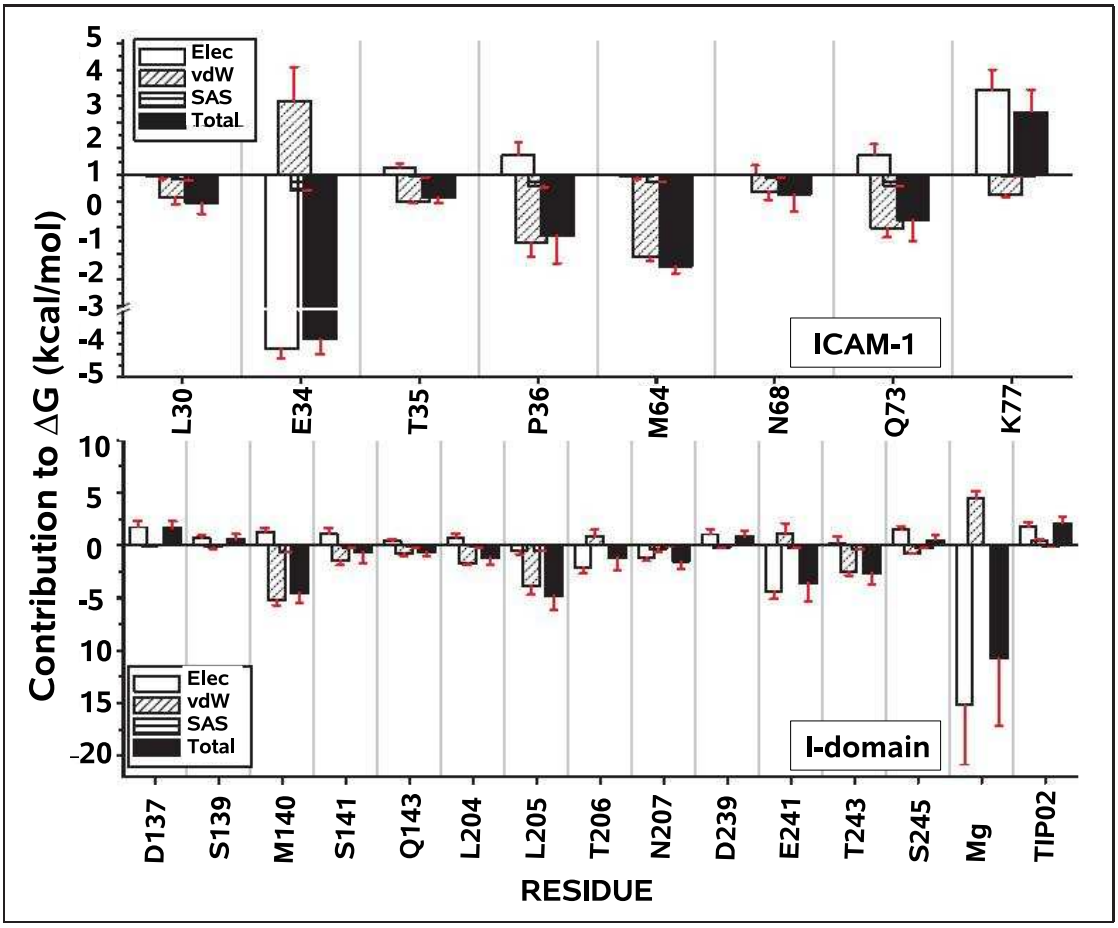


Fig. 5.12: Contribution of each amino acid in ICAM-1 to High Affinity complex formation with LFA-1. Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; TOTAL = Total binding free energy (Total Elec. + Total NP)

5.2.3 Amino acid efficiency and ligand binding

A different, instructive way to identify amino acids that have an important role in binding is to compute the amino acid efficiency, a quantity defined here in analogy to ligand efficiency. Ligand efficiency is defined as the ratio of affinity to molecular size, a common measure of size being the number of non-hydrogen atoms. The notion of ligand efficiency has recently attracted a considerable interest in drug design studies [254, 255], particularly in the context of fragment based drug design. In fragment based drug design, the goal is to find small molecular building blocks that bind to a medicinal target. Subsequently, several building blocks can be linked or iteratively increased in size while keeping the molecular weight in the druggable range (Lipinsky limit) [256]. Small ligands with high binding efficiency are thus sought as a good starting point for lead identification and drug design. By analogy with this concept, efficient amino acids were identified in the binding interface between the low affinity, medium affinity and high affinity I-domains and ICAM-1, by dividing the free energy contribution of each amino acid by its number of non-hydrogen atoms (i.e. 4 for Gly, 5 for Ala, 14 for Trp etc.). Amino acid efficiency can provide an assessment of the binding potential of a particular hotspot interaction. In drug design strategies aimed at protein-protein interfaces, the identification of efficient amino acids could help identify key interactions and prioritize the region of the protein-protein interface that could be targeted by small molecules. In order to gain physical insight in the origins of the amino acid efficiency, each individual contribution to the binding free energy was divided, as well as the total.

Low affinity complex: The most efficient amino acids in the stabilization of the low affinity complex are listed in Table **5.10**: L30, K50, Y66 and N68 in ICAM-1 and T243, S245, K263 and H264 in the L.A. I-domain.

Medium Affinity complex: The most efficient amino acids in the stabilization of the medium affinity complex are listed in Table **5.11**: E34 and P36 in ICAM-1 and M140, L205 and the Mg^{2+} cation in the M.A. I-domain.

High Affinity complex: The most efficient amino acids in the stabilization of the high affinity complex are listed in Table **5.12**: E34 and M64 in ICAM-1 and M140, L205 and the Mg^{2+} cation in the H.A. I-domain.

As observed, again, the residues that take part in the formation of the low affinity complex are different to those contributing to the formation of the medium or high affinity complexes. In general, the absolute value of the efficiency of the amino acids on each of the three I-domains is better than the efficiency of residues in ICAM-1. If Tables **5.11** and **5.12** are compared for the M.A. and H.A. complexes, it is observed that the identity of the most efficient amino acids are the same in many cases. In most of the cases, the van der Waals contribution is the term that dominates over the rest, except for E34 in ICAM-1 and the Mg^{2+} cation in the I-domain, with a dominant electrostatic term.

One of the methodologies followed in drug design studies, consists of mimicking the electronic and geometric properties of the natural ligand of the target molecule. Among the properties of a drug candidate, its size and molecular weight are of key importance for its Absorption, Distribution, Metabolism and Excretion (ADME hereafter) characteristics. In this sense, a molecular weight lower than 500MW is desirable [256]. To design a drug candidate of small molecular weight but good binding properties, it is interesting to identify those residues in the

natural ligand that most efficiently bind to the target molecule. Upon identification, an effort should be made to generate their synthetic equivalents.

Out of the efficiency results obtained from the per-residue decomposition analysis of the free energy of binding, it is concluded that a good drug candidate that binds to the high affinity αL I-domain should contain, at least, a carboxylate group mimicking E34 in ICAM-1. In addition, this group should be surrounded by apolar groups that would mimic the aliphatic chains of residues M64, P36, Q73 and T35.

Table **5.10**: Residues with a significant efficiency in ICAM-1 and the low affinity αL I-domain, in the LA. Complex

	Total Elec.	SAS	vdW	TOTAL
Low Affinity Complex				
ICAM-1				
LEU 30	-0.05	-0.05	-0.33	-0.43
ILE 33	-0.06	0.00	-0.04	-0.10
GLU 34	0.12	-0.05	-0.23	-0.15
PRO 36	0.05	-0.04	-0.23	-0.23
LEU 37	-0.10	0.00	-0.07	-0.17
PRO 38	-0.01	-0.02	-0.09	-0.11
LYS 50	-0.59	-0.01	0.11	-0.49
MET 64	0.00	-0.03	-0.13	-0.16
TYR 66	-0.51	-0.01	-0.03	-0.55
ASN 68	-0.20	-0.02	-0.21	-0.44
PRO 70	-0.21	-0.01	0.00	-0.22
I-domain				
ASP 137	-0.22	0.00	0.00	-0.22
SER 141	0.04	-0.02	-0.16	-0.14
LEU 205	-0.02	-0.03	-0.19	-0.25
THR 243	-0.32	-0.05	-0.35	-0.71
ASP 244	-0.20	-0.01	-0.16	-0.38
SER 245	-0.61	-0.05	-0.06	-0.72
LYS 263	0.01	-0.04	-0.38	-0.42
HSE 264	-0.52	-0.03	-0.10	-0.64
THR 267	0.17	-0.06	-0.32	-0.21
TIP02	-0.53	-0.01	0.18	-0.36

Total Elec = Total Electrostatic term (Interaction + Desolvation); SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; TOTAL = Total binding free energy; vdW = vdW Interaction

Table **5.11**: Residues with a significant efficiency in ICAM-1 and the medium affinity αL I-domain, in the M.A. Complex

	Total Elec.	SAS	vdW	TOTAL
Medium Affinity Complex				
ICAM-1				
LEU 30	-0.01	-0.03	-0.13	-0.17
GLU 34	-3.41	-0.07	0.33	-3.16
PRO 36	0.00	-0.07	-0.46	-0.54
PRO 38	-0.02	-0.02	-0.13	-0.16
MET 64	0.06	-0.04	-0.37	-0.35
I-domain				
MET 140	0.15	-0.08	-0.58	-0.51
LEU 205	-0.05	-0.07	-0.46	-0.58
ARG 207	-0.18	-0.01	-0.06	-0.24
GLU 241	-0.28	-0.03	0.06	-0.26
THR 243	0.33	-0.05	-0.37	-0.10
Mg	-22.19	0.00	4.63	-17.56

Total Elec = Total Electrostatic term (Interaction + Desolvation); SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; TOTAL = Total binding free energy; vdW = vdW Interaction

Table **5.12**: Residues with a significant efficiency in ICAM-1 and the high affinity αL I-domain, in the H.A. Complex

	Total Elec.	SAS	vdW	TOTAL
High Affinity Complex				
ICAM-1				
GLU 34	-3.73	-0.06	0.31	-3.48
THR 35	0.04	-0.01	-0.15	-0.12
PRO 36	0.11	-0.06	-0.37	-0.33
MET 64	-0.01	-0.03	-0.39	-0.43
GLN 73	0.08	-0.05	-0.22	-0.19
I-domain				
MET 140	0.16	-0.08	-0.65	-0.57
SER 141	0.18	-0.04	-0.25	-0.11
LEU 204	0.09	-0.02	-0.21	-0.15
LEU 205	-0.06	-0.06	-0.48	-0.60
LEU 206	-0.29	0.00	0.12	-0.17
ASN 207	-0.14	-0.01	-0.04	-0.19
GLU 241	-0.49	-0.03	0.12	-0.40
THR 243	0.03	-0.05	-0.36	-0.38
Mg	-15.19	0.00	4.43	-10.76

Total Elec = Total Electrostatic term (Interaction + Desolvation); SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; TOTAL = Total binding free energy; vdW = vdW Interaction

5.3 Binding of Small Antagonists to the αL I-domain

As a first approach to study the ability to bind to the MIDAS motif of compounds designed in our research group, one of the in-house LFA-1 antagonists was docked to this site at the top of the αL I-domain. MD simulations were run and a free energy decomposition analysis performed. The results presented here guided the process of designing a second family of LFA-1 antagonists, which should, according to the below presented predictions, display a stronger binding to the MIDAS motif.

5.3.1 Choice and parametrization of the LFA-1 antagonist

Among the LFA-1 antagonists synthesized by previous members of the group, compound HC0303 (Figure 5.13) was chosen as the one to be docked to the αL I-domain. This compound had previously been subjected to *in vitro* and *in vivo* experiments in the laboratories of the company Pharmakine Ltd., experiments that will be summarized later in this thesis book (see Chapter 6). At this point it will be just mentioned that compound HC0303 was shown to be an excellent inhibitor of the LFA-1 dependent cell adhesion, as well as to importantly reduce the number and size of metastases in a colon cancer murine model. As this compound had been designed mimicking the electronic and geometric features of key residues in ICAM-1, it is expected that it exerts its inhibitory activity by binding to the MIDAS motif, as ICAM-1 does. Therefore, this compound was docked to the MIDAS motif of the high affinity αL I-domain, MD simulations were run and the interactions that were established were analyzed, as before, following the MM/PBSA protocol and decomposing the free energy of binding in a per-residue basis.

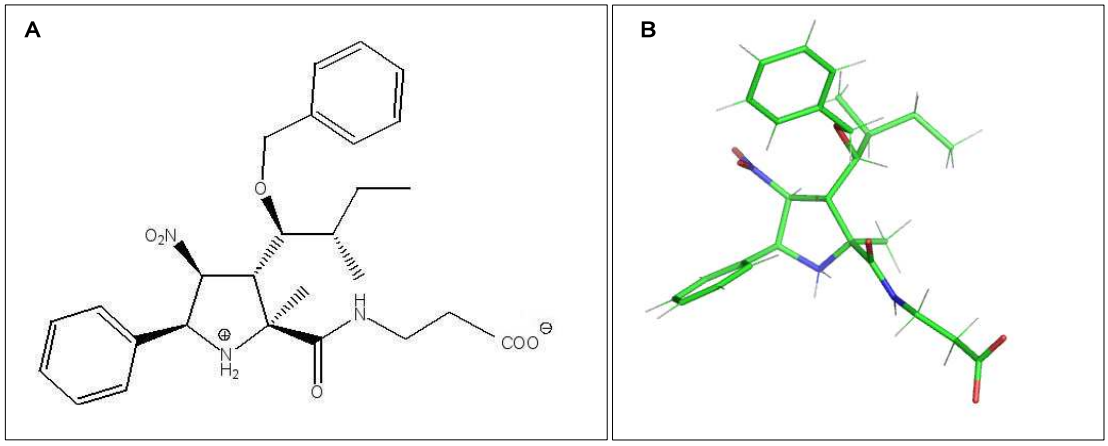


Fig. 5.13: A) Schematic diagram of compound HC0303 and B) stick representation of the minimized structure of compound HC0303.

Parametrization of HC0303: Compound HC0303 basically consists of a pyrrolidine ring with different type of substituents. Among the latter phenyl groups are found, as well as carboxylate groups, nitro groups and aliphatic carbons. The parameters for compound HC0303 are not included in the CHARMM force field. Nevertheless, as similar chemical groups can be found on the force field, atom types were assigned straightforwardly and charges derived by analogy with similar chemical groups present in the CHARMM topology files. The partial charge on each atom (see atoms in Figure 5.14) are found in Table **5.13**. A "software-based" parametrization is still required to properly characterize the compound. These will be, therefore, preliminary results on the "in silico" behavior of the most promising LFA-1 antagonist designed in-house, with respect to its binding to the MIDAS motif of the high affinity αL I-domain.

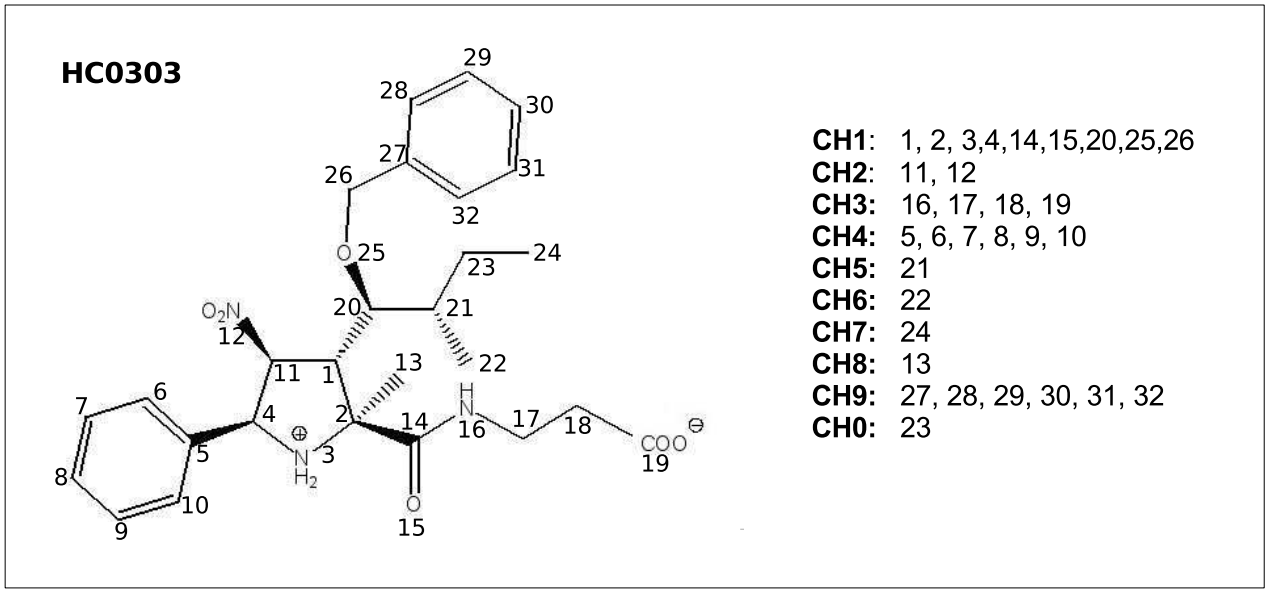


Fig. 5.14: Left: Diagram of the LFA-1 antagonist HC0303, where groups of atoms are numbered from 1 to 32. Right: atom names of the atoms that belong to each group of atoms numbered.

5.3.2 MD simulation of the high affinity αL I-domain complexed with HC0303

Docking of compound HC0303 to the high affinity αL I-domain: To generate the structure of the complex between HC0303 and the H.A. αL I-domain a Monte Carlo procedure was followed. Compound HC0303 was drawn towards the I-domain, being the latter free to adapt to the approaching compound via conformational rearrangements of backbone and sidechain atoms. A virtual axis was defined between the Mg^{2+} cation in the high affinity I-domain structure and atom C30 in the carboxylate group in compound HC0303 (see Figure 5.14). Along this axis was guided the displacement of the ligand towards the I-domain during the Monte Carlo procedure. The initial distance between the two edges of the axis was set to 8 $\mathring{A}$. The displacement towards the MIDAS motif of compound HC0303 (referenced in atom C30) after each one of the 6 loops that were run during the docking procedure was set to 1$\mathring{A}$. Each loop was 5000 steps long (1 femtosecond/step), during which the ligand could be translated a maximum of 1 $\mathring{A}$ and rotated a maximum of 10 degrees. In addition, the maximum change allowed to happen in torsion angles in the ligand was set to 20 degrees. After every rotation, translation or torsion movement, the system was minimized during 20, 20 and 10 steps, respectively, using the steepest descent algorithm. Simultaneously, all the atoms in the I-domain located at a maximum distance of 12 $\mathring{A}$ from any atom in compound

Table **5.13**: Parameters for compound HC0303. In each column, the atom names, atom types and atom partial charges are found.

Name	Type	Charge	Name	Type	Charge	Name	Type	Charge
C01	CP2	-0.10	H01	HA	0.10	C02	CP1	0.42
NP3	NP	-0.56	HP1	HC	0.36	HP2	HC	0.36
C04	CP3	0.32	H04	HA	0.10	C23	C	0.38
O25	O	-0.38	NP0	NX	0.23	O	OC	-0.17
O62	OC	-0.17	C05	CP2	0.01	H05	HA	0.10
N24	NH1	-0.50	H24	H	0.28	C27	CT2	0.02
H71	HB	0.10	H72	HB	0.10	C28	CT2	-0.28
H81	HA	0.09	H82	HA	0.09	C30	CC	0.62
OM1	OC	-0.76	O32	OC	-0.76	C12	CA	0.00
C13	CA	-0.10	H13	HP	0.10	C14	CA	-0.10
H14	HP	0.10	C15	CA	-0.10	H15	HP	0.10
C16	CA	-0.10	H16	HP	0.10	C17	CA	-0.10
H17	HP	0.10	C33	CN8	0.05	H33	HN8	0.10
O45	ON6	-0.30	C46	CN8	-0.05	H61	HN8	0.10
H62	HN8	0.10	C35	CT1	-0.10	H35	HA	0.10
C41	CT3	-0.30	H41	HA	0.10	H42	HA	0.10
H43	HA	0.10	C37	CT3	-0.30	H30	HA	0.10
H31	HA	0.10	H32	HA	0.10	C20	CT3	-0.30
H21	HA	0.10	H22	HA	0.10	H23	HA	0.10
C49	CA	0.00	C50	CA	0.10	HF1	HP	-0.10
C51	CA	-0.10	H51	HP	0.10	C52	CA	-0.10
H52	HP	0.10	C53	CA	-0.10	H53	HP	0.10
C54	CA	0.10	HF2	HP	-0.10	CEX	CT2	-0.20
HEX	HA	0.10	HES	HA	0.10			

HC0303 were allowed to rearrange and adapt to the presence of the latter. At the end of the sixth loop all the constrains in the system were released and a short simulation was run for 2000 steps. In Figure 5.15 shown is the structure of the complex generated after the Monte Carlo docking procedure, which was used as starting geometry in the MD simulation run.

Rearrangements during the MD simulation: From the MD simulations of compound HC0303 bound to the H.A. αL I-domain, the rearrangements that occur are characterized in relation to the relative position of both counterparts. First, in Figure 5.16 shown is the time series of the RMS deviation values of the backbone atoms of the αL I-domain, of the heavy atoms in compound HC0303, as well as the time-series of the RMSD values of backbone and heavy atoms, respectively, of both the I-domain and the compound. The latter, informs about the global reorientation of one partner with respect to the other. As observed in Figure 5.16, the MD simulation was 2.5 nsec long and the RMSD of the I-domain stabilized early in the dynamics. The RMSD value of compound HC0303 during the trajectory is larger and less stable than for the backbone atoms of the I-domain. This is due to the flexibility of the substituents of the pyrrol ring in the compound. As observed (dots in Figure 5.16), the relative orientation of compound HC0303 with respect to the high affinity I-domain also varies a bit, as the RMSD values of LFA-1 or compound HC0303, separately, are shorter than the ones for all the backbone atoms in the system.

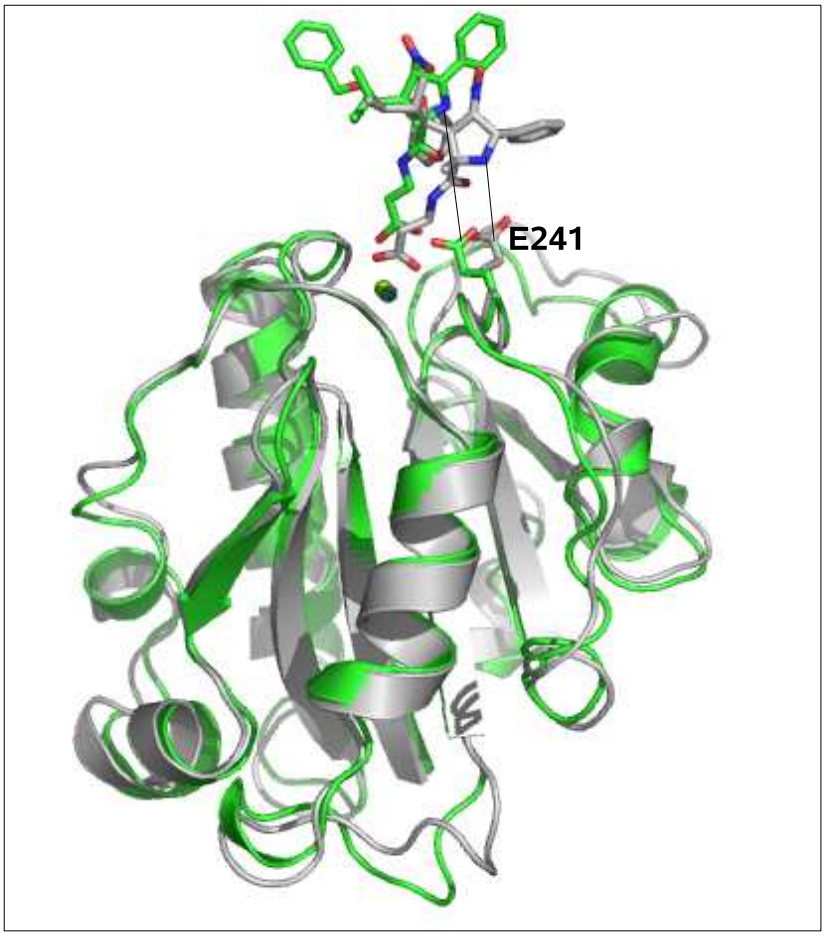


Fig. 5.15: Cartoon representation of the H.A. I-domain-HC0303 complex, before (green) and after (white) the MD simulation. Solid line: desired interaction between the CD atom of E241 in the I-domain and the NH_2^+ group in the pirrol ring in HC0303.

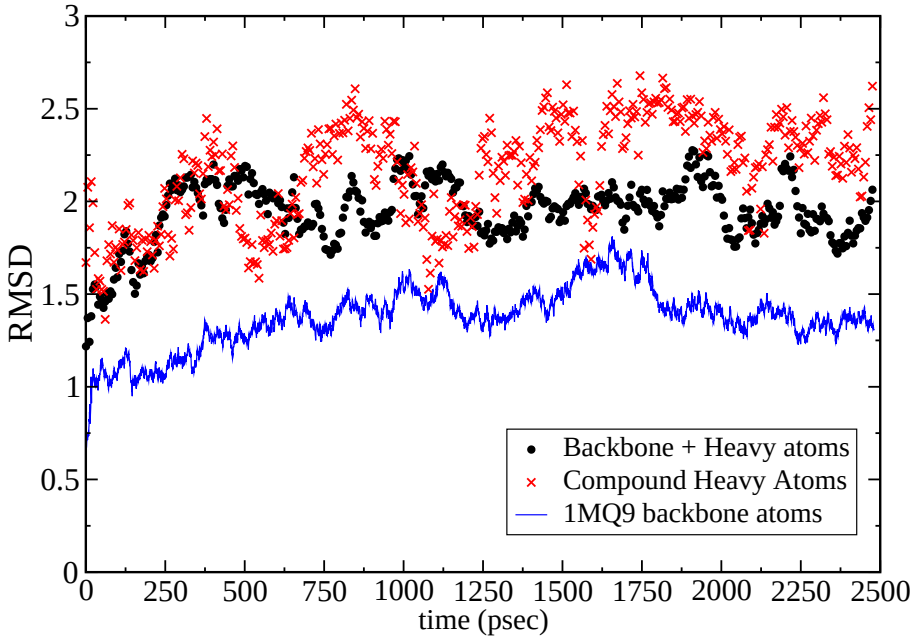


Fig. 5.16: Time series of the RMSD values of backbone or heavy atoms of the complex between 1MQ9 and HC3030 (dots), of backbone atoms of the H.A. αL I-domain (solid line) or heavy atoms in HC3030 (crosses).

Key interactions: In order to easier compare aminoacids in ICAM-1 to groups of atoms in compound HC0303, the latter was subdivided in several groups/residues, depicted in Figure 5.17. The geometric and chemical properties of compound HC0303 were selected in order to try to mimic, at least, a couple of those key interactions observed in the complex of the I-domain with its natural ligand ICAM-1. Such interactions are i) the coordination of E34 in ICAM-1 to the Mg^{2+} cation and ii) the salt bridge between K39 in ICAM-1 and E241 in the I-domain. In compound HC0303, group CH3 would mimic E34 and the positive charge in the pirrol ring in group CH1 would mimic K39.

The presence or absence of these two interactions could first be established by a visual inspection of the complex. As observed in Figure 5.15, while the interaction with the metal cation seems to be established, the distance between the $NH2^{+}$ group in HC0303 and E241 seems to be too large as to make a salt bridge. For clarity, in Figure 5.18 shown are the time series of the distances between these two atom groups during the MD simulation. The distance of atom C30 and the Mg^{2+} cation is around 4 $\mathring{A}$, and is kept quite constant during the whole trajectory. On the contrary, the distance between atom NP3 in HC0303 and atom CD in residue E241 of the I-domain fluctuates between 9 and 13 $\mathring{A}$. Despite the long-range nature of the electrostatic contacts, it can already be assumed that there is no salt bridge between these two groups of atoms.

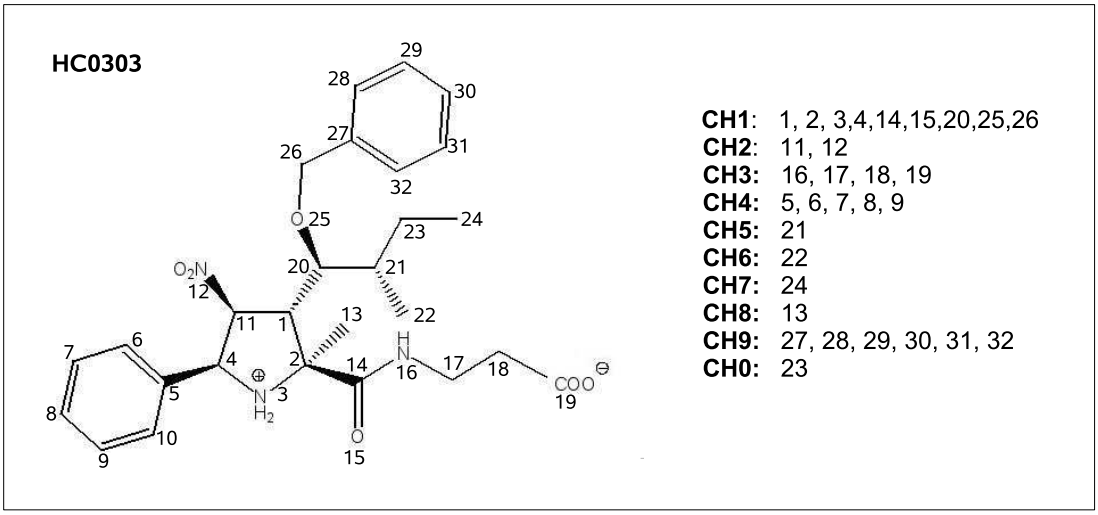


Fig. 5.17: Left: Diagram of the LFA-1 antagonist HC0303, where groups of atoms are numbered from 1 to 32. Right: Residue names and atom groups that belong to each residue defined.

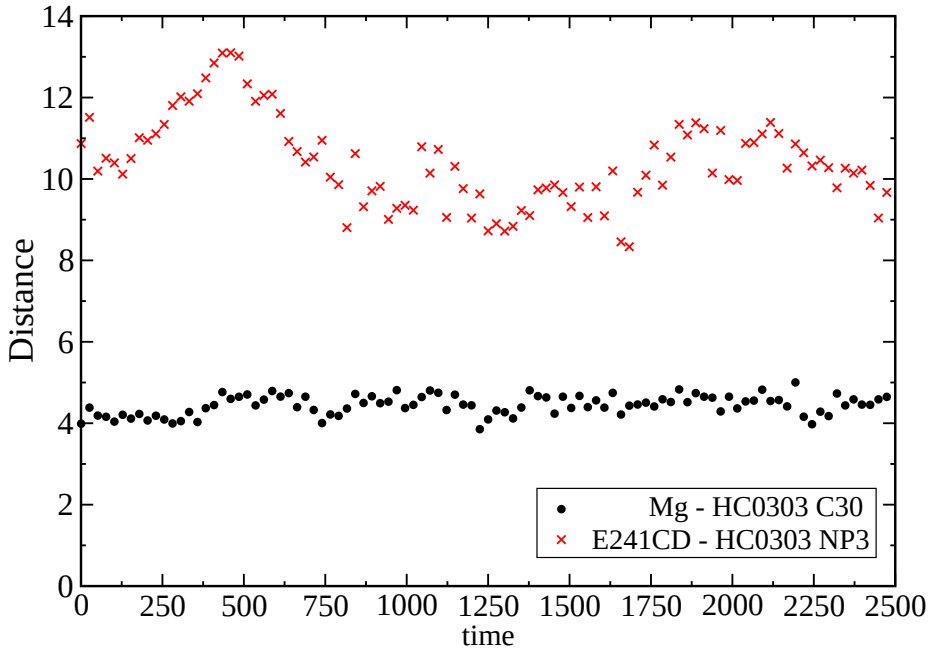


Fig. 5.18: Time (in psec) series of the distances (in $\mathring{A}$) between i) the NP3 atom of group CH3 in compound HC0303 and CD atom of residue E241 in the H.A. I-domain (crosses) and ii) C30 atom of group CH3 in compound HC0303 and the Mg^{2+} cation (dots).

In summary, it is concluded that only one of the two expected and most important electrostatic contacts is taking place between compound HC0303 and the High Affinity I-domain. Nevertheless, the complex is stable during more than 2 nsec of MD simulation. What type of non-obvious contacts are stablished so that the complex is not broken? What is the affinity of the H.A. I-domain towards compound HC0303? These and other questions were addressed by performing a free energy analysis and its per-residue decomposition. The results of such an analysis are shown in the following paragraphs.

5.3.3 Total binding free energy

During the MD simulation, the complex between compound HC0303 and the H.A. I-domain was stable, which is a first clue that suggests that our compound may exert its *in vivo* inhibitory activity by binding to the MIDAS motif. In other words it can be said that the computational results obtained in the present study do not contradict that hypothesis. An additional key piece of information would be that of the affinity or strength of the binding. A comparison of this data to the affinity results obtained with ICAM-1 could help understand some aspects of the possible competitiveness of HC0303 and ICAM-1. With this goal in mind, a free energy analysis was performed and the results decomposed in a per-residue basis, as before, in order to estimate the magnitude and nature of this promising interaction and to localize the most important hotspots.

In Table **5.14** shown is the total free energy value of compound HC0303 binding to the high affinity I-domain (averaged over a number of hotspots extracted from the MD simulation). The total free energy of HC0303 binding to the high affinity αL I-domain was found to be -22.23 kcal/mol. Comparing to data in Table **5.2** it is observed that compound HC0303 binding to the H.A. I-domain is 40 kcal/mol less favorable than ICAM-1 binding. In the same way, ICAM-1 binding to the L.A. I-domain is more favorable than HC0303 binding to the H.A. I-domain. In principle, compound HC0303 should find it difficult to compete with ICAM-1 for binding to the the I-domain. It is necessary, therefore, to consider the possibility of controlling the serum concentration of the inhibitor to be competitive. What is clear is that, as will be explained in chapter 6, compound HC0303 was proved to be an excellent inhibitor of the LFA-1 dependent adhesion of cancer cells to endothelial cells in "in vitro" cell adhesion assays.

Nevertheless, according to our calculations, it seems mandatory to redesign a new and more efficient family of LFA-1 antagonists that would permit to decrease the dose of compound and, therefore, the toxicity of a potential therapy based on this chemical.

Table **5.14**: Contribution of compound HC0303 and the H.A. αL I-domain to the free energy of complex formation, dissected into its components.

Protein	Elec. Int.	Desolv.	Total Elec.	vdW	SAS	Total NP	TOTAL
CH3030	-28.73	27.21	-1.52	-2.61	-8.14	-10.75	-12.27
H. A. LFA	-36.95	29.81	-7.15	-2.30	-0.51	-2.82	-9.96
COMPLEX	-65.68	57.01	-8.67	-4.91	-8.65	-13.56	-22.23

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar ; sum of vdW and SAS components; TOTAL = Total binding free energy (Total Elec. + Total NP)

A complex formed by important electrostatic contacts, and further stabilized by a hydrophobic environment: According to our results, between compound HC0303 and the H.A. I-domain, important electrostatic contacts are created, whose contribution sum up -65.68 kcal/mol (Table **5.14**). Due to the high electrostatic desolvation penalty paid by the complex (+57.01 kcal/mol), the total electrostatic component of the interaction is significantly smaller in magnitude, but still favorable (-8.67 kcal/mol). In addition, this term is still relatively important with respect to the vdW contacts. The latter, favor the interaction between HC0303 and the H.A. I-domain significantly (-4.91 kcal/mol), but less than the total electrostatic term.

The most interesting aspect may be the important contribution to the total free energy of the desolvation of hydrophobic groups. As observed in Table 5.2, the Solvent Accessible Surface term (SAS), which is proportional to the cost of the desolvation of hydrophobic residues (see methods), is -8.65 kcal/mol. In addition, it is compound HC0303 and not the I-domain which contributes with the 100 percent of the SAS term of the complex. This value is as significant as the total electrostatics term (-8.67 kcal/mol), and represents the 40 percent of the total binding free energy value.

These results are coherent with (and represent an additional support to) our quantum results presented in chapter 3. In the latter it was concluded that the permittivity of the environment, i.e. the hydrophobicity of the media, plays a key role on ligand binding regulation: the higher the hydrophobicity, the higher the affinity.

5.3.4 Per-residue decomposition of the total binding free energy

In order to identify valuable residues in compound HC0303 or residues that should be modified or removed, a per-residue decomposition of the free energy of binding was performed. With this goal in mind, compound HC0303 was subdivided in a set of groups of atoms, shown in Figure 5.17. The protocol followed was the same as the one used in previous cases. The per-residue decomposition and the contribution of each term, is shown in Table **5.15**.

Hotspots in the I-domain: Among the residues that contribute favorably to the free energy of binding we find M140, L205, T206 and the Mg^{2+} cation (Table **5.15**). Among them, the Mg^{2+} cation, with a total contribution of -10 kcal/mol, makes the most important interaction. All the residues (but L205) that contribute favorably have the total electrostatic component as the most important term. I.e, the favorable contacts between the H.A. I-domain and compound HC0303 are electrostatic in nature. In addition, residue L205 also makes significant favorable vdW contacts with compound HC0303.

The residues that contributed unfavorably to the free energy of binding are: D137, L204, D239, E241, T243. All of them create unfavorable electrostatic contacts that, in addition, pay a relatively important electrostatic desolvation penalty. On the other hand, interestingly, all of them create favorable vdW contacts, but only T243 in a significant way.

Hotspots in compound HC0303: All the atom groups in compound HC0303 contribute favorably to the binding free energy. Among them, group CH3, with a total contribution of -4.93 kcal/mol, makes the most important contact. The latter is mainly electrostatic in nature (-28.83 kcal/mol) but, unfortunately, the high electrostatic desolvation penalty paid (26.50 kcal/mol), decreases its total electrostatic contribution to -2.33 kcal/mol. This group of atoms also makes favorable vdW contacts with the I-domain. Its contribution (-1.44 kcal/mol), together with the favorable contribution of the desolvation of its aliphatic chain (-1.17 kcal/mol), also favors binding. The rest of the groups favor binding to the H.A. I-domain due to a favorable desolvation of their hydrophobic or non-polar parts. CH2 and CH4, with a total contribution each of -1.10 and -1.22 kcal/mol respectively, are the groups in HC0303 that favor binding to a higher extent.

As predicted before, with a visual inspection of the complex, there is no significant interaction between residue E241 in the I-domain and the NH2 in HC0303. At least, the contacts made are not electrostatic in nature, and that was the aim when the molecule was designed. These results suggest that the spatial distribution of the positive charge in the compound is not optimum. It was therefore decided to design a second generation of the family of compounds represented by HC0303, where the disposition of the positive and negative charges in the molecules is different (see later).

In summary, the MD simulation and the free energy decomposition analysis of the complex between the H.A. I-domain and the LFA-1 antagonist HC0303, provide us with valuable information with respect to the inhibitory activity of the latter. In this sense, it is observed that our compound is relatively well designed, as all the contacts established with the I-domain favor complex stabilization. In addition, it is seen that most of the groups in compound HC0303 function as they were expected to function: i) group CH3 makes the most important contact, a relatively strong electrostatic interaction with the Mg^{2+} cation in the I-domain; ii) the rest of the groups provide the necessary hydrophobic environment that strengthen the interaction (as predicted also by our DFT studies; see Chapter 3).

It was also learned that the antagonists could and should still be optimized, in order to generate a family of LFA-1 antagonist competitive with ICAM-1. In this sense, the interaction between the positive charge in the group CH1 (pyrrol ring) and E241 in the I-domain is not optimized, according to our calculations. The modifications introduced in this first family of inhibitors are explained below.

Table **5.15**: Significant contribution of each amino acid in LFA-1 to the total free energy of binding between 1MQ9 and Compound CH3030

1MQ9	Elec. Int.	Desolv.	Elec.	vdW	SAS	NP	Total
D137	1.15 (0.19)	0.48 (0.05)	1.64 (0.20)	-0.01 (0.00)	0.00(0.00)	-0.0	1.62 (0.21)
M140	-0.71 (0.26)	0.10 (0.03)	-0.61 (0.24)	-0.28 (0.07)	-0.04(0.01)	-0.32	-0.93 (0.32)
L204	0.75 (0.28)	0.52 (0.15)	1.27 (0.40)	-0.26 (0.05)	-0.00(0.05)	-0.26	1.00 (0.45)
L205	-1.28 (0.20)	0.63 (0.04)	-0.65 (0.22)	-1.197 (0.24)	-0.18(0.05)	-1.38	-2.03 (0.51)
T206	-4.38 (0.67)	1.94 (0.35)	-2.44 (0.67)	0.406 (0.79)	-0.00(0.00)	0.40	-2.04 (1.45)
D239	0.61 (0.13)	0.35 (0.14)	0.96 (0.23)	-0.019 (0.01)	0.00(0.00)	-0.02	0.94 (0.24)
E241	0.25 (0.15)	0.71 (0.17)	0.96 (0.28)	-0.219 (0.06)	-0.02 (0.01)	-0.24	0.72 (0.35)
T243	0.34 (0.46)	2.73 (0.37)	3.08 (0.81)	-0.961 (0.17)	-0.17(0.02)	-1.14	1.94 (1.00)
Mg^{2+}	-29.81 (3.53)	19.12 (1.89)	-10.69 (1.81)	-0.018 (0.00)	0.00(0.00)	-0.02	-10.71 (1.82)
TIP1	0.30 (1.05)	0.71 (0.53)	1.01 (0.67)	-0.110 (0.16)	-0.02(0.01)	-0.13	0.88 (0.84)
TIP2	-3.63 (0.53)	2.01 (0.26)	-1.63 (0.38)	1.083 (0.55)	-0.06(0.01)	1.03	-0.62 (0.94)
HC0303	Elec. Int.	Desolv.	Elec.	vdW	SAS	NP	Total
CH1	0.00 (0.16)	0.65 (0.14)	0.65 (0.19)	-0.322 (0.05)	-1.16(0.01)	-1.48	-0.83 (0.25)
CH2	0.02 (0.01)	0.00 (0.00)	0.02 (0.06)	-0.322 (0.05)	-0.80(0.01)	-1.12	-1.10 (0.07)
CH3	-28.83 (4.14)	26.50 (2.32)	-2.33 (2.86)	-1.436 (1.15)	-1.17(0.02)	-2.60	-4.93 (4.03)
CH4	0.06 (0.02)	0.01 (0.00)	0.07 (0.02)	-0.090 (0.02)	-1.19(0.02)	-1.28	-1.22 (0.06)
CH5	0.00 (0.00)	0.00(0.00)	0.00 (0.00)	-0.020 (0.01)	-0.84(0.00)	-0.86	-0.85 (0.02)
CH6	-0.01 (0.00)	0.00 (0.00)	-0.01 (0.00)	-0.008 (0.00)	-0.56(0.02)	-0.57	-0.58 (0.03)
CH7	0.01 (0.04)	0.01 (0.01)	0.02 (0.05)	-0.110 (0.12)	-0.49(0.12)	-0.60	-0.58 (0.28)
CH8	0.05 (0.04)	0.02 (0.01)	0.07 (0.05)	-0.218 (0.09)	-0.67(0.03)	-0.89	-0.82 (0.17)
CH9	-0.01 (0.01)	0.00 (0.00)	-0.01 (0.01)	-0.030 (0.02)	-0.55(0.09)	-0.58	-0.59 (0.12)
CH0	-0.01 (0.04)	0.01 (0.01)	0.00 (0.04)	-0.054 (0.02)	-0.70(0.03)	-0.76	-0.75 (0.10)

Elec. Int. = Electrostatic Interaction (no desolvation included); Desolv = Electrostatic component of desolvation; Total Elec = Total Electrostatic term (Interaction + Desolvation); vdW = vdW Interaction ; SAS = Solvent Accessible Surface or Non-electrostatic component of desolvation; Total NP = Total Non-polar ; sum of vdW and SAS components; TOTAL = Total binding free energy (Total Elec. + Total NP)

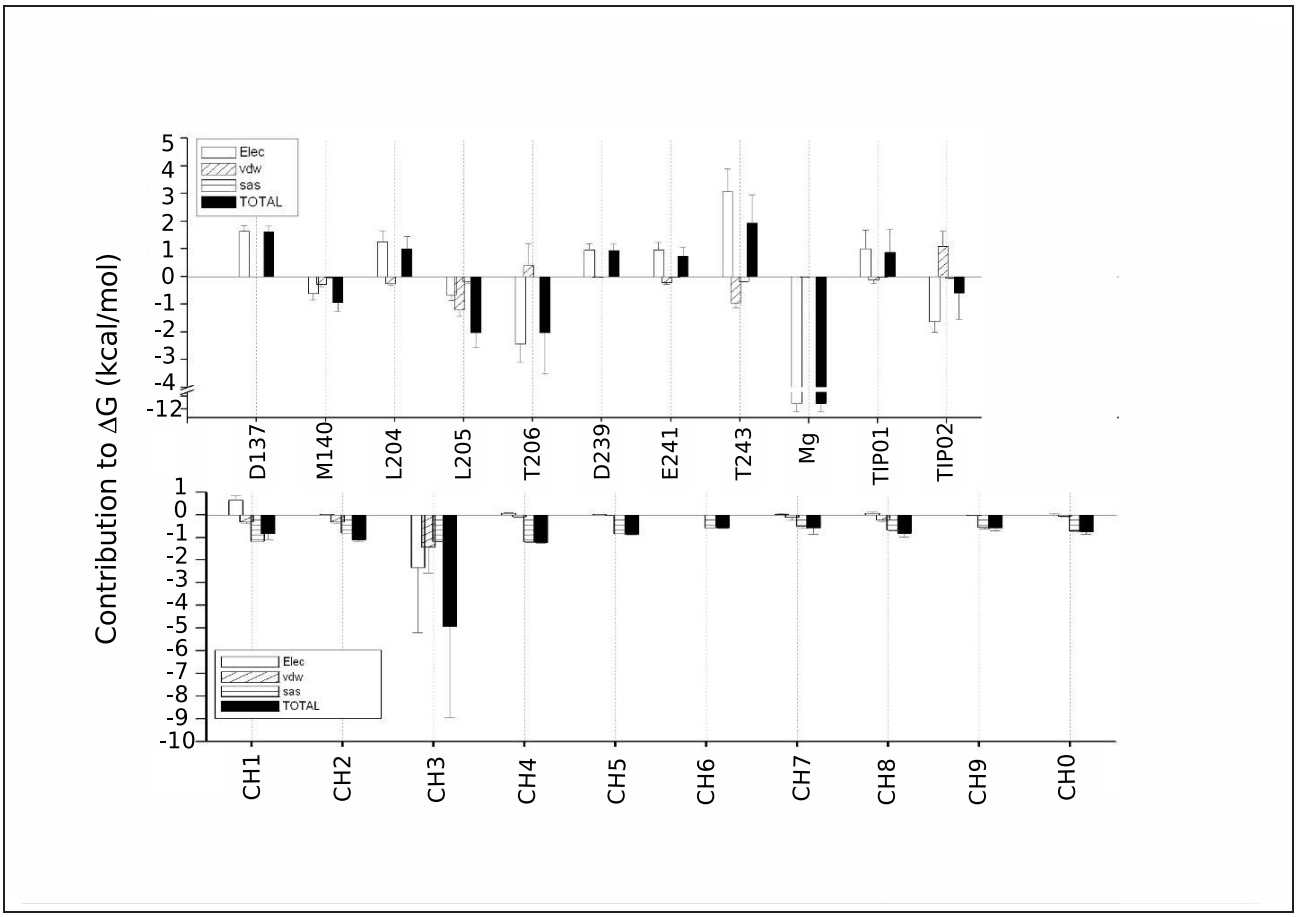


Fig. 5.19: Amino acids in the H.A. I-domain (top) and residues in compound HC0303 (bottom) with a significant contribution to the free energy of binding.

5.3.5 Design of a second generation of LFA-1 antagonists

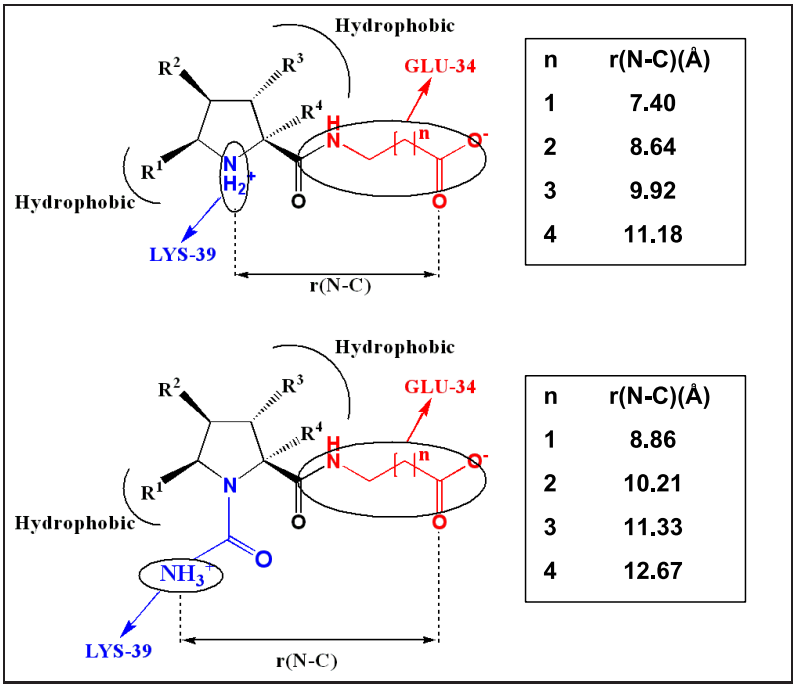
According to the results of the free energy decomposition analysis and the MD simulation, in order to improve the efficiency of HC0303, a modification of the spatial relative disposition of the positive and negative charges in the compound needs to be carried out. Basing our design on the HC0303 binding mode observed in the MD simulation, two possible alternatives may be worth trying: i) An enlargement of the CH3 group and/or ii) the exclusion of the NH2 group from the pyrrolidine ring and its inclusion at the end of an aliphatic chain in the same position of the ring. A simplified representation of the compounds that would result from these actions are shown in Figure 5.20. The number of -CH2 groups (n) will determine the length of the carboxylate group and/or that of the imine group.

As observed in Figure 5.20-top, when n=1 compound HC0303 is obtained. Compounds where n=2,3,4, would represent the second generation of this family of inhibitors. The approximate distance between the N of the imine and the C atom of the carboxylate increases from 7.40 to 11.18 $\mathring{A}$, from n=1 to n=4. It is expected that the longer the length of the aliphatic chain behind the carboxylate group, the higher its flexibility and, therefore, it may be easier to optimize the interaction with E241, without loosing strength on the interaction with the

Mg^{2+} cation. With the same idea, the family of inhibitors depicted in Figure 5.20-bottom was designed. In this case, not only the length of the carboxylate can be enlarged, but the disposition of the $NH2^{+}$ group of the imine is such that the possibilities of both interactions to be optimized at the same time increases substantially.

In Figure 5.21-left, a stick representation is shown of a reduced model of the first family of inhibitors (such as HC0303) docked to the MIDAS motif; in Figure 5.21-right, the same type of representation is shown for the second family of inhibitors. As observed, the molecules represented by the molecular model shown in Figure 5.21-left, are able to make good contacts with the Mg^{2+} cation (green sphere). On the contrary, there is no salt bridge created between the positive charge in the pyrrol ring (highlighted with an asterisk), and the negative charge in E241 in the I-domain. The second family of inhibitors, proposed after having analyzed the MD and FED results, and represented by the molecular model shown in Figure 5.21-right, should be able to better optimize these two interactions simultaneously. In addition, it is believed that an intramolecular hydrogen bond is established, which would not only favor the orientation of the substituents as shown in the figure, but would also decrease the flexibility and therefore the entropic cost of the binding of these type of compounds to the target protein.

In summary, the behavior of compound HC0303 was predicted complexed with the H.A. αL I-domain. In that sense, the complex is, in principle, stable. According to the per-residue decomposition of the free energy of binding all the residues defined in HC0303 make favorable contacts with the protein. The nature of the contacts established are the ones desired, i.e., electrostatic with the Mg^{2+} cation and hydrophobic to strengthen the complex. On the contrary, the salt bridge between the positively charged imine group in the compound and the negatively charged carboxylate groups in residue E241 of the I-domain is not established. As a consequence, a second family of inhibitors was designed which could, according to our predictions, optimize both types of electrostatic contacts simultaneously. The new family of compounds was synthesized by Yosu Vara in our group, and were subjected to *in vivo* and *in vitro* results in the laboratories Dominion Pharmakine Ltd., as was done with the first family of LFA-1 antagonists. The results of such experiments are explained in Chapter 7. For the impatient reader, it is announced that they showed to represent an improved family of LFA-1 antagonists.



n	r(N-C)(Å)
1	7.40
2	8.64
3	9.92
4	11.18

n	r(N-C)(Å)
1	8.86
2	10.21
3	11.33
4	12.67

Fig. 5.20: Top) First generation of LFA-1 inhibitors (n=0,1) designed by former members of the group, and Subgroup I (n=2,3,4) of the second generation of LFA- inhibitors designed as a results of this thesis project. Bottom) Subgroup II of the second family of inhibitors designed as a results of this thesis project. R1, R2, R3, R4 and R5 :Apolar groups used as substituents of the pyrrol ring.

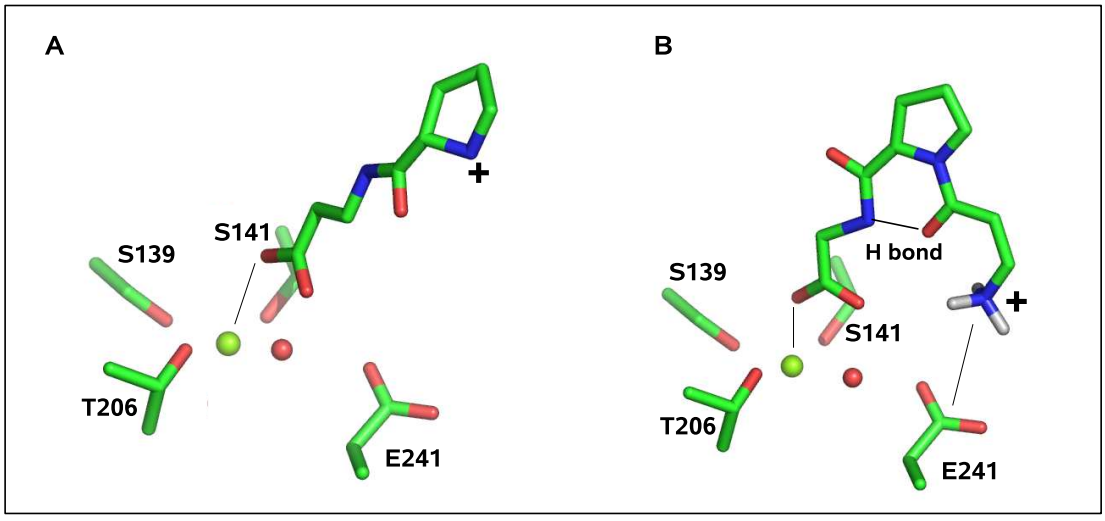
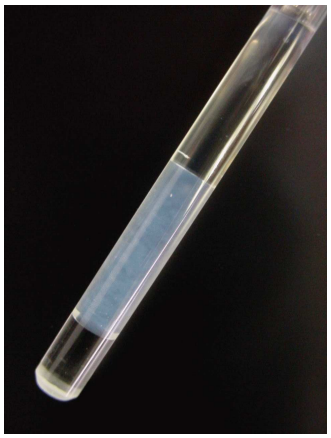


Fig. 5.21: Reduced models of the first and second family of LFA-1 antagonists, interacting with key residues in the I-domain. Key interactions are highlighted with a solid line. The Mg^{2+} cation is a green sphere, and the red sphere represents an equatorial water molecule oxygen coordinated to the metal.

Chapter 6

Identifying the binding pocket in LFA-1

Contents



Introduction to Chapter 6

At least two alternative therapies against cancer metastases and autoimmune diseases are under development so far. One of them consists of intravein administration of monoclonal antibodies against integrins VLA-4 and LFA1 [257, 258, 35], which is pretty expensive and unfriendly for the patient. An alternative therapy consists of the oral administration of low molecular weight $\alpha L\beta 2$ or $\alpha 4\beta 1$ antagonists, which would represent a cheaper and a patient friendly option. In addition, these small molecular weight inhibitors are designed to target specific sites in the integrins, which avoids or at least decreases undesired symptoms of the chemotherapy. The small molecules designed and synthesized in our group fall into this second type of inhibitors.

As part of the thesis projects of former members of the group, small molecular weight inhibitors were designed and synthesized against the interaction between integrin VLA-4 and its counterreceptor VCAM-1. The similarities that exist between VCAM-1 and ICAM-1 concerning their VLA-4 and LFA-1 binding epitopes, respectively, moved former members of the group to design a primary family of inhibitors against the LFA-1/ICAM-1 interaction, using the core structure of the antagonists of VLA-4 as scaffold to build antagonists against integrin LFA-1. In addition, as a result of this thesis work, the primary family of LFA-1 antagonists was optimized, giving rise to a secondary family of LFA-1 inhibitors. The inhibitory activity of the VLA-4 antagonists and of the Subgroup I of a second family of LFA-1 antagonists had already been tested *in vitro* and *in vivo* by the time the experiments presented in this chapter were carried out. The same tests on Subgroup II of the the second family of LFA-1 antagonists designed and synthesized as a results of the work done in this thesis project, and the NMR experiments presented in this chapter, were carried out simultaneously.

In the present part of this thesis work, NMR techniques were used to test the binding of the VLA-4 antagonists and the first and second family of LFA-1 antagonists to both the wild type and active mutant αL I-domains. In the first section of this chapter, the reader will find a summary on the design process and the nature of the inhibitors designed by former members of the group (Dr. Aizpea Zubia, Dr. Eneko Aldaba and Yosu Vara phD student) as antagonists of either integrin VLA-4 or LFA-1. This summary will be followed by a brief overview of the *in vitro* and *in vivo* results obtained with the above mentioned compounds. In the last two sections of this chapter, the reader will find i) the results of the experimental part of this thesis project and ii) a brief summary of the reports found in the literature concerning the localization of the binding site of LFA-1 antagonists designed by other authors.

6.1 Background on "in-house" VLA-4 and LFA-1 antagonists

6.1.1 α4 antagonists

As mentioned in Chapter 1, integrin VLA-4 is composed of the α4 and β1 subunits. VCAM-1 presumably binds to the N-terminal extracellular domain of the α4 subunit of VLA-4. There is no solved structure of the N-terminal domain of the α4 subunit, but it has been predicted to consist of a seven four-stranded β-sheets propeller domain, with a putative Mg^{2+} binding site in the center. No I-domain is inserted [35] in the β propeller domain. On the other hand, domains D1 and D2 of VCAM-1 have been solved by crystallography [73] (Figure 6.1-left) and, in addition, directed mutagenesis studies indicate that the peptidic chain connecting the beta strands in domain D1 (Figure 6.1-left), and residue ASP40 in particular, is crucial for VCAM-1 binding to VLA-4 [75]. In the solvated molecule, the carboxymethyl group of ASP40 has a high degree of conformational freedom and is surrounded by a highly hydrophobic environment, formed by ILE39 and SER41 (Figure 6.1, B). Taking these electronic and geometric features of VCAM-1 into account, inhibitors based on chiral nitroprolines (Figure 6.1 C and Figure 6.2) were designed and synthesized as part of the thesis projects of former members of the group (Silvia Vivanco [107], Aizpea Zubia [106] and Eneko Aldaba [109]). This type of molecules consist of a cyclic structure attached to a carboxymethylated group that could serve to bind VLA-4 by mimicking ASP40 in VCAM-1. In addition, apolar groups attached as substituents to the central ring mimic the hydrophobic environment around ASP40 in domain D1 of VCAM-1 (Figure 6.1, D and E).

These molecules were subjected to *in vitro* and *in vivo* assays in the laboratories of Dominion Pharmakine Ltd [1] in order to evaluate their inhibitory capacity. These results will we shown later in this chapter, together with the equivalent results obtained for inhibitors designed against the LFA-1/ICAM-1 interaction.

6.1.2 αL antagonists

In contrast to α4 in VLA-4, the N-terminal extracellular domain of the αL subunit of LFA-1 has an inserted domain that includes a metal ion-dependent adhesion site whose structure has been solved [60]. In addition, as observed in Figure 6.3, integrin binding sequences in ICAM-1 and VCAM-1 have a high degree of similarity.

The length of the carboxylate group: One of the most (surprisingly) significant differences between the sequences in Figure 6.3 is found in the length of the residues designed to anchor to the Mg^{2+} cation in each integrin, which in ICAM-1 is the GLU34 residue, longer than the ASP40 residue present in VCAM-1. Since the molecules designed to inhibit the VCAM-1/VLA-4 interaction showed excellent inhibitory activity (see later), together with the sequence similarities shown in Figure 6.3 the following possibility was considered: the use of the cyclic structure of the inhibitors depicted in Figure 6.2 as scaffolds to design inhibitors against the LFA-1/ICAM-1 interaction by introducing changes in the substituents (Figure 6.4) in a way that this new family of inhibitors could fit into the binding pocket of the αL I-domain. As mentioned in Chapter 5, GLU34 in ICAM-1 interacts with the Mg^{2+} cation

[1] We are grateful to Dr. Lorea Mendoza (Dominion Pharmakine, Ltd.) and Prof. Fernando Vidal-Vanaclocha (Medical School of the University of the Basque Country) for these biological studies

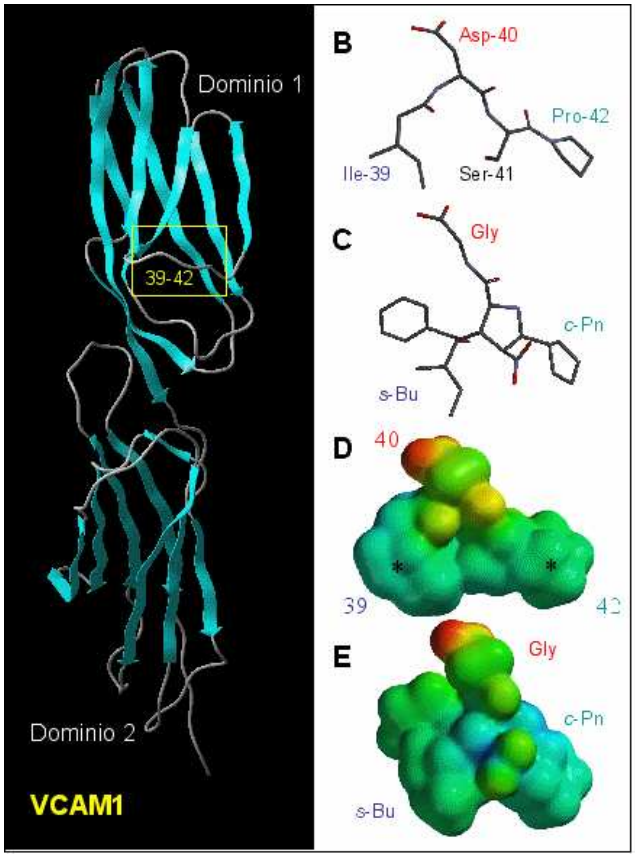


Fig. 6.1: Left, Cartoon representation of the crystal structure of domains 1 and 2 of VCAM-1 (PDB ID=1IJ9 [73]). Right: (B and C) Stick representation of the residues in VCAM-1 and in the VLA-4 antagonist used to bind to VLA-4, respectively. (D and E) Electronic densities of the VLA-4 binding epitope in VCAM-1 and in one VLA-4 antagonist, respectively.

strongly, and LYS39 in ICAM-1 forms a salt bridge with GLU241 in LFA-1. Therefore, the changes in substituents with respect to VLA-4 antagonists mainly consisted of a modification of the length of the carboxylate group to mimic the longer carboxylate of GLU34 in ICAM-1 (Figure 6.4-top, n=0,1). This new family of inhibitors designed by former members of the group is referred to as the primary family of LFA-1 antagonists through this thesis book.

On the other hand, and as concluded in Chapter 5, modifications were still needed in the primary family of LFA-1 antagonists, since the interaction with the Mg^{2+} cation and the salt bridge with GLU241 were not being established simultaneously. In this sense, it was estimated that a combination of amino and carboxy groups at a distance of ca. 8-10 $\mathring{A}$ in the inhibitor should mimic the electronic features of these key residues of ICAM-1. A whole series of inhibitors covering a range of distances and possible orientations between these two functional groups were designed and synthesized. This is referred to as the secondary family of LFA-1 antagonists and consists of two subgroups of molecules; subgroup I, formed by inhibitors shown in Figure 6.4-top, where n=2,3,4 and subgroup II, formed by inhibitors shown in Figure 6.4-bottom. A central scaffold such as a pyrrolidine ring should, again, provide the appropriate relative disposition between key functional groups, and the necessary hydrophobic environment is again provided by the apolar substituents of the ring.

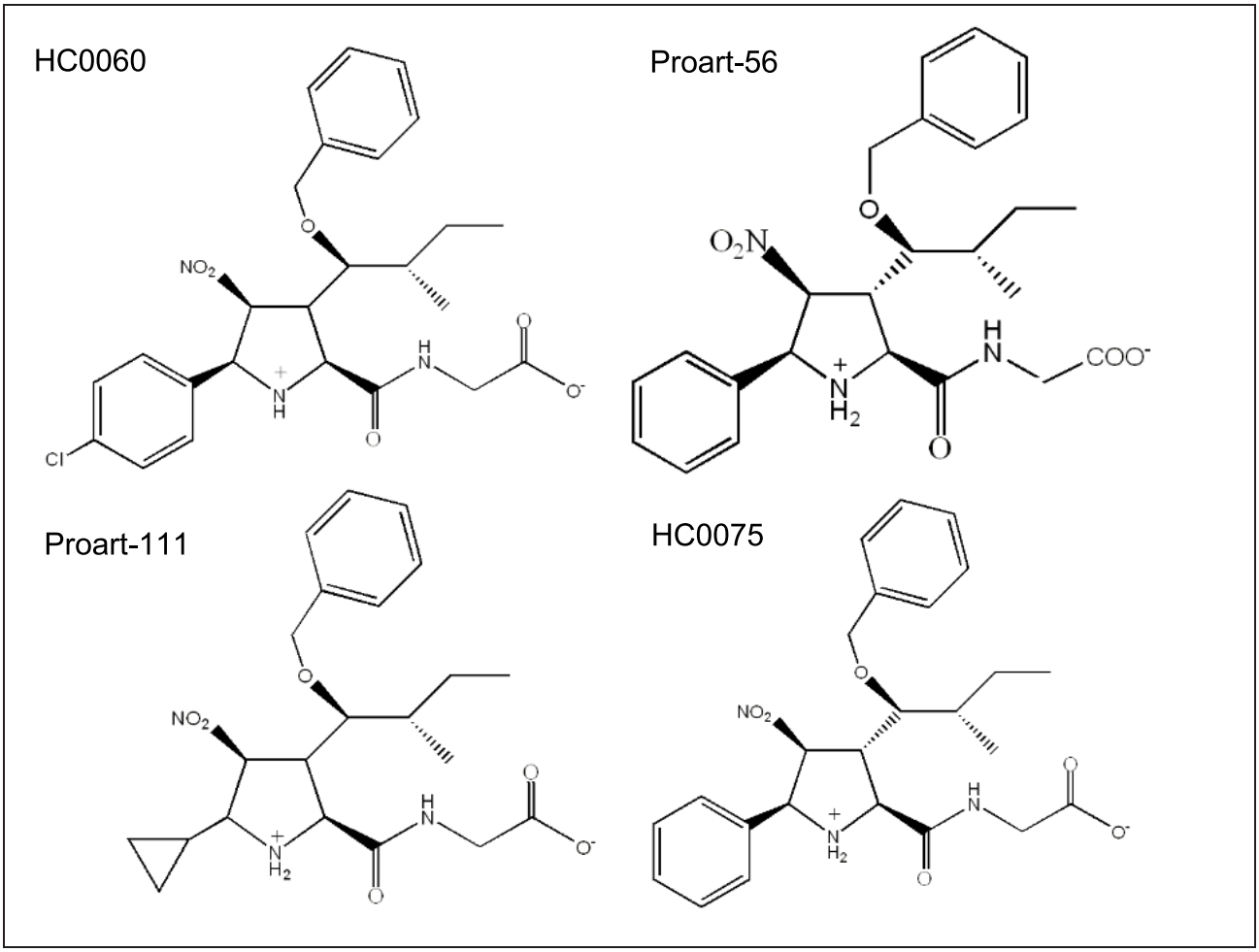


Fig. 6.2: Some of the chiral nitroprolines designed and synthesized by former members of the group as VLA-4 antagonists.

GLN 32 – ILE 33 – GLU 34 – THR 35 – PRO 36 – LEU 37	**in ICAM-1**
GLN 38 – ILE 39 – ASP 40 – SER 41 – PRO 42 – LEU 43	**in VCAM-1**

Fig. 6.3: LFA-1 binding residues in ICAM-1 (top) and VLA-4 binding residues in VCAM-1 (bottom). Equivalent or identical residues in both sequences are surrounded by a box.

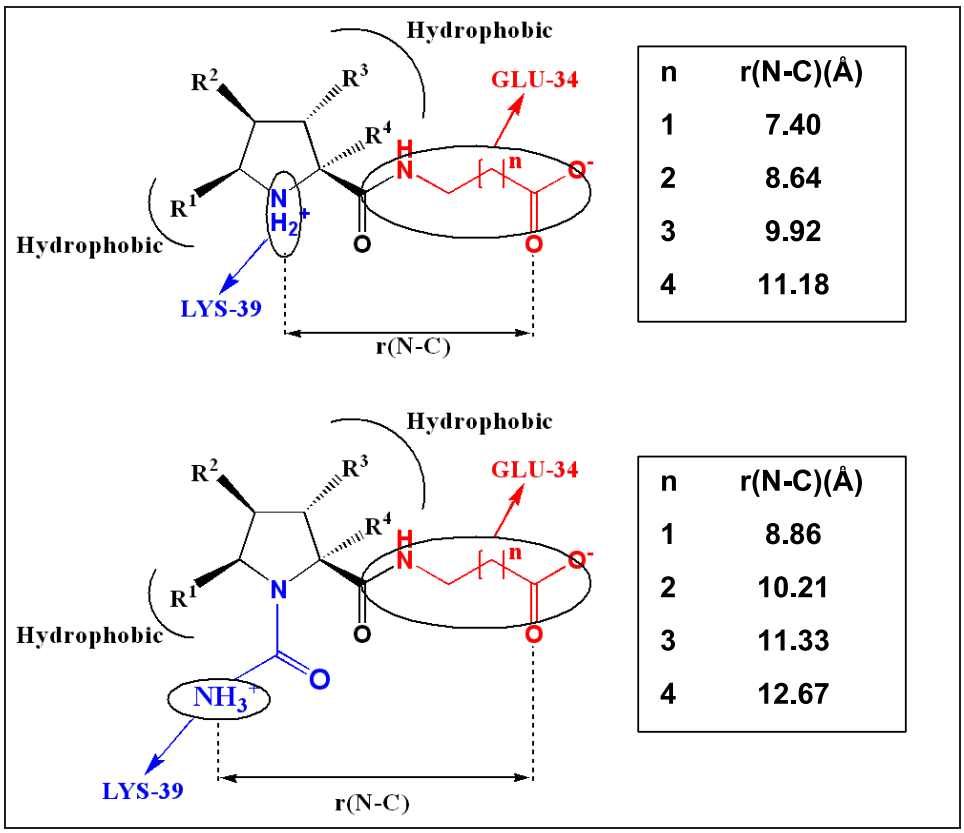


Fig. 6.4: Top: Primary family (n=0,1) and subgroup I of the secondary (n=2,3,4) family of the LFA-1 antagonists designed and synthesized by former members of the group. Bottom: Subgroup II of the secondary family of LFA-1 antagonists designed and synthesized as a result of the present thesis project.

6.1.3 Summary of the *in vitro* and *in vivo* results (I)

The inhibitory activity of some VLA-4 and LFA-1 antagonists selected from the primary family of compounds designed by former members of the group was evaluated through several established models on which integrin VLA-4 or integrin LFA-1, respectively, are involved [2]. The relatively high cost of these type of experiments forced us to choose among the most promising candidates to be subjected to these types of tests. Some of the most significant results are summarized in the lines below.

In Figure 6.5, three examples of the integrin antagonists subjected to the *in vitro* and/or *in vivo* experiments are shown. Compounds HC0225 and HC0303 (also known as endo-anti-55a and endo-anti-63, respectively) are LFA-1 antagonists, and compounds HC0075 (also known as endo-anti-39a) is a VLA-4 antagonist (notice the length of the carboxylate substituent).

[2]Work done on the Medical School of the University of the Basque Country, by the group of Professor Vidal-Vanaclocha

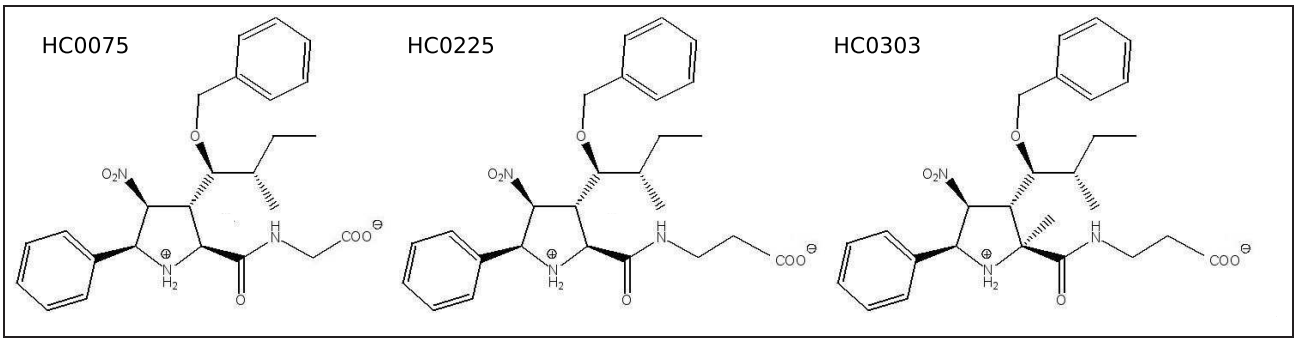


Fig. 6.5: Some of the compounds subjected to *in vitro* and *in vivo* tests. HC0303 and HC0225, were found to be good LFA-1 antagonists. HC0075, was found to be a good VLA-4 antagonist, but showed no activity against LFA-1.

Cell adhesion experiments

Melanoma cell adhesion to hepatic endothelial sinusoidal cells: The *in vitro* inhibitory activity of VLA-4 antagonists (such as HC0075) and LFA-1 antagonists (such as HC0225 and HC0303) was determined on B16 melanoma (B16M hereafter) cell adhesion to hepatic sinusoidal endothelial cells (HSE hereafter). *In vivo*, cancer cells induce the overexpression of ligands on endothelial cells; in these *in vitro* cell adhesion experiments, the overexpression of VCAM-1 or ICAM-1 by the HSE was induced by preincubating the latter with Tumor Necrosis Factor-α (TNF-α hereafter) or Vascular Endothelial Growth Factor (VEGF hereafter), respectively.

Regardless the integrin studied, incubation of HSE cells with TNF-α or VEGF significantly increased B16M cell adherence to HSE, compared to untreated cells (data not shown). In the case of VLA-4 dependent adhesion, compound HC0075 showed the highest inhibitory activity, whereas compounds HC0303 and HC0225 were the most effective ones inhibiting the LFA-1 dependent adhesion. Interestingly, the VLA-4 antagonist HC0075 was not able to inhibit the LFA-1 dependent adhesion, which supports the idea that compounds with a short carboxylic arm such as the one in HC0075, are not able to properly mimic E34 in ICAM-1 and bind the MIDAS motif, whereas a compound with a longer carboxylic arm as that in compounds HC0303 or HC0225, is able to do so. In addition, despite the fact that compound HC0225 showed an acceptable ability to inhibit the LFA-1 dependent adhesion, the methyl group in position 2 of the pyrrolidine ring in compound HC0303, was shown to substantially increase the inhibitory capacity of this family of inhibitors.

Melanoma cell adhesion to immobilized VCAM-1 and lymphocyte adhesion to immobilized ICAM-1: In order to confirm that the adhesion observed in the previous set of experiments was due to the interaction of VLA-4 with VCAM-1 or LFA-1 with ICAM-1, a second set of experiments was performed. B16M cells pre-incubated with Interleukine-18 (IL-18 hereafter) and human peripheral blood lymphocytes (PBLs hereafter) pre-incubated with phorbolmyristate acetate (PMA), respectively, were added to wells containing immobilized VCAM-1 or ICAM-1, respectively. IL-18 and PMA induce the overexpression of integrin VLA-4 and LFA-1, respectively.

The VLA-4 antagonist HC0075 decreased the adhesion of tumoral B16M cells to immobilized VCAM-1, both in the case of IL-18 treated or non-treated B16M cells (data not shown). In the case of LFA-1, both anti-murine LFA-1 monoclonal antibody [3] and compound HC0303 significantly decreased PMA-induced cancer cell adhesion to immobilized ICAM-1 (data not shown). Adhesion of B16M cells to untreated HSE cells was not affected by the inhibitors, which suggests that compounds such as HC0303 inhibit adhesion of the activated LFA-1 to ICAM-1, and to a lesser extent, that of the inactive LFA-1 to ICAM-1.

"In vivo" experiments

In vitro effect of compound HC0303 on *in vivo* metastatic behavior of cancer cells was tested in a murine model of colon cancer. To do so, male mice were intrasplenically injected with viable cancer cells (Co26-CC), in the presence or absence of compound HC0303. Briefly, the number, density and volume of the liver metastases were importantly reduced or abrogated in mice given HC0303-pretreated colon cancer cells in comparison to those receiving untreated cells (data not shown).

In summary, these *in vitro* and *in vivo* experiments show that a promising family of inhibitors was designed by former members of the group, against both the VLA-4/VCAM-1 or LFA-1/ICAM-1 interactions. In addition, it has been observed that compounds that inhibit the VLA-4/VCAM-1 interaction, do not inhibit the LFA-1/ICAM-1 interaction (see compound HC0075), probably due to the length of the carboxylate substituent in the antagonists. Last, it seems that LFA-1 antagonists display their inhibitory activity when integrin LFA-1 is activated, which is crucial for the specificity of a potential treatment of metastasic processes with this family of LFA-1 antagonists.

[3] An antibody is a large Y-shaped glycoprotein used by the immune system to identify and neutralize foreign objects like bacteria and viruses. Monoclonal antibodies are identical antibodies used in cancer treatments due to the fact that they bind only to cancer cell-specific antigens and induce an immunological response against the target cancer cell.

6.1.4 Summary of the *in vitro* and *in vivo* results (II)

The inhibitory activity of the secondary family of LFA-1 inhibitors, designed as a result of this thesis work, was evaluated as before. Some of the most significant results are summarized in the lines below, and compared to results obtained with the best candidate of the previous generation of LFA-1 antagonists designed in the group (compound HC0303).

In Figure 6.6 and Figure 6.7, shown are 6 compounds that belong to the subgroup I and 4 compounds that belong to the subgroup II, respectively, of the secondary family of the LFA-1 antagonist. All of them were subjected to the *in vitro* or/and *in vivo* tests explained below.

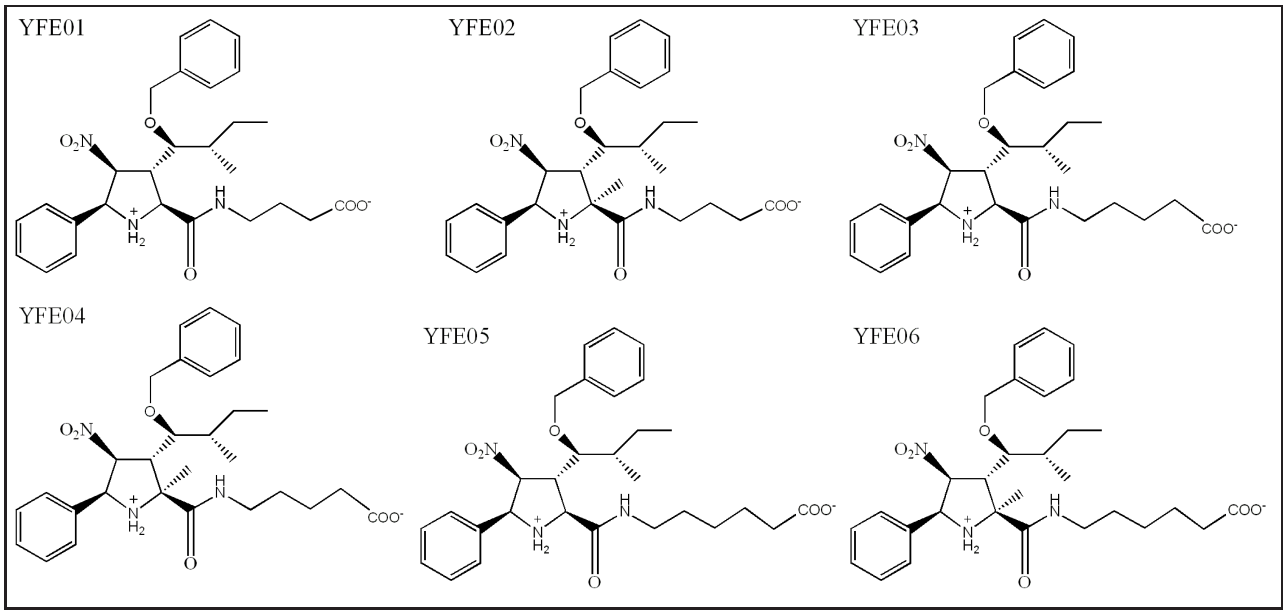


Fig. 6.6: Subgroup I of the secondary family of LFA-1 antagonists designed in this thesis work.

Cell adhesion experiments

Inhibitory effect of LFA-1 antagonist cancer cell adhesion to HSE cells. The *in vitro* inhibitory activity of subgroup I and subgroup II of the secondary family of LFA-1 antagonists was determined on cancer cell (CT26 cells) adhesion to hepatic sinusoidal endothelial cells. The overexpression of LFA-1 by the cancer cells was induced by preincubating the latter with PMA.

As observed in Figure 6.8, incubation of CT26 cells with PMA significantly increased adherence to HSE cells, compared to untreated cells. Compounds YFE05 and YFE10 were the most effective ones inhibiting cancer cell adhesion to the HSE, up to 65% and 48% of inhibition, respectively. But, more importantly, both compounds were shown to be more efficient than compound HC0303, which inhibited adhesion up to 40% and that was proved to be the best inhibitor among the primary family of LFA-1 antagonists in previous experiments.

These results indicate that predictions done in this thesis work with respect to the necessity of an adjustment of the relative disposition of the positively charged amine group and the negatively charged carboxylate groups in the inhibitors, were correct. At least one member of each of the two subgroups designed as possible improved generations of the first subfamily

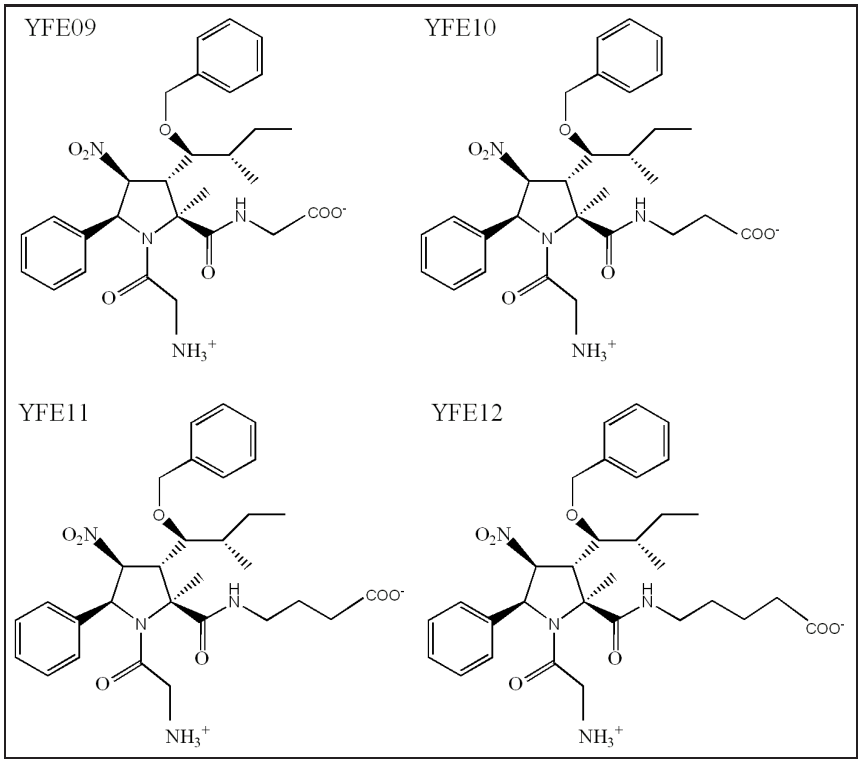


Fig. 6.7: Subgroup II of the secondary family of LFA-1 antagonists designed in this thesis work.

of LFA-1 antagonists, showed to be better than HC0303. An increase in the length of the carboxylate group (YFE05) or the new disposition of the amine group out of the pyrrolidine ring in position 1 (YFE10) was shown to substantially increase the inhibitory capacity of this family of inhibitors. On the contrary, compound YFE03 seems to be an activating agent of the CT26 cell adhesion to HSE cells. Nevertheless, MD simulation and docking studies are still required to analyze what are the particular geometric features in compounds YFE05 and YFE10 that make of them the best hits among all of the rest of the compounds in the same subgroups.

Inhibitory effect of LFA-1 antagonist on leukocyte cell adhesion to immobilized ICAM-1. In order to confirm that the adhesion observed in the previous set of experiments was due to the interaction of LFA-1 with ICAM-1, PBLs pre-incubated with PMA, in the absence or presence of one of our inhibitors, were added to wells containing immobilized ICAM-1. Some of the results are shown in Figure 6.9.

As observed, compounds listed in Figure 6.9 significantly decreased PMA-induced cancer cell adhesion to immobilized ICAM-1. In particular compound YFE05, again, displayed the best behavior among all the compounds tested, decreasing adhesion in a 73%. The inhibition provoked by compound HC0303 was lower (%50), but still high. The present set of experiments did not point to any activating role of compound YFE03, as the previous set of experiments did.

As the only known ligand for ICAM-1 is integrin LFA-1, these results indicate that inhibition is related to compound binding to LFA-1 in a way that interaction to ICAM-1 is blocked.

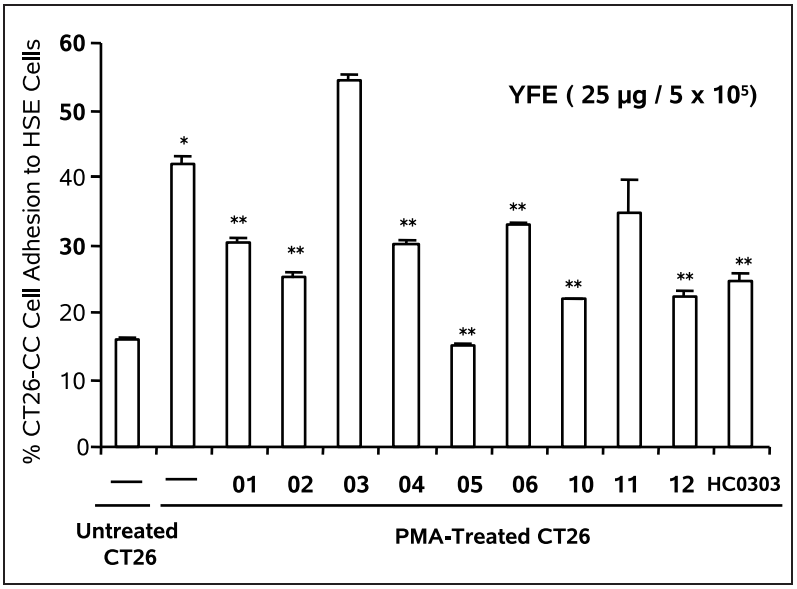


Fig. 6.8: Cancer cell adhesion to HSE cells. CT26 cells were preincubated with either basal medium or 50 nM PMA. Some cells were treated with 5 g/10^5 cells LFA-1 antagonists/PAMAMG3.5 30 minutes before cell adhesion. Once treatments finished, cells were labeled with BCECF-AM and adhesion assay to a rhICAM-1-coated plate was performed. The results are the mean SD of three separate experiments, each in sextuplicate (n= 18). Differences in the percent of adhering cells with respect to (*) basal medium or (**) PMA-treated CT26 cells were statistically significant ($p < 0.01$) by ANOVA and Bonferronis post-hoc test.

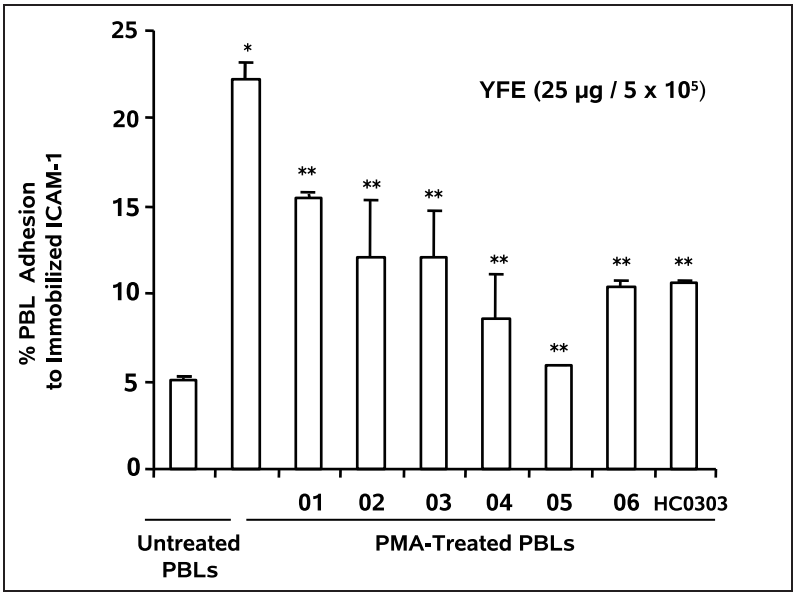


Fig. 6.9: PBL cells were preincubated with either basal medium or 50 nM PMA. Some cells were treated with 5 g/105 cells LFA-1 antagonists/PAMAMG3.5 30 minutes before cell adhesion. Once treatments finished, cells were labeled with BCECF-AM and adhesion assay to a rhICAM-1-coated plate was performed. The results are the mean SD of three separate experiments, each in sextuplicate (n= 18). Differences in the percent of adhering cells with respect to (*) basal medium or (**) PMA-treated BPL cells were statistically significant (p¡0.01) by ANOVA and Bonferronis post-hoc test

"In vivo" experiments

As compounds YFE05 and YFE10 were the best hits in the *in vitro* cell adhesion experiments summarized above, their *in vivo* metastatic behavior was tested in a murine model of colon cancer. To do so, male mice were intrasplenically injected with viable CT26 cancer cells in the presence and absence of compounds YFE05 or YFE10. Results are shown if Figure 6.10 and explained in the paragraphs below.

Figure 6.10 shows tissue which contains primary tumors of mice treated with compounds YFE05 (A) or compound YFE10 (B), together with control tissue with primary tumors not treated with YFE05 or YFE10. Plot C, summarizes the results obtained after a visual inspection of the mentioned tissues. As observed, the number, density and volume of the primary tumors were importantly reduced in mice given YFE05 or YFE10-pretreated colon cancer cells in comparison to those receiving untreated cells. Whereas YFE05 reduced the volume of the primary tumors in a 85% with respect to the control experiment, compound YFE10 reduced the volume in a 77%.
Plot D in Figure 6.10 shows the expression level of protein Ki67[4], whose expression by cancer cells is importantly related to the its malignancy. As observed, compounds YFE05 and YFE10 significantly decrease the expression of the Ki67 antigen, up to a 38% with respect to the control experiment. It is therefore concluded that these two compounds provoke not only a substantial decrease on the number, density and volume of the primary tumors, but the primary tumors still present in the liver are less malignant. It is believed that this antiproliferative capacity of compounds YFE05 and compound YFE10 are related to the inhibition of the angiogenic process that guides blood vessels formation.
The reader must take into account that most, if not all, the conclusions of this thesis project were obtained by focusing on the αL I-domain of the integrin as an isolated unit, whereas the experiments summarized above were carried out with cells expressing the intact heterodimer. This "difference", as will be explained later in this chapter, may be of great importance when trying to properly understand the interaction between small molecular weight inhibitors and integrin LFA-1. In addition this idea may be crucial to properly interpret the NMR results shown later in this chapter.

[4]Ki67 is a big nuclear protein responsible for keeping the cellular cycle alive. The Ki67 antigen is considered a marker of the proliferation of neoplasic cells. There is an important correlation between the Ki67 index and the malignancy of a cancer cell.

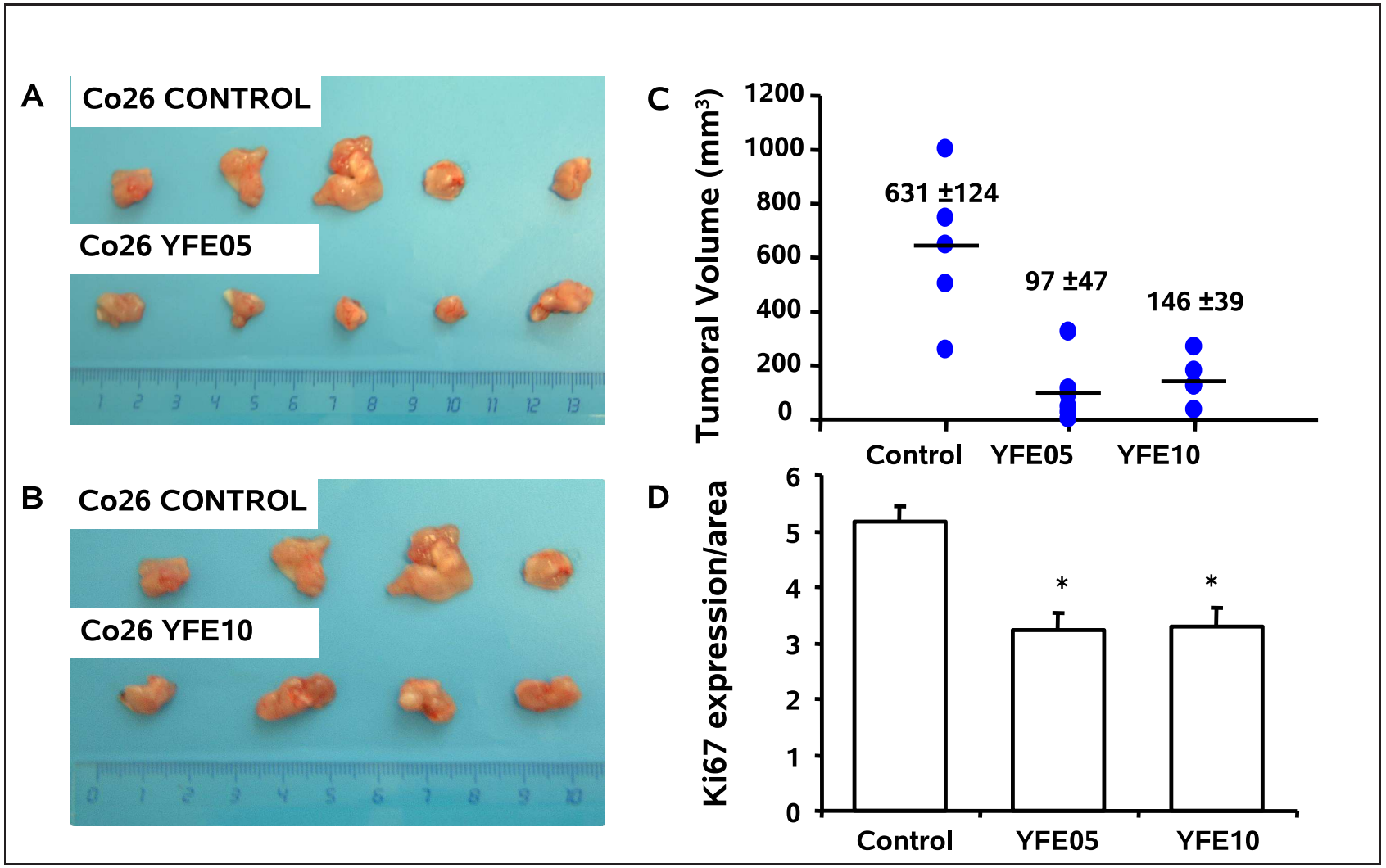


Fig. 6.10: LFA-1 antagonists inhibits CT26M tumor growth. CT26-CC cells ($5x10^5$) were subcutaneously injected into C57BL/6J mice. Mice received 2.5 mg/kg YFE05/PAMAMG3.5 and YFE10/PAMAMG3.5 daily. Primary tumors were removed on day 15th and tumor volume (C), and KI67-expressing melanoma cells (D) were determined by inmunohistochemistry. Data represent values SE. Differences in average values were statistically significant with respect to control mice by ANOVA and Tamhanes post-hoc test (*p¡0.05).

6.2 NMR results

The goal of the experimental part of the project was to localize the site in LFA-1 to which a set of LFA-1 antagonist bound in solution. As mentioned earlier in this chapter, the compounds were designed to bind to the *open* MIDAS motif of the active αL I-domain, the MIDAS motif being, therefore, the expected binding pocket to be identified. As a negative control, the binding of several VLA-4 antagonists was tested. For some of the LFA-1 and VLA-4 antagonists tested in this experimental part, *in vivo* and *in vitro* results were already available, which pointed to some of the compounds as being excellent inhibitors of the LFA-1 or VLA-4 dependent adhesion, respectively. Needless to say that working with the intact integrin would have represented a better choice, but its purification requires unaffordable resources for the group.

As mentioned in Chapter 1, integrin LFA-1 is constitutively expressed in leukocytes in a default inactive state, and is quickly transformed into the active state when required. In cancer cells, the integrin is supposed to be activated. This activation of the integrin seems to involve a conformational change of the I-domain in the αL subunit. Shimaoka and coworkers [60] were able to cristallyze the isolated αL I-domain in a number of different states; among them we find the wild type I-domain, which showed to have very low affinity for ICAM-1 and the high affinity mutant. In the latter, mutation of two specific residues to cysteines (K287C/K294C) generates a disulfide bond that provokes a 10 $\mathring{A}$ downward movement of the C-terminal α7 helix. The affinity of this construct for ICAM-1 was measured and found to be 10000 times the one of the wild type I-domain [60]. It was therefore suggested that the activation of LFA-1 is accompanied by the displacement of the C terminal α7 helix. In addition, they were able to cocrystallize the mutated construct complexed with ICAM-1 and observed that the MIDAS motif was in the *open* conformation, whereas in the ligand-free wild type αL I-domain the MIDAS motif is in the *closed* conformation. Whether the close-to-open conformational change in the MIDAS motif precedes or follows ICAM-1 binding needs to be clarified yet.

In this study, NMR techniques were used to test the binding of the VLA-4 and LFA-1 antagonists both to the wild type and active mutant αL I-domains. The *in vivo* and *in vitro* experiments carried out in the laboratories Dominion Pharmakine Ltd. with the intact heterodimer showed that our compounds were specially good at inhibiting adhesion of activated LFA-1 to ICAM-1, compared to the adhesion of non-activated LFA-1 to ICAM-1. Nonetheless, it was considered interesting to check the existence of a binding pocket for our compounds in the wild type αL I-domain too, which is the state of the integrin under resting conditions *in vivo*. Binding of VLA-4 antagonists to the αL I-domains was used as control experiments for absence of binding.

6.3 Results: Detecting binding of compounds

6.3.1 Acquisition and assignment of HSQC spectra

The first step to take when identifying a binding site of a ligand in a protein using NMR methodology, usually consists of the acquisition of a HSQC spectrum of the ligand-free protein. Each signal in a regular $^{1}H^{15}N$-HSQC spectrum is usually assigned to an amino acid of the protein studied. By doing so, a fingerprint of the isolated protein in solution is obtained, which may be modified by the presence of a ligand in a second $^{1}H^{15}N$-HSQC experiment.

For the assignment of the backbone of the wild type αL I-domain, HNCACB and CBCA(CO)NH experiments[5] were acquired and processed. The complete backbone assignment is found in Figure 6.11 and agrees with that obtained by Kriwacki and coworkers [259] (not shown), which indicates that the protein has been successfully purified in the desired conformation.

The assignment of the backbone of the active mutant αL I-domain was not finished by the time this thesis project was written. For this reason, the results concerning the active mutant αL I-domain were interpreted by comparing its $^{1}H^{15}N$-HSQC spectrum to that of the wild type I-domain. Nevertheless, as observed in Figure 6.11, the $^{1}H^{15}N$-HSQC spectra of both the wild type and active mutant I-domains are easily superimposable, so that residues in the mutant αL I-domain could be tentatively assigned in a reliable way.

Interpretation of the $^{1}H^{15}N$-HSQC spectra

A careful analysis of the $^{1}H^{15}N$-HSQC spectra of the wild type and the active mutant αL I-domains provided interesting information concerning the structure of the constructs purified, which is summarized below.

The active mutant αL I-domain does not have an open MIDAS motif: As observed in Figure 6.11, the signals corresponding to S139, S141, T206 and D239 in the wild type I-domain spectrum (red signals) do not undergo significant chemical shift perturbation upon introduction of the disulfides to generate the mutant I-domain (black signals). For a significant chemical shift perturbation, see signal of residue T291 at the top right panel. This observation implies that the MIDAS residues in the wild type I-domain do not suffer any change in their environment which promotes a chemical shift perturbation in their NMR signals. The closed-to-open conformational change would imply that at least residues T206 and D239 observe a significant change in their electronic environment, since they move from the first to the second coordination shell of the Mg^{2+} cation and viceversa, respectively. In other words, the MIDAS motif, which was supposed to be in the *closed* conformation in the wild type I-domain, seems to be still in the *closed* conformation in the active mutant too.

[5]The HNCACB 3D experiment resolves amide proton/nitrogen correlations; the third dimension contains the ^{13}C chemical shifts of the Cα and Cβ resonances of a given residue and of the preceeding residue. The CBCA(CO)NH 3D experiment correlates Cα and Cβ resonances to the backbone amide ^{1}H and ^{15}N resonances of the next residue.

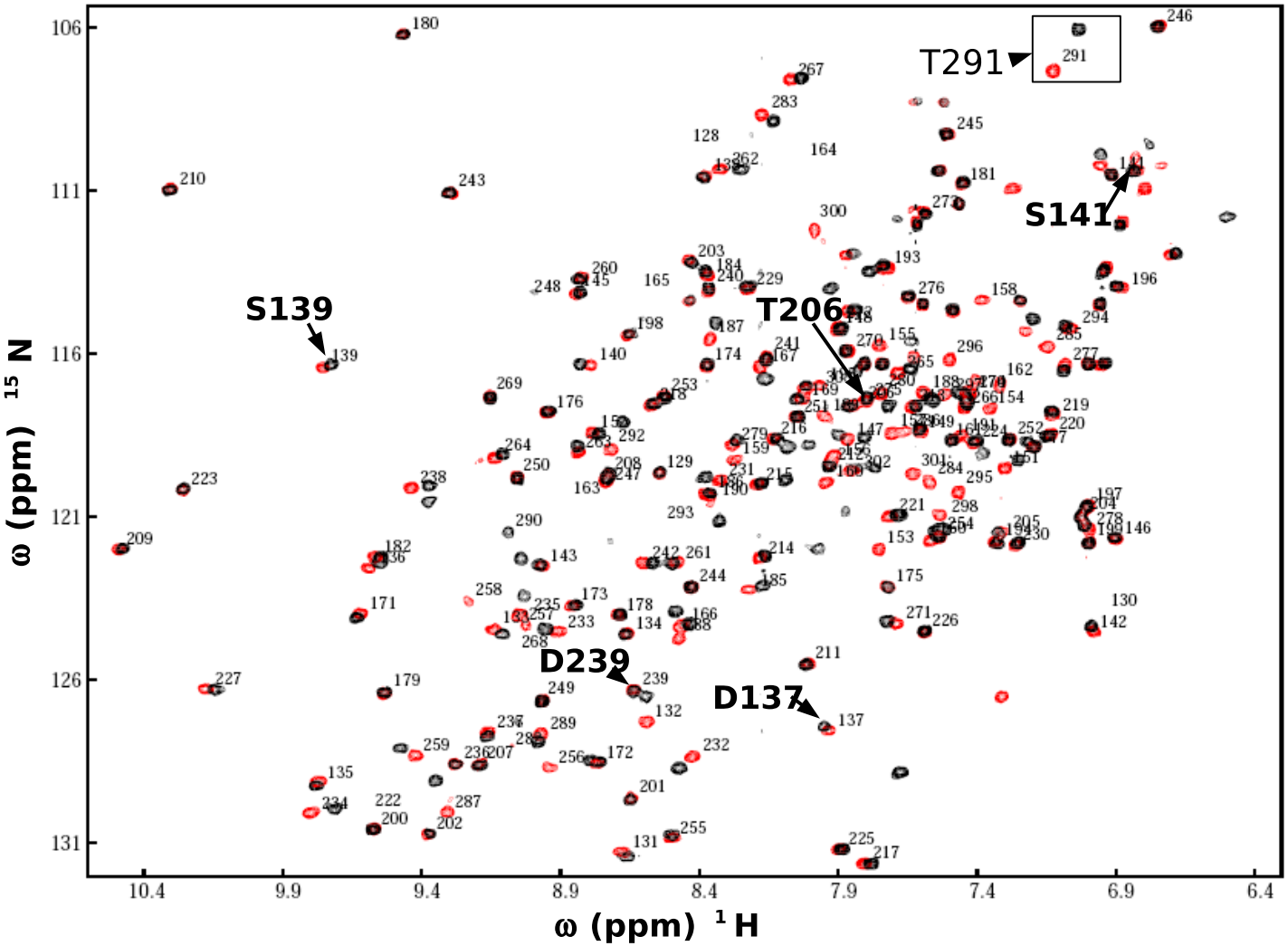


Fig. 6.11: Overlay of the ^{15}N-HSQC spectra of the wild type (red) and mutant (black) αL I domains. To each signal, a residue number has been assigned according to the reported sequence of the protein (black labels). Highlighted in with a black arrow: Residues coordinated to the metal cation either in the *closed* or the *open* MIDAS motif. Highlighted in with a red arrow: Residue 291, as an example of a residue undergoing a significant chemical shift perturbation.

The active mutant undergoes a change on the C-terminal α7 helix: In Figure 6.11 it is observed that upon introduction of disulfides to generate the active mutant αL I-domain, some residues undergo significant perturbations on their chemical shifts or, alternatively, their signal is not detected anymore. These residues correspond to two epitopes in the I-domain: i) residues 258 to 262 and ii) most of the residues comprised between residue numbers 284 and 300. If these residues are localized in the structure of the mutant αL I-domain [178], as shown in Figure 6.12, it is observed that the residues affected by the presence of disulfide bonds are, as expected, localized to the C-terminal α7 helix, as well as to the preceding α6-β6 loop, β6 strand and β6-α7 loop, and the spatially close β5 strand. These results suggest that the C-terminal α7 helix has undergone the desired conformational changes, and represents a positive control of the presence and effect of the disulfide bond. On the contrary, the disulfide bond itself, according to the results summarized in the paragraph above, is not able to provoke the closed-to-open conformational change on the MIDAS motif.

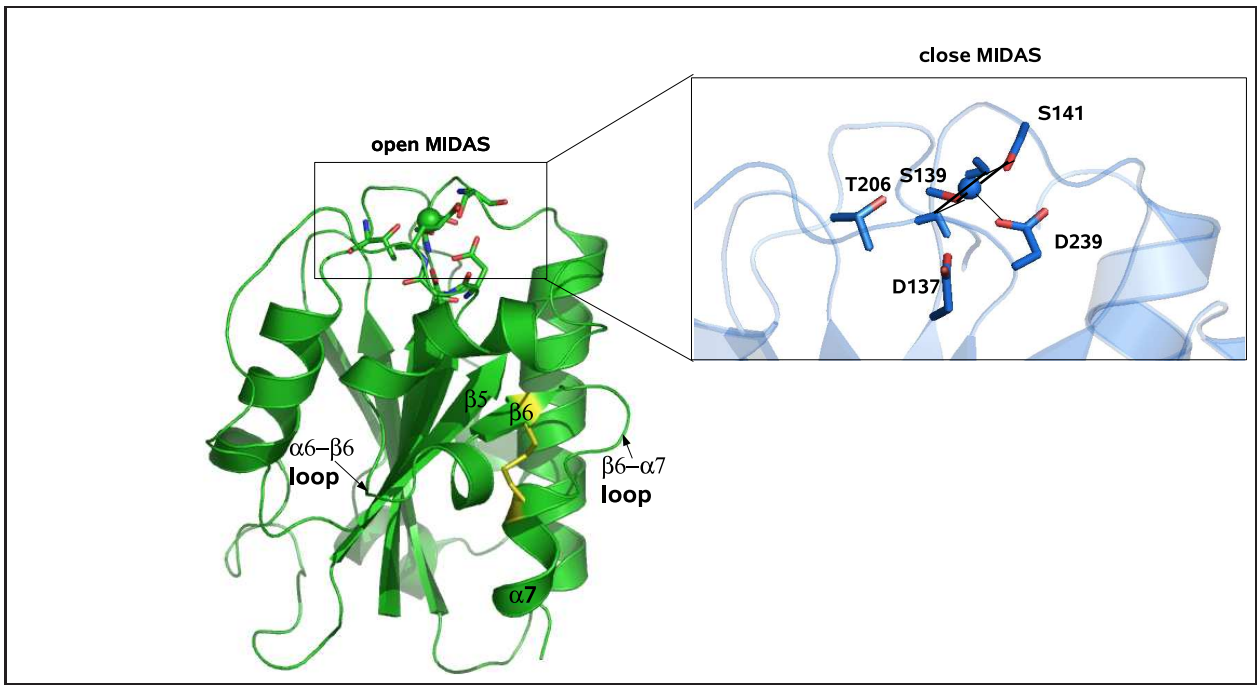


Fig. 6.12: Cartoon representation of the crystal structure of the high affinity mutant αL I-domain. In stick, yellow: CYS 287 and CYS 294. Right panel: *closed* MIDAS motif present in the low affinity αL I-domain. It is proposed that the latter is present in the high affinity mutant αL I-domain, instead of the *open* MIDAS motif reported in the presence of ICAM-1.

It is therefore proposed for the first time that the structure known as the high-affinity mutant αL I-domain has a MIDAS in the *closed* conformation, despite the presence of the disulfide bond that is known to provoke, also in this case, the downward movement of the C-terminal α7 helix. It is further hypothesizes that, in the present of ICAM-1 or a pseudoligand, the MIDAS motif may undergo the closed-to-open conformational change. In fact, the crystal structure of this high affinity mutant in the presence of a pseudoligand (PDB ID=1MQ9) and in the presence of ICAM-3 (PDB ID = 1T0P), have an *open* MIDAS motif.
.

6.3.2 Chemical shift perturbation experiments

Chemical Shift Perturbation (CSP hereafter) experiments were performed with the wild type and active mutant αL I-domains in order to detect binding, pinpoint binding site and measure binding constants (Kd). With these purposes, seventeen different ^{15}N-HSQC spectra were acquired, one per antagonist added (15 LFA-1 antagonists + 2 VLA-4 antagonists) to the NMR sample containing wild type αL I-domain. By comparing the spectrum of the isolated I-domain in the presence and in the absence of a given compound, the binding pocket in the wild type I-domain was identified. The same procedure was followed to locate the binding site of of these compounds in the active mutant I-domain. The results are summarized in the following paragraphs.

Identifying the binding site

All the compounds tested bind to the same allosteric site in the wild type I-domain: In Figure 6.13, 18 different ^{15}N-HSQC spectra of the wild type I-domain are overlaid. The latter correspond to one spectrum for each of the 17 compounds added to wild type I-domain sample (yfeo6, yfeo1, yfeo2, hco225, yfe03, yfe04, yfe05, hco289, hco211, hco230, hco287, hco234, hco303, hco207, hco233 against LFA-1 and, as controls, hco060 and proart111) (signals in any color but red), plus the reference spectrum in the absence of compounds (signals in red).

As observed in Figure 6.13, every compound clearly interacted with the I-domain, since all of them provoked chemical shift perturbations of several residues. All the compounds in the family affected the chemical shifts of the same residues, but to a different degree, as observed on the right panel in Figure 6.13. Visual analysis of the spectra revealed that compound YFEO2 (blue spectra in Figure 6.13) caused, in most of the affected signals, the largest chemical shift deviation. Due to the structural similarity, it was assumed that all the compounds in our library bound in a similar orientation to the binding pocket, and therefore that YFE02 was the compound that most strongly bound to the wild type I-domain. Subsequent analysis of the binding properties of this set of compounds was therefore performed using compound YFE02 as the most representative member of the family. The interpretation of the results of these experiments, is given in the paragraphs below.

LFA-1 and VLA-4 antagonists, bind to the IDAS site in the wild type αL I-domain: The distance each peak had moved due to the presence of YFEO2 was calculated using a Perl script written in-house [6] and Equation 6.1, where CDS stands for Chemical Shift Difference and $\Delta\delta_{HN}$ and $\Delta\delta_N$ represent the chemical shift difference in the proton and nitrogen axis, respectively. The corresponding distances are plotted in Figure 6.14. All those peaks that had shifted a distance higher than the experimental error[7]+the average distance shift, were considered to be significantly perturbed by the presence of YFE02. Likewise, those peaks which had broadened to the point of disappearance were considered significantly affected.

$$CSD = \sqrt{\frac{1}{2}\left((\Delta\delta_{HN})^2 + (\Delta\delta_N/5)^2\right)} \tag{6.1}$$

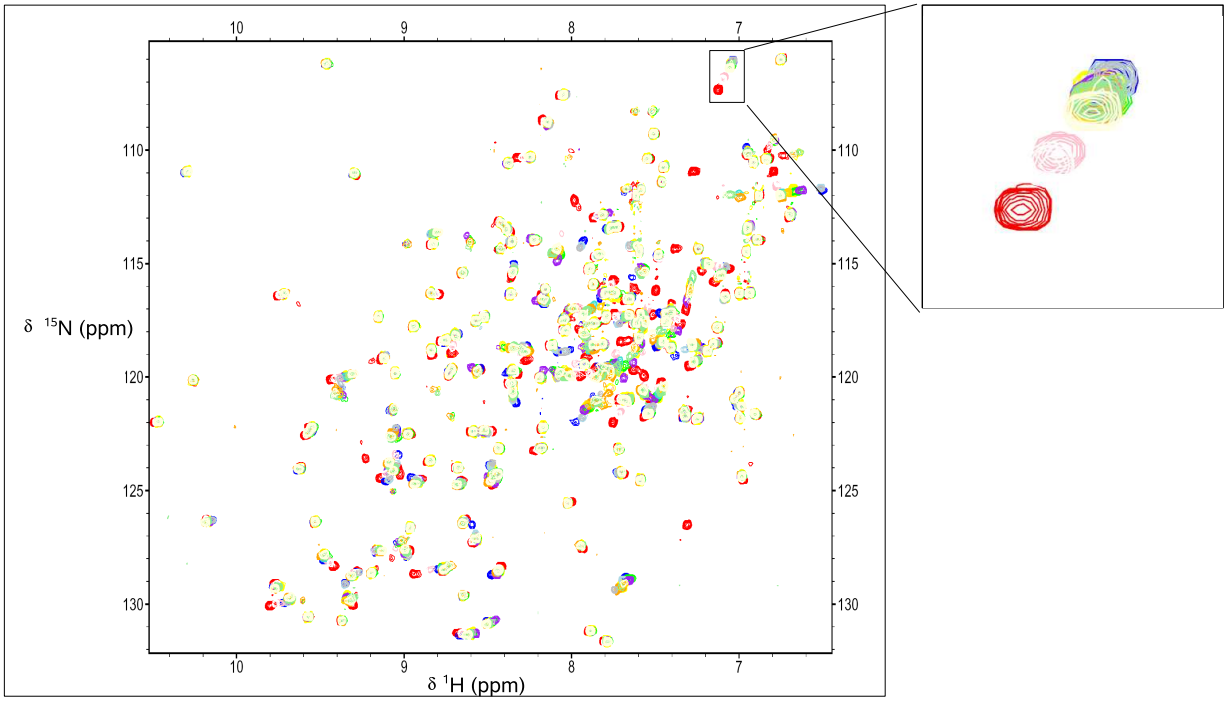


Fig. 6.13: Overlay of multiple HSQC spectra, in the absence of inhibitors (red) or in the presence of: YFE02 (blue), HC0225 (yellow), YFE03 (green), YFE04 (light blue), YFE05 (purple), HC0289 (pink), HC0211 (turquoise), HC0230 (orange), HC0233 (light yellow), HC0287 (white), HC0234 (gold), HC0303 (gray), HC0207 (light green). Left, the complete spectrum. Right, a magnified area of the spectrum, showing the perturbation of the chemical shift of the signal corresponding to residue 291.

The residues corresponding to shifted or disappeared peaks mapped to a discrete area on the structure of wild type αL I-domain (C-terminal α7 helix, Figure 6.15). Surprisingly, none of the MIDAS site residues were significantly perturbed (S139, S141, M140, D239, T206), so it is improbable that the compound bound there. Instead the compound appeared to bind to the

[6] We are grateful to Tahl Zimmerman, PhD student from the Centro Nacional de Investigaciones Oncológicas

[7] The experimental error is the sweepwidth/#points in the proton and nitrogen dimensions (spectral resolutions) plugged into Equation 6.1 where dH is the spectral resolution in the proton dimension and dN is the spectral resolution in the nitrogen dimension. The experimental error for the spectra of the LFA1 is 0.0157

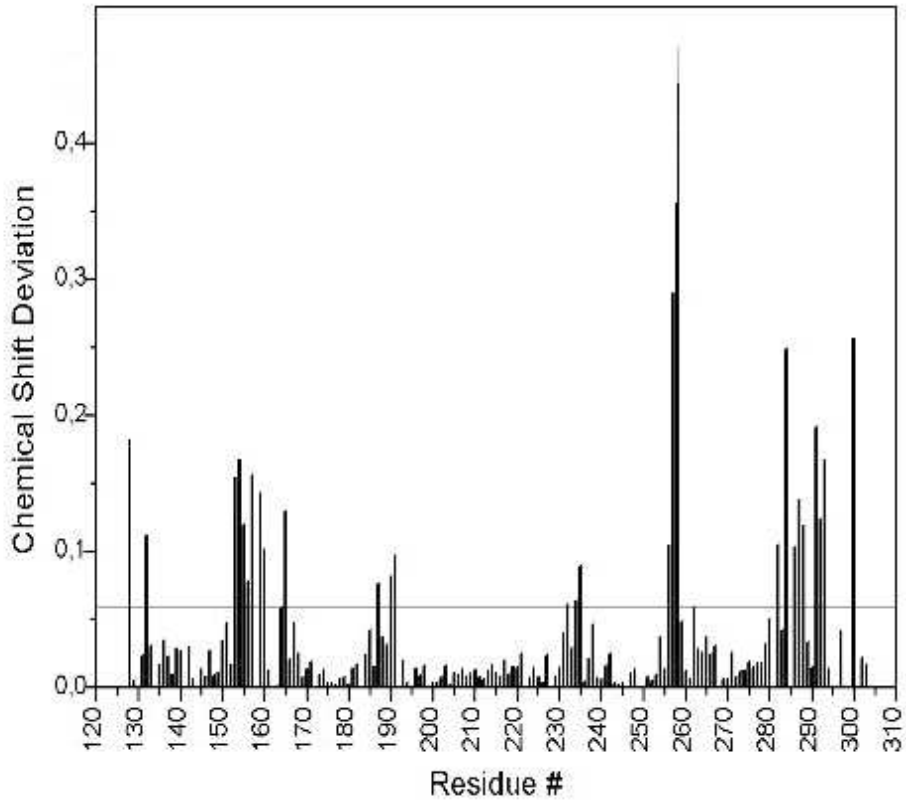


Fig. 6.14: Chemical shift perturbation of residues in the wild type I-domain upon addition of compound YFE02. The horizontal line indicates the average distance shifted+error

allosteric or IDAS site of the wild type αL I-domain.

All the compounds tested seem to bind to the IDAS site of the wild type αL I-domain in the same way, regardless their nature as VLA-4 or LFA-1 antagonists. Compound HC0075, for example, which was shown to be a good VLA-4 inhibitor and a bad LFA-1 inhibitor in the *in vitro* and *in vivo* experiments with the intact heterodimeric LFA-1 (see previous section), binds to the IDAS of the αL I-domain in the same manner as LFA-1 antagonists do.

The results explained above were surprising in two different senses: First, it was expected that LFA-1 antagonists, such as YFE02, would bind the MIDAS motif at the top of the domain and not the IDAS site. Second, it was not expected that VLA-4 antagonists, such as HC0075, would bind to the same site in LFA-1. Taking the *in vivo* results as reference, it appears that the isolated αL I-domain has a very "sticky" pocket in the surroundings of the C-terminal α7 helix. Different type of molecules, regardless of their chemical nature or geometry, were shown to bind to the IDAS site in the wild type I-domain, which is a highly hydrophobic pocket [8]. In fact, a modest high-throughput screening was made with several compound taken from Aldrich, and it was not possible to find a single molecule that would not bind the IDAS site. In light of this data, it may be reasonable to think that the isolated wild type αL I-domain is not an appropriate biological unit to be considered representative of the intact LFA-1 heterodimer found *in vivo*.

[8] Work done by PhD student Tahl Zimmerman at the CNIO

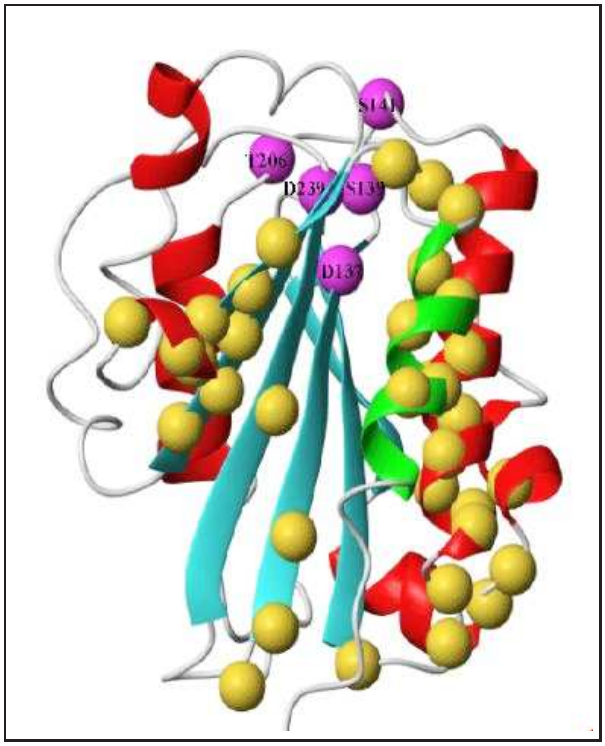


Fig. 6.15: Cartoon representation of the wild type αL I-domain. The Cα atoms of the residues of the MIDAS motif are highlighted as magenta spheres, whereas the Cα atoms of the residues that underwent significant chemical shift perturbation upon YFE02 addition are shown as yellow spheres. The C-terminal α7 helix is depicted in green.

In-house designed compounds do not bind to the MIDAS motif of the mutant αL I-domain: The original idea when the mutant αL I-domain was purified was to check the binding of in-house compounds to the *open* MIDAS motif that was expected to be present in the mutant protein. However, as explained above, the MIDAS motif was found to be in the *closed* conformation regardless of the position of the C-terminal α7 helix. Nevertheless, the binding of the compounds to this mutant conformer was tested in order to find out whether our compounds would be now able to interact with the metal in the *closed* MIDAS motif and provoke the closed-to-open conformational change. According to Figure 6.12, the C-terminal α7 helix may have undergone the desired (or expected) conformational change, priming (preparing) the I-domain for receiving the ligand. If the ligand were ICAM-1, and according to the crystal structures published to date, it seems reasonable to expect that the closed-to-open conformational change would occur. Do the small molecular weight compounds induce the same change in the conformation of the MIDAS motif? The answer to this question is found in the paragraphs below.

^{15}N-HSQC spectra of the active mutant in the presence and absence of compound YFE02, YFE05, YFE06, YFE10, YFE11, YFE12 and HC0303 were acquired and results compared to those obtained with the wild type αL I-domain (Figure 6.13). In Figure 6.16 the results obtained with compound YFE02 are shown, since equivalent results were again obtained with the rest of the compounds tested. The binding profile of YFE02 to the mutant αL I-domain resulted to be very similar to that of its binding to the wild type I-domain (Figure 6.14). As observed, the residues affected by the presence of YFE02 were roughly the same in both cases

but the chemical shift perturbation of these residues was much smaller in the case of the active mutant that in the case of the wild type I-domain (compare scales of Y axis in both plots). One of the hypotheses that could explain these differences is based on the presence of the disulfide bond in the mutant αL I-domain. The disulfide bond is formed between residues 287 and 294 and as observed in Figure 6.12, it sits right in the center of the IDAS site. It is thus easy to understand that the binding of the compounds to the IDAS site in the active mutant αL I-domain may be weaker than their binding to the wild type I-domain. Nevertheless, the presence of the disulfide bond does not completely block the entrance of the compounds, which still prefer this highly hydrophobic site to the MIDAS motif.

As previously mentioned, the backbone of the active form of the αL I-domain was not completely assigned yet by the time this thesis project was written but, instead, reasonable assumptions were made upon overlaying the wild type assignment onto the active form spectrum, as in Figure 6.11. For most of the residues, this approximation provided an adequate estimation of their chemical shifts. Nevertheless, these results need to be considered as preliminary, until a rigorous assignment of the mutant αL I-domain is performed.

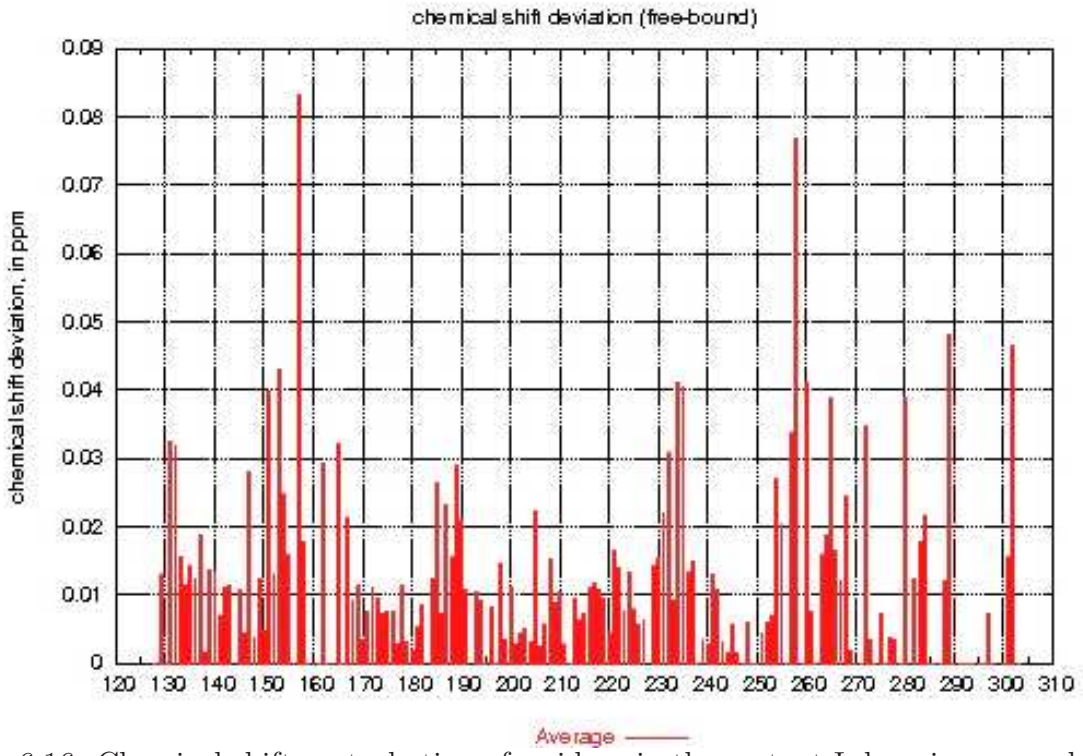


Fig. 6.16: Chemical shift perturbation of residues in the mutant I-domain upon addition of compound YFE02.

YFE02 binds exclusively to the αL I-domain: To make sure that compound YFEO2 had certain specificity for the I-domain, its binding was also tested against the I-domain of integrin α2 and the 6kD protein G. In both cases YFE02 showed no binding at all, as no signal of the reference ^{15}N-HSQC spectra of these proteins was perturbed in the presence of compound (data not shown).

Estimation of the binding constant

The binding is in the micromolar range: In order to estimate the Kd of YFE02 and HC0303 binding to the IDAS site in the αL I-domain, CSP experiments were performed adding increasing amounts of compound to a fixed amount of singly labeled wild type αL I-domain. The signal of residue T291 was used to calculate the chemical shift deviation at different concentrations of compounds. As shown in Figure 6.13 this residue displayed an adequate behavior to measure binding, since its chemical shift perturbation values upon addition of increasing amounts of compound were measurable with a significant reliability. As observed in Figure 6.17, this signal displayed a measurable chemical shift perturbation pattern. The chemical shift perturbation values were plotted against the compound concentration, getting the curve shown in Figure 6.18. As shown, the binding constants for YFE02 and HC0303 were found to be in the low micromolar range, 7.43 micromolar and 10.05 micromolar, respectively.

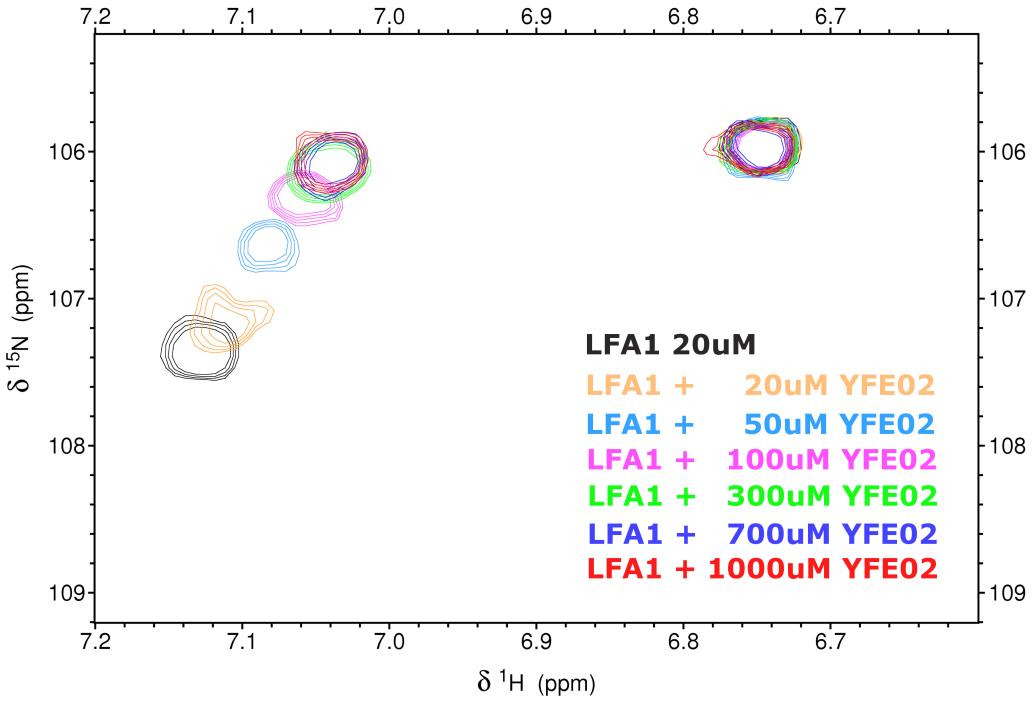


Fig. 6.17: Fraction of the seven ^{15N}-HSQC spectra obtained and overlaid in the CSP experiments with the wild type αL I-domain and compound YFE02. The signal on the left corresponds to residue T291, which displayed the highest chemical shift perturbation values, and was used in the CSP measurements. Right signal: Residue 246, as an example of a signal in the ^{15N}-HSQC spectra of the wild type αL I-domain that did not display any significant chemical shift perturbation upon addition of compound YFE02. Almost identical results were obtained with compound HC0303 (data not shown).

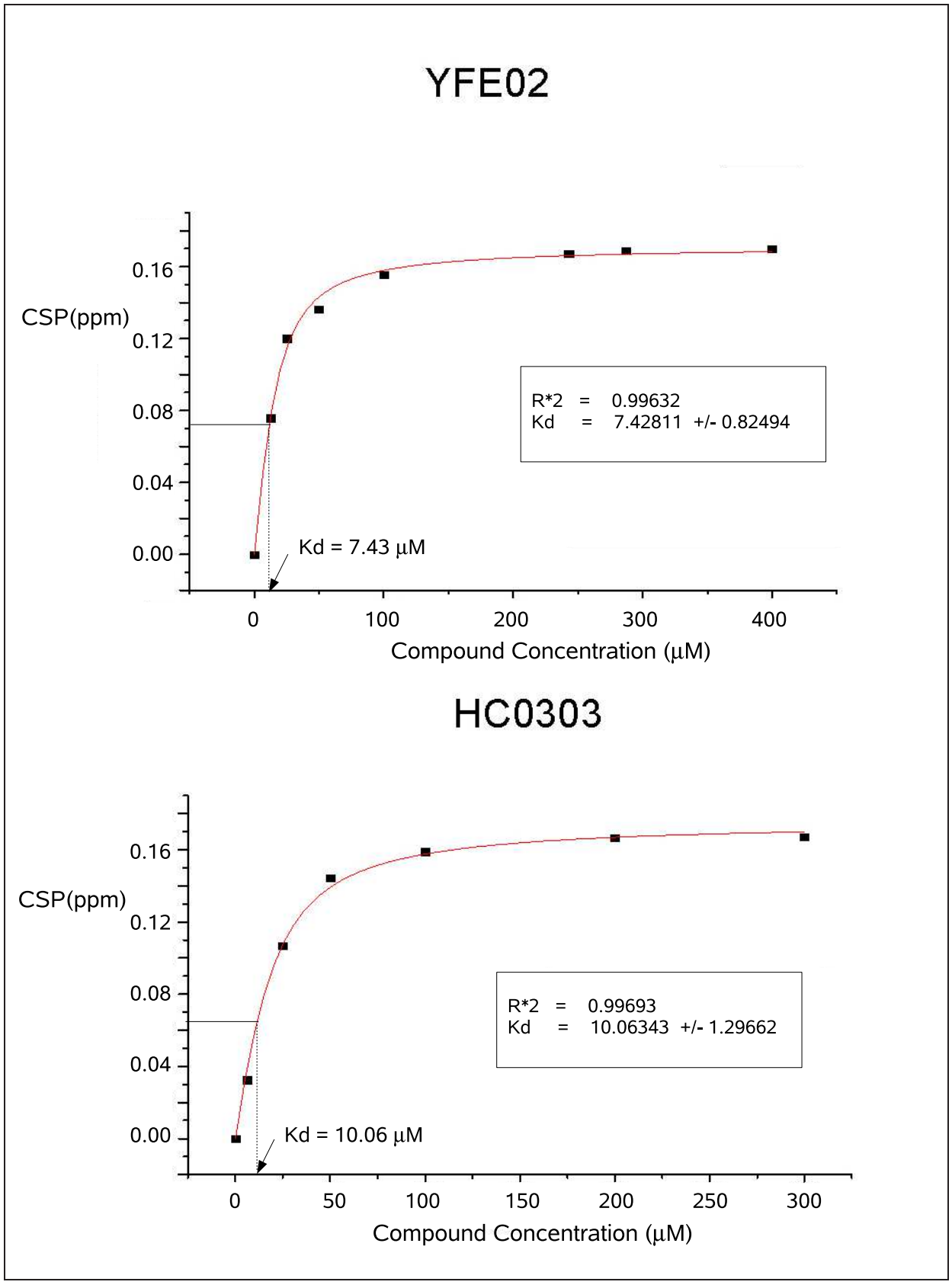


Fig. 6.18: Titration curves of compounds YFE02 (top) and HC0303 (bottom). The CSP values were plotted against compound concentration.

6.3.3 Surface plasmon resonance (SPR) experiments

SPR experiments were carried out with a twofold goal in mind: i) as a control of the quality of the purified constructs, their binding constants to ICAM-1 (Domain1 to Domain5) were measured and compared to the reported ones[60] and ii) to estimate the capacity of our compounds to inhibit this interaction. To accomplish the first goal, ICAM-1 was immobilized in a chip and either the wild type or the mutant αL I-domain were passed over it while binding was measured. To accomplish the second goal, the first set of experiments was repeated in the presence of compound in the solution containing the I-domain. The results of these two types of experiments are summarized in the lines below.

The wild type I-domain forms a low affinity complex with ICAM-1: As a first control, the binding of either the wild type or active mutant I-domain to the immobilized ICAM-1 was measured. To do so, ICAM-1 (D1-D5) was immobilized on a chip, and solutions containing mutant or wild-type αL I-domain were passed through the chip. In Figure 6.19, the signals recorded are gathered, corresponding to the wild-type (brown) and the mutant (blue) I-domains. The mutant I-domain showed a relatively strong binding to ICAM-1, compared to the binding of the wild type I-domain to ICAM-1, which could not be measured in a reliable way, despite the fact that the concentration of the wild-type construct used was up to 4 times higher than the concentration of the mutant I-domain. As observed, the binding was still too weak to be detected by means of this technique, which, in addition, is not highly sensitive. Nevertheless, it is clear from the shape and relative magnitudes of these two curves, that the wild-type I-domain binds with very low affinity to ICAM-1, whereas the mutant I-domain binds to ICAM-1 with much higher affinity.

The mutant αL I-domain forms a metal dependent complex with ICAM-1: In order to check that the observed binding of the mutant αL I-domain was achieved via the MIDAS motif, the experiments above were repeated in the presence of ethylenediamine tetraacetic acid (EDTA hereafter), an agent known to chelate cations such as Mg^{2+}. If the MIDAS motif were involved in the binding to ICAM-1, the presence of EDTA should disrupt it. In Figure 6.19, the light-blue line represents the binding of the mutant αL I-domain to ICAM-1 in the presence of EDTA. As observed, the binding has been completely abolished. Therefore, the binding of the mutant αL I-domain to ICAM-1 is done via the MIDAS motif, as expected. This result suggests, that the purified construct is equivalent to that reported in the literature, being these results comparable to those of others.

The binding constant of the mutant αL I-domain to ICAM-1 is consistent with that reported in the literature: As a second control of the quality of the purified constructs, the Kd the active mutant binding to ICAM-1 was estimated and compared to the results reported by other authors [60]. To do so, increasing concentrations of the mutant αL I-domain were passed through the chip were ICAM-1 was immobilized. A time dependent signal was recorded for each concentration, and the results are plotted in Figure 6.21.

The value of the maxima of each curve can be plotted against the concentration of mutant I-domain, which yield a curve from which the binding Kd was estimated to be around 0.69 μM. Other authors reported a kd of 0.15 μM [60] using the same technique. The slightly different values may be related to the fact that in this study ICAM-1 contained 5 domains (D1-D5), whereas in the mentioned study it only consisted of domains 1 and 2 (D1-D2).

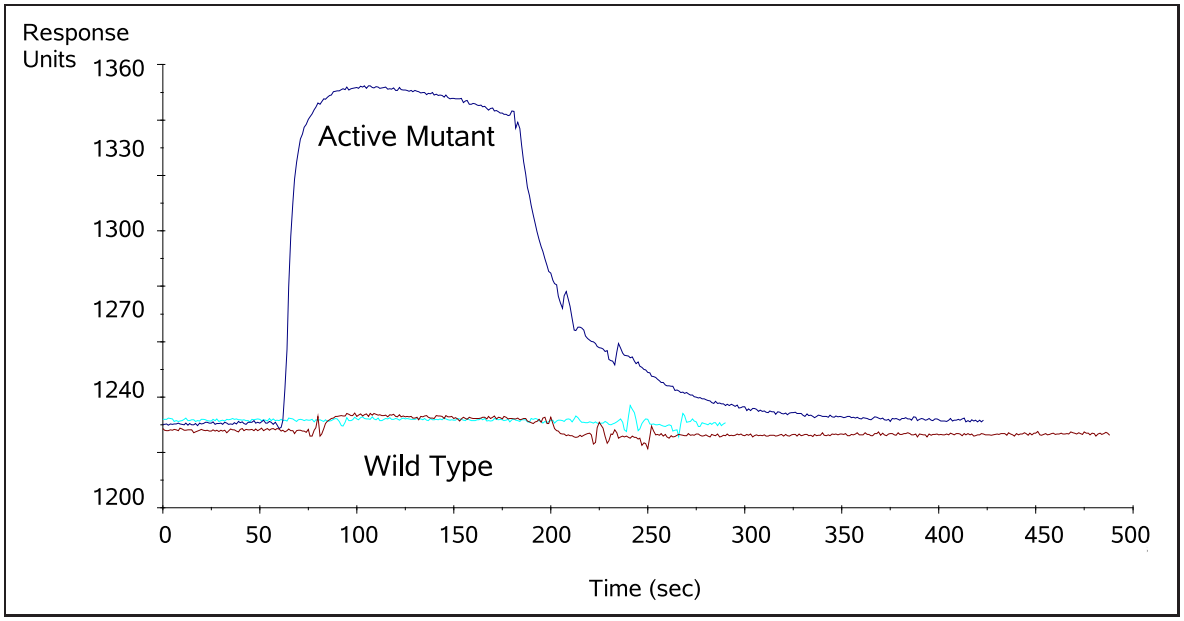


Fig. 6.19: Surface Plasmon Resonance Sensogram. Immobilized ICAM-1 binding to the wild type I-domain (brown), the active mutant I-domain (blue) and the active mutant I-domain in the presence of EDTA (light blue). Concentration of wild type I-domain = Four times concentration of the active mutant I-domain.

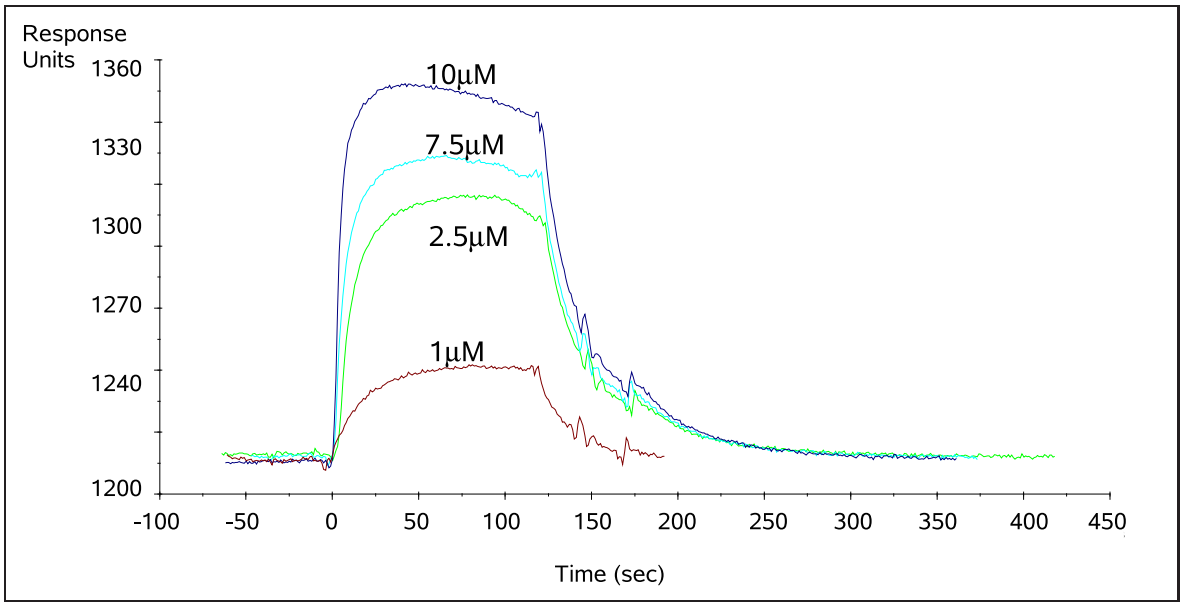


Fig. 6.20: Surface Plasmon Resonance Sensograms. Different concentrations of the mutant αL I-domain were passed through a chip that contained a fixed concentration of immobilized ICAM-1 (D1D5). From the maxima of these curves, the kd of binding was estimated as shown in Figure 6.21.

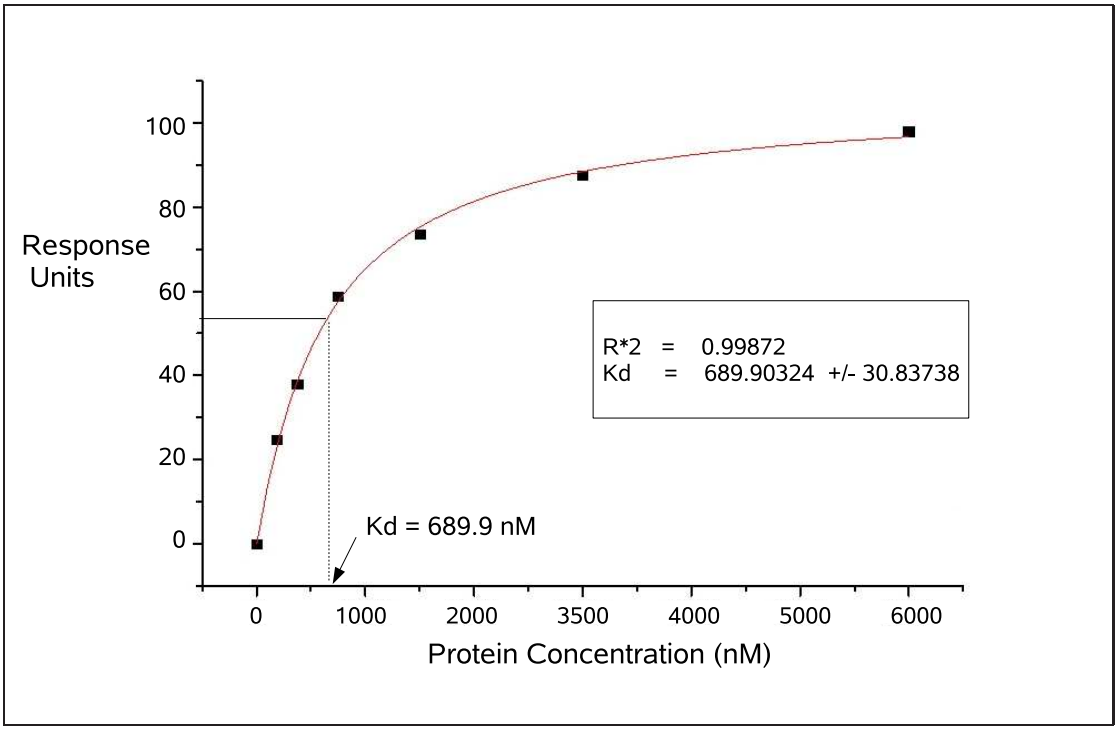


Fig. 6.21: Estimation of the Kd of the binding of the mutant I-domain to ICAM-1. Kd= 0.69uM.

Our compounds do not inhibit mutant I-domain binding to ICAM-1: As mentioned earlier in the text, the ability of our compounds to inhibit adhesion of the LFA-1 heterodimer to ICAM-1 was successfully tested in *in vivo* assays and in cell-free and cell-based *in vitro* adhesion assays. Later, during this thesis project, it has been found that in the isolated wild type or isolated active mutant I-domains, the allosteric site underneath the C-terminal $\alpha 7$ helix is the binding pocket for our compounds. Linking these two type of results together, it could be assumed that our compounds behave as allosteric inhibitors. This conclusion, however, would have been reached indirectly, since the biological unit used differs from one experiment to the other. It was therefore decided to perform an **inhibition** experiment with our compounds and the isolated αL I-domain. If inhibition of binding were observed, the allosteric character of our inhibitors would be confirmed.

With this goal in mind, we repeated the experiments in Figure 6.19 with the mutant I-domain, in the presence and absence of compounds YFE02 and HC0303. The low signal observed with this technique in response to wild type I-domain binding to ICAM-1, did not allow us to check the capacity of our compounds to inhibit this low affinity interaction. In Figure 6.22, shown is the SPR time dependent response recorded upon having flown the mutant αL I-domain over the immobilized ICAM-1, in the absence (red) and presence (black) of compound YFE02. As observed, YFE02 is not able to inhibit the interaction, despite the fact that we know that this compound binds to the allosteric site of the mutant I-domain. The results with compounds HC0303 were equivalent. I.e, apparently, despite the fact that our compounds bind to the isolated mutant αL I-domain, this binding does not inhibit the interaction of the mutant αL I-domain with the immobilized ICAM-1.

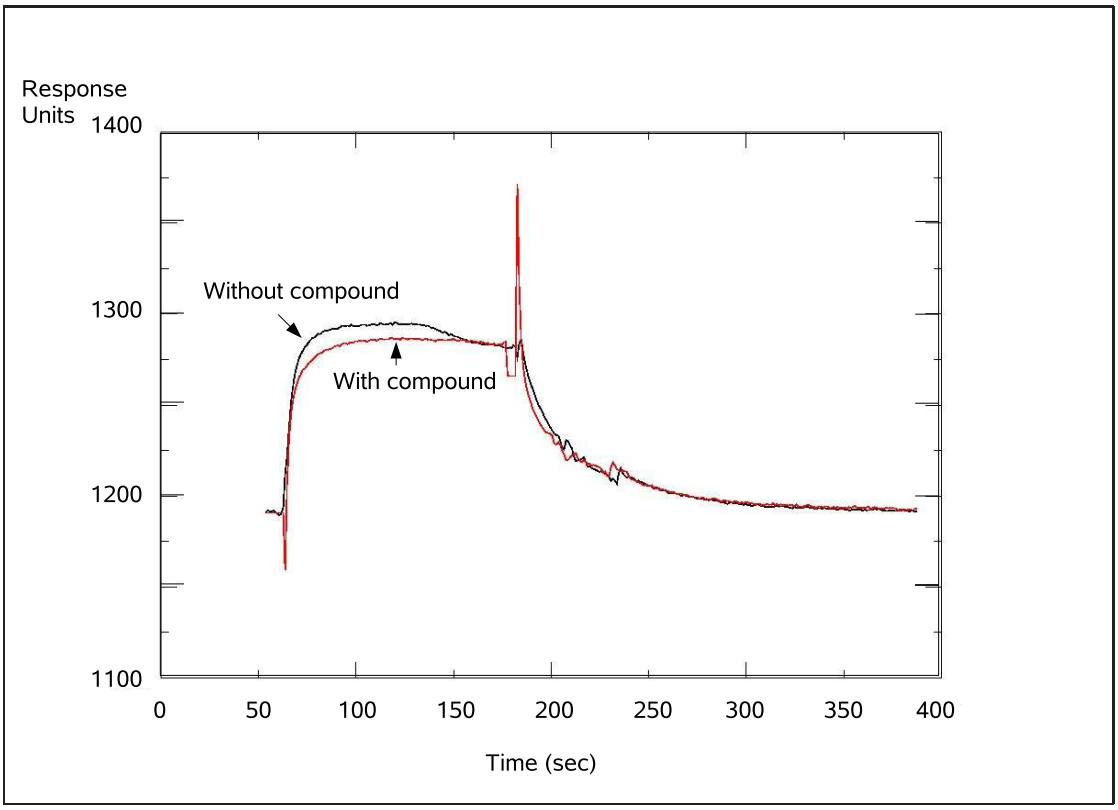


Fig. 6.22: Surface Plasmon Resonance Sensogram in the presence and absence of compound YFE02.

These results represent a new evidence for the hypothesis to be tested, i.e., that the isolated αL I-domain is not the appropriate biological unit for measuring either binding or inhibition capacity of low molecular weight compounds.

6.4 Discussion: The isolated αL I-domain

According to the backbone assignment of the wild type I-domain and the preliminary observations with the spectrum of the mutant I-domain, we can conclude that in the latter the MIDAS motif is in the *closed* conformation. This is not in agreement with the idea of the existence of a link between the position of the C-terminal α7 helix and the coordination mode of the metal. We suggest for the fist time that, in the absence of ICAM-1 or a pseudoligand, the disulfide bond is able to promote the downward displacement of the helix, but not the change in the coordination of the metal. In addition, the crystal structures of the medium affinity I-domain in complex with ICAM-1 (PDB ID=1MQ8 [60]), the high affinity I-domain in complex with a pseudoligand (PDB ID=1MQ9 [60]), and the high affinity I-domain in complex with ICAM-3 (PDB ID=1T0P [78]) display a MIDAS motif in the *open* conformation. These structures also suggest that only the presence of ICAM-1 is able to provoke the close-to-open conformational change in the MIDAS motif of an isolated αL I-domain primed with a disulfide bond.

The binding experiments between our compounds and the isolated wild type or active mutant αL I-domain showed that both the VLA-4 antagonists and LFA-1 antagonists bind to the same place underneath the C-terminal α7 helix, known as the I-domain allosteric site or IDAS site. The binding constant of the LFA-1 antagonists, such as compounds YFE02 and HC0303, to this allosteric site in the wild type I-domain was found to be 7.43 uM and 10.06 uM, respectively. On the other hand, these compounds were not able to inhibit the binding of the active mutant αL I-domain to immobilized ICAM-1. These results are not consistent with the *in vitro* and *in vivo* results obtained with the intact heterodimer, which revealed that compounds designed to inhibit the VLA-4 dependent adhesion, such as HC0225, were shown to be unable to inhibit the interaction between LFA-1 and ICAM-1. On the contrary, according to the NMR results, HC0225 binds to the allosteric site in the wild type and active mutant I-domain in a way comparable to the binding of YFE02 or HC0303, known to inhibit the LFA-1 dependent adhesion.

Since the 15N HSQC spectrum of the wild type I-domain is equivalent to that reported by Huth and coworkers [178] and the binding constant of the mutant I-domain to immobilized ICAM-1 is comparable to that obtained by Shimaoka and coworkers [60], it is reasonable to think that the proteins purified in this work are viable. In addition, the NMR and SPR experiments were successfully performed. Consequently we believe that the answer to the inconsistencies might lay in the fact that the isolated αL I-domain may not properly represent the ICAM-1 binding domain in the intact integrin. An important support to this hypothesis was found after having carefully analyzed the works published by other authors, which are summarized below.

6.5 Reviewing the Literature

A through reading of the literature concerning the chemical nature, structure and binding site of small molecular weight inhibitors of the LFA-1/ICAM-1 interaction reveals a controversy regarding the conclusions reached by different authors. These conclusions depended not only on the type of assays chosen to analyze the inhibitory activity of these molecules, but on the nature of the biological unit used to "represent" the behavior of the *in vivo* intact LFA-1 heterodimer. Consequently, some authors identified the binding site of a specific compound in an **isolated** αL I-domain, while other authors showed that, for the same compound as before, results of experiments performed with either the **intact heterodimer** in cell free assays or with cells expressing the **intact heterodimer**, suggest an alternative binding site. Furthermore, some authors mixed results obtained from binding assays (using the isolated I-domain) and results obtained from inhibition assays (using the intact heterodimer), which create confusion regarding the actual binding site of an inhibitor. We strongly believe that joining together both types of results to suggest that the binding site found in the isolated I-domain is the place were the compound bound in the inhibitory assays could be sometimes a dangerous source of ambiguity. In the following paragraphs, we will briefly describe the results published to date in relation to LFA-1/ICAM-1 inhibitors [9].

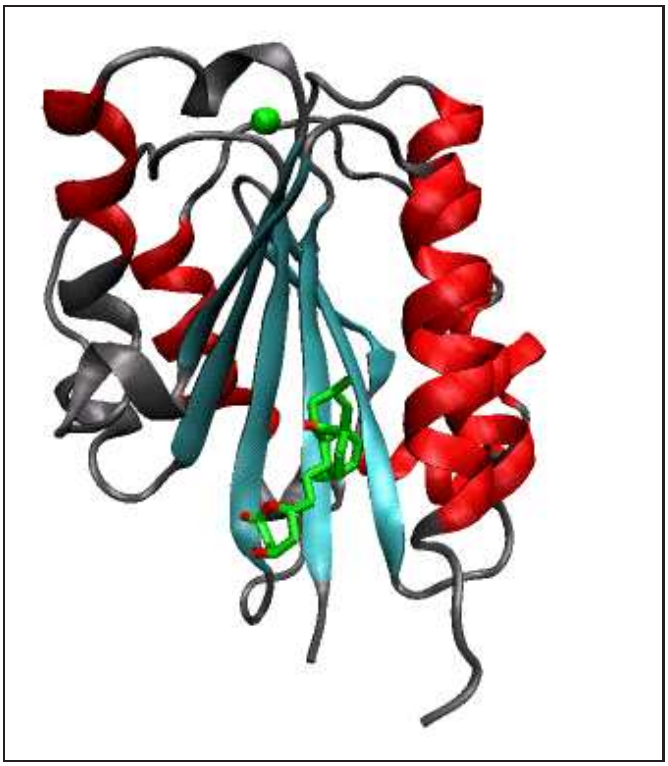


Fig. 6.23: Lovastatin inhibitors bound to the IDAS site of the wild type αL I-domain

[9] I based the summary on the review by Shimaoka et al.[260], and will try to highlight its possible weak points

6.5.1 The αI allosteric inhibitors

The molecules belonging to this family of inhibitors are believed to bind to the hydrophobic groove that exists underneath the C-terminal $\alpha 7$ helix of the isolated I-domain of the αL subunit of LFA-1 (Figure 6.23). The molecules that have been described to date as being members of this group are: LFA703 (Patented by Novartis), BIRT0377 (Patented by Boehringer-Ingelheim), A-286982 (Patented by ICOS/Abbott/BioGen), Lovastatin and LFA878.

According to Shimaoka and coworkers [260], authors such as Kallen and coworkers [42], Last-Barney and coworkers [261] and Liu and coworkers [262] demonstrate, based on crystallographic arguments, that the inhibitors mentioned above bind to the IDAS site, providing a structural basis for αI allosteric inhibitors. In the following lines, the work done by these authors is summarized, in order to try to highlight possible weak points.

o Statin derived inhibitors

Lovastatin: Kallen and coworkers [42] tried to locate the binding site of a molecule called lovastatin, both in its lactone and hydroxyacid form (Figure 6.24, left and right, respectively), using the isolated αL I-domain as the biological unit for LFA-1. This molecule was identified as a possible inhibitor after a high throughput screening experiment. Under physiological conditions, lactone ring of lovastatin undergoes hydrolysis into the hydroxyacid form. The unique difference among them is the existence of a carboxylate group in the latter, that may provide a coordinating group to the metal of the MIDAS motif in the αL I-domain.

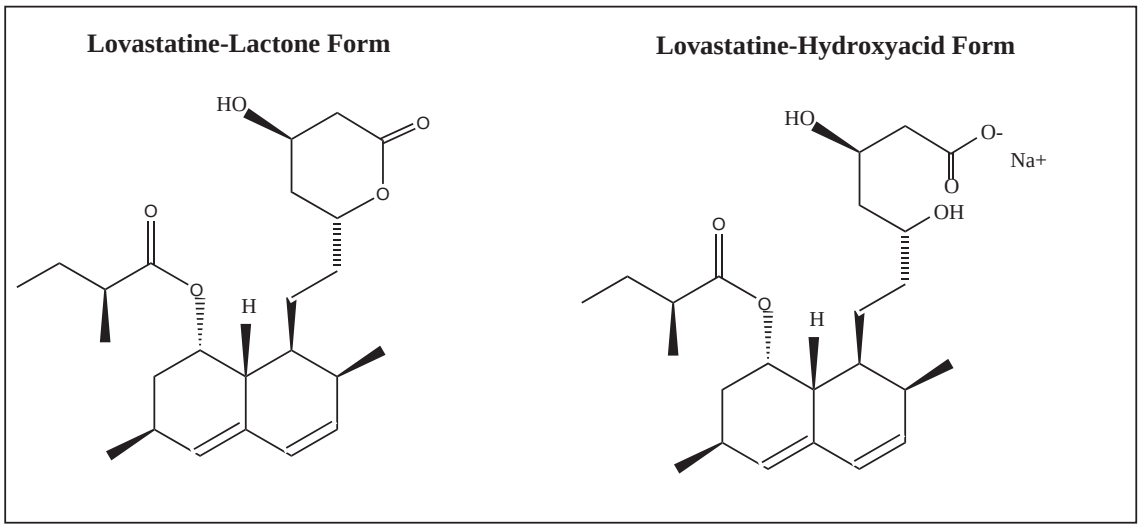


Fig. 6.24: Lactone (left) and Hydroxyacid (right) forms of lovastatin.

As inhibition experiments the authors performed cell-based assays where cells expressing the LFA-1 **heterodimer** (activated by addition of Mn^{2+} cation [263, 264, 265]) were added to wells containing immobilized ICAM-1, in the presence and absence of inhibitor. The lactone form of

lovastatin was able to inhibit adhesion four times better than the hydroxyacid form. Alternatively, in cell-free assays, the same authors immobilized the αLβ2 heterodimer, and binding to ICAM-1 was measured. Once again, the lactone form of lovastatin displayed the best inhibitory capacity among the two. Due to the existence of a carboxylic group in the hydroxyacid form of lovastatin (and the lack of it on the lactone form), the authors suggested that the inhibition observed could not be a consequence of any of the lovastatin forms binding to the MIDAS site of the αL I-domain as, and according to the authors, if this were to be the binding pocket for both type of molecules, the hydroxyacid form would have shown a higher inhibitory activity.

We strongly believe that, despite the fact that this hypothesis is properly based on the structural features of ICAM-1, the experiments do not directly prove that the inhibition is due to lovastatin binding somewhere different to the MIDAS motif in the αL I-domain. Further experimentation was still needed at this point.

To check that their assumptions were true, Kallen et al.[42] purified the wild type αL I-domain and observed, both by NMR experiments and X-ray crystallography, that both the lactone and the hydroxyacid form of lovastatin bind to the allosteric site in the αL I-domain or IDAS site (Figure 6.23). In the NMR experiments, the two forms of lovastatin showed no significant differences on the chemical shift perturbation analysis, suggesting no big differences on the binding mode of both molecules, as opposed to the cell-based and cell-free assays with the intact heterodimer, where the lactone form showed to have a stronger inhibitory capacity. Similarly, and paraphrasing the authors, "the X-ray structure of the I-domain/lovastatin complex does not provide a clear answer" to the question of why does the lactone form of lovastatin represent a better inhibitor than the hydroxyacid form.

In summary, Kallen and coworkers performed **inhibition** experiments with the *intact heterodimer*, but used the *isolated αL I-domain* to locate the **binding site** of lovastatin molecules. Competition experiments with the intact integrin or inhibition experiments with the isolated I-domain would be of great interest. Hence, we think that it is still not completely clear whether lovastatin is an allosteric inhibitor or a competitive inhibitor. In addition, Kallen and coworkers also state that lovastatin may bind to other regions of the heterodimer, apart from the IDAS site.

o Hyndantoin based LFA-1 antagonists

BIRT 377: Last Barney and coworkers [261] concluded, after having performed photoaffinity labeling experiments [10] with the isolated αL I-domain, that the hydantoin based LFA-1 antagonist BIRT377 (Figure 6.25), binds to the allosteric site in the αL I-domain (the IDAS site).

BIRT377 is a lipophilic compound that had previously been shown by Kelly and coworkers [266] to be a potent antagonist of LFA-1 mediated cell adhesion. BIRT377 was not designed by mimicking any epitope in ICAM-1 and its binding site was completely unknown. By means of

[10]Photoaffinity labeling experiments consist on the substitution of selected amino acids of proteins by analogue photoactive moieties that are activated upon exposure to ultraviolet light. This technique permits the identification of protein-nucleotide interactions at the molecular level.

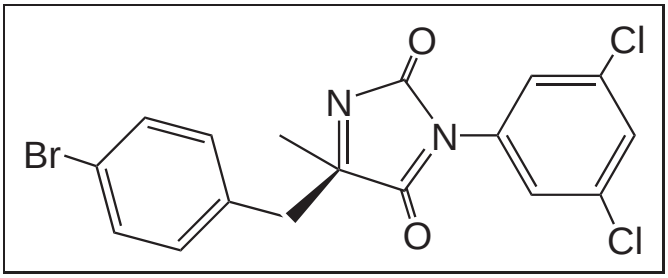


Fig. 6.25: Hyndantoin based inhibitor BIRT377

cell-based and cell-free assays, Kelly and coworkers observed that BIRT377 was able to inhibit the interaction between the intact LFA-1 heterodimer and ICAM-1, but was not able to inhibit the interaction between the intact MAC-1 heterodimer and ICAM-1, or VLA-4 to immobilized fibronectin. Both LFA-1 and MAC-1 integrins share a $\beta 2$ subunit, and differ on the type of α subunit (αL or αM, respectively).These experiments suggest that BIRT377 inhibits LFA-1 dependent adhesion by binding somewhere in the αL subunit. Experiments with a set of monoclonal antibodies (mAbs hereafter) and the intact LFA-1 heterodimer, suggested that the binding pocket of this compound might be located either in the αL I-domain of the integrin, or in a remote allosteric site of the α subunit. Nevertheless, both α subunits contain an I-domain, contrary to VLA-4.

In order to specifically locate the binding site of BIRT377 in the αL subunit, two years later, Woska and coworkers[267] measured binding of the monoclonal antibody (mAb hereafter) mAb24 to LFA-1. This mAb is known to bind uniquely to the high affinity conformation of the I-domain. They observed that, in the presence of BIRT377, mAb24 did not bind to LFA-1. The authors proposed that this was due to BIRT377 binding to the allosteric site underneath the C-terminal $\alpha 7$ helix in the αL I-domain (IDAS site). The binding of BIRT377 would avoid the low affinity to high affinity conformational change of the I-domain.

In the same year, Last-Barney et al.[261] concluded, by means of photoaffinity labeling experiments using the isolated αL I-domain, that BIRT377 binds to the IDAS site. As Kelly et al. had previously shown that BIRT377 was a good *in vivo* and *in vitro* LFA-1/ICAM-1 inhibitor, Last-Barney and coworkers concluded that BIRT377 is an αL allosteric inhibitor that blocks LFA-1 dependent adhesion by binding to the IDAS site.

LFA878 and LFA703: Weitz-Schmidt et al.[268], cocrystallized the isolated wild type I-domain and the statin derivative compound LFA878 (Figure 6.27), and found that it binds to a subpocket in the IDAS site of the isolated αL I-domain. Nevertheless, this subpocket in the IDAS site had not been previously described and was different from that of Lovastatin, that was also cocrystallized in this work bound to the isolated wild-type I-domain.

They also performed mAb binding assays and observed that a second compound, LFA703, can modify the binding pattern of some mAbs known to bind both to the αL I-domain and to the $\beta 2$ I-like domain. On the contrary, LFA878 only varied the binding pattern of mAb known to bind to the β I-like domain. They explain the different behavior between both compounds by assuming that LFA703 binds the IDAS site in such a way that it also blocks mAb binding to the nearby I-like domain. On the contrary, they propose that LFA878 binds to the IDAS site in such a way that the $\beta 2$ I-like domain stays unaffected.

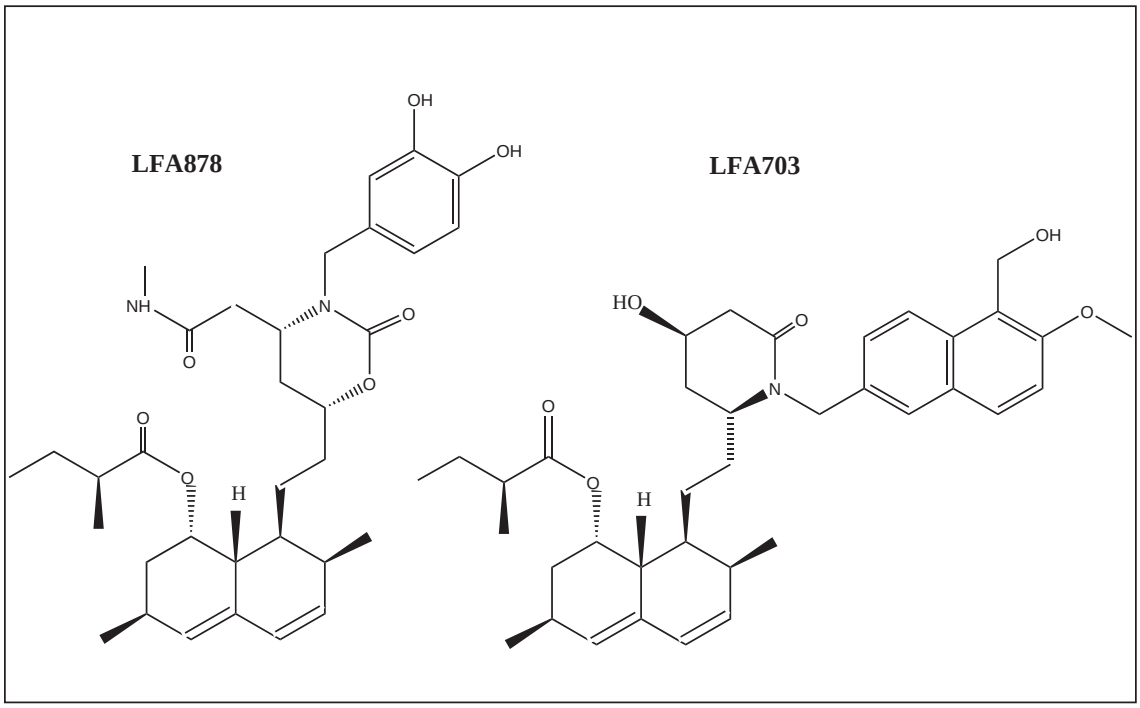


Fig. 6.26: Hyndantoin based inhibitors: LFA878 (left) and LFA703 (right).

As in the previous example, it seems pretty reasonable to propose further experimentation in order get to unambiguous conclusions.

o p-arylthiocinnamide-based inhibitors

Liu and coworkers [262] used NMR techniques to locate the binding site of p-arylthiocinnamide-based compounds in the isolated αL I-domain, which was found to be, once again, the IDAS site (Kd=20uM) (NMR). This compound was the best hit of a systematic search for leads to inhibit the binding of full length LFA-1 to ICAM-1; this study was performed by the same authors[269], being the binding Kd in this case 2.3uM. According to the authors, the Kd was pretty similar to that found by NMR and, therefore, they concluded that the inhibition observed with the intact integrin was due to the binding of the compounds to the IDAS site. Again, direct and unambiguous proves of the latter are still missing.

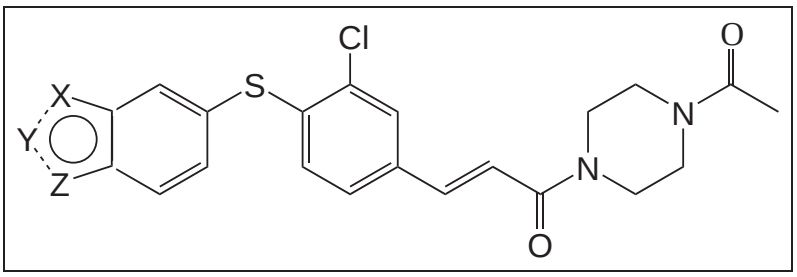


Fig. 6.27: p-arylthiocinnamide-based inhibitors

6.5.2 The αL and αM allosteric inhibitors

Compounds that belong to this family remain less defined with respect to their binding site and mode of action. Some of them have been patented as LFA-1 antagonist by Genentech (Burdick 1999, Burdick 2001) or dual-acting LFA-1 and MAC-1 inhibitors[270] or LFA-1 antagonists[271] by Roche.

Gadek and coworkers [192] synthesized a set of compounds that mimic the electronic and geometric features of the residues that ICAM-1 uses to bind to LFA-1. The inhibitory activity of these compounds was tested by performing *in vivo* assays and *in vitro* cell assays where the integrin was expressed as an intact heterodimer. Results showed that the compounds were excellent inhibitors of the LFA-1/ICAM-1 interaction. Out of these results, the authors concluded that these compounds may bind to the MIDAS site in the αL I-domain. This competitive inhibitory capacity, though, was not directly tested yet at this point. In addition, they performed Enzyme-Linked Immunosorbent Assay experiments (ELISA hereafter) [11] were adhesion of cells expressing the full length integrin to wells containing immobilized ICAM-1 molecules was measured, demonstrating that inhibition was due to the blockade of the LFA-1/ICAM-1 interaction; i.e., the specificity of these compound as LFA-1 antagonists was tested. Compound 3c was able to bind to LFA-1 but not to MAC-1. These two integrins differ in the α subunit, but they both share a common β2 subunit. This specificity indicates that compound 3 might not bind to the β I-like domain in LFA-1, as suggested later by Shimaoka and coworkers [272]. On the contrary, the presence of an I-domain in the αL and αM subunits, points to compound 3 binding to the αL I-domain and not the αM I-domain. In summary, by the time this work by Gadek and coworkers was published [192], it was not unambiguously known whether compound 3 binds to the MIDAS motif (despite the fact that it mimics ICAM-1) or if compound 3 binds to the hydrophobic groove underneath the C-terminal α helix, or somewhere else.

Welzenbach and coworkers [273] studied the behavior of the structurally related dual αLβ2 and αMβ2 inhibitor from Roche, shown in Figure 6.28-XVA143, instead of studying the behavior of compound 3 by Gadek and coworkers. They showed that XVA143 did not perturb NMR chemical shifts of isolated αL or αM I domains, but did perturb antibody mediated binding to the β2 I-like domain in LFA-1 and MAC-1 holoreceptors. It was therefore suggested that these inhibitors represent a novel class of antagonists of I domain-containing integrins, which bind to the β2 I-like domain MIDAS motif near a key regulatory interface with the αL and αM subunits. Upon binding to that site, this inhibitors would apparently block the communication of a conformational change to the I domain, while at the same time activating conformational rearrangements elsewhere in integrins. All these conclusion obtained with compound XVA143 were extrapolated to Gadek's compound 3.

[11] ELISA experiments can be used to measure binding between two proteins, by combining the specificity of antibodies with the sensitivity of simple enzyme assays. Antibodies or antigens are coupled to one of the proteins of interest, which is also coupled to an easily assayed enzyme that possesses a high turnover number.

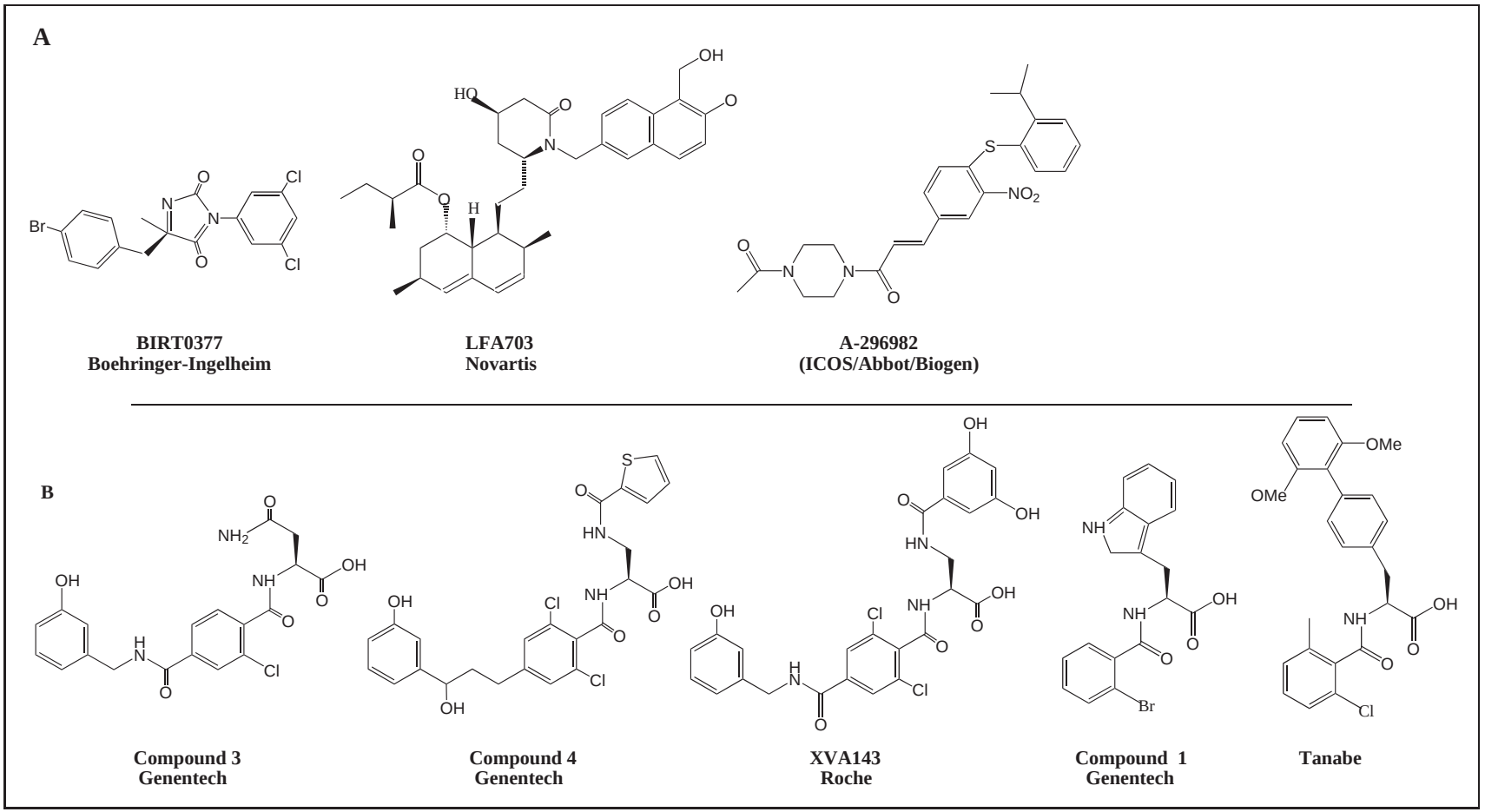


Fig. 6.28: Top) αL allosteric inhibitors and Bottom) αL-αM allosteric inhibitors (According to Shimaoka and coworkers [260]).

Later, Shimaoka and coworkers published a report [272] which strongly opposed Gadek's characterization of compound 3 [192]. In this report, Shimaoka and coworkers [272] summarized the results obtained with different inhibition experiments, where instead of the intact αLβ2 heterodimer, the isolated αL I-domain was used as biological unit. In this type of assays, Gadek's compound 3, which had already been patented by Roche, displayed no inhibitory activity on the adhesion of isolated αL I-domain to immobilized ICAM-1. Alternatively, they also performed binding assays (note that inhibition assays were not performed) of Gadek's compounds to the full length integrin. Shimaoka and coworkers [272] observed that Gadek's compounds were able to stabilize the interaction between the αL and β2 subunits. Shimaoka and coworkers therefore concluded that Gadek's compound 3, instead of an αL competitive inhibitor, is a β2 allosteric inhibitor.

In view of Shimaoka's report[272], Keating and coworkers[274] (together with Dr. Gadek), repeated some of the experiments performed by Shimaoka and coworkers[272], concluding that some of the results obtained by Dr. Shimaoka were overrated. In fact, Keating and coworkers[274] also observed binding of compound 3 to the β2 subunit of the LFA-1 heterodimer, but much weaker than that suggested in [272]. Interestingly, this binding to the β2 subunit was only observed when working with the intact heterodimer, and not when working with the isolated β2 domain. In view of these results, Keating and coworkers [274] modified their previous characterization of compound 3, adding that the latter was able to bind both to the αL I-domain and to the β2 I-like domain. To further check the competitive character of their compounds, Keating and coworkers[274] performed competition experiments with the intact heterodimer and compound 3 and showed that the binding of compound 3 was competitive with ICAM-1 and cation dependent, which suggests that compound 3 binds to a metal binding site, such as the MIDAS motif in the αL subunit and/or the putative MIDAS motif on the β2 I-like domain. It was therefore suggested that compound 3 binds to the MIDAS motif of the αL I-domain of the intact heterodimer. In this sense, there was no cross linking of compound 3 with the isolated α I-domain. Again, there is a hint here that points to the need of the using the intact αLβ2 heterodimer in order to properly characterize the binding of small molecular weight inhibitors to LFA-1.

Chapter 7

General conclusions.

Contents

7.1 Conclusions. Chapter by Chapter

The motivation for this study arose from the fact that integrins VLA-4 and integrin LFA-1 have lately become targets of the treatments of autoimmune and inflammatory diseases. A therapy to fight these pathologies consists of the oral administration of low molecular weight compounds that bind specific sites of the integrins. Previous members of the Organic Chemistry Group I, designed and synthesized promising families of LFA-1 and VLA-4 antagonists, that showed excellent *in vitro* and *in vivo* antimetastasic activity.

This thesis project was focused on the understanding of the main aspects of the interaction between integrin LFA-1 and its counterreceptor ICAM-1. The knowledge acquired was applied to the design of an improved family of small molecular weight LFA-1 antagonists. These results are particularly interesting in the therapeutic treatment of metastasis of gastrointestinal carcinoma [96, 97], melanoma [98, 99, 100], lymphoma [101, 102, 103, 104] and colon cancer [105]. In the following paragraphs, the results obtained and the conclusions reached in this thesis work are summarized.

Chapter 3. Quantum Mechanics: A DFT study on the MIDAS motif

The divalent cation (Mg^{2+}/Mn^{2+}) present in the Metal Ion Dependent Adhesion Site (MIDAS) of the ligand binding domain (I-domain) of integrin LFA-1 binds negatively charged residues in extracellular ligands (ICAM-1 and ICAM-3). Since integrins possess several ligand-affinity modulating mechanisms, it has been proved difficult under experimental conditions to determine the relative contributions of different individual phenomena on the affinity of the MIDAS motif for ligands. The Density Functional Theory studies carried out in this thesis project provided clarifying information with respect to the affinity modulating mechanism on integrins.

Affinity regulated by *closed*-to-*open* conformational changes in the MIDAS motif.

It has been quantified for the first time the change (gain) in ligand affinity related to the *closed* to *open* conformational change in the MIDAS motif, observing that the *open* MIDAS binds glutamate-type residues with a higher affinity than the *closed* MIDAS motif. A major finding in this respect is that the surroundings of the MIDAS motif contribute to the regulation of the affinity towards ligand as much as or to a higher extent than the conformational change mentioned above. In this respect, the inclusion of second shell ligands in the MIDAS model and the increase in the polarity around the MIDAS and the approaching acetate molecule, dramatically decrease the strength of this interaction. This behavior helps to understand why the MIDAS motif has only been observed in its *open* state once bound to ICAM-1, and why the Mg^{2+} ion in the MIDAS motif and the anionic E34 residue in ICAM-1 are mainly surrounded by hydrophobic residues in the protein.

Replacement of cations in the MIDAS motif.

Our results indicate an interesting dependence of metal exchange energy on the conformation of the MIDAS and nature of the metal. Studies were performed with biologically relevant and experimentally tested cations: Zn^{+2}, Ca^{+2}, Mg^{+2} and Mn^{+2}. In the case of Zn^{+2}, regardless the conformation of the MIDAS, *closed* or *open*, or of the polarity of the environment, exchange was found to be always favorable. On the contrary, exchange possibilities for Ca^{2+} cation was found to strongly depend on the conformation of the MIDAS, being thermodynamically unfavorable for the *closed* MIDAS (for any dielectric), and favorable in the case of the *open* MIDAS conformation and intermediate permittivities. Last, the replacement of Mg^{+2} by Mn^{+2} is favored for both MIDAS conformations, and for either polar or more hydrophobic environments (except water).

Affinity regulated by the metal cation.

Our conclusions related to the effect of the different metals (Mg^{2+}, Zn^{2+}, Ca^{2+} and Mn^{2+}), help to clarify the controversial experimental results obtained by other authors for the role of these metals on ligand binding activity of integrins. This controversy may arise from the difficulty to isolate the intrinsic effect of the metals from many other factors that affect ligand binding activity in integrins. From our results, we observed that Mg^{2+} may be the most abundant but not necessarily the best cation for the purposes of entering the MIDAS and strongly binding to ICAM-1. Other ions, such as Zn^{2+}, and especially Mn^{2+}, could coordinate to the MIDAS motif and substantially favor ICAM-1 binding. On the other hand, too strong bindings are not necessarily desired from a physiological point of view, as there must eventually be a release of ligand at certain point of the biological process taking place. On the contrary, the inhibitory effect of Ca^{2+} cation may be due to the fact that it is not easily accommodated in an octahedral geometry of the MIDAS and its binding may lead to a structural distortion of the MIDAS and the surrounding environment.

Synthesis of new inhibitors.

Concerning the synthesis of new inhibitors of the LFA-1/ICAM-1 interaction, we concluded that carboxylic acids included in the metal binding moiety would improve the affinity for the LFA-1 I-domain, whereas electron withdrawing groups (alfa-halo carboxy donors) or aromatic rings (benzoates or heteroaromatic analogs) should be avoided because they lead to lower binding affinities. On the contrary, nitronates surrounded by hydrophobic groups, may yield binding affinities as good as or even better than the carboxylic acids.

Chapter 4. DFT vs QMC: Combined study on Mn^{2+}

Results summarized in Chapter 3 suggest that Mn^{2+}, when inserted in the MIDAS motif, is the most efficient cation to bind to extracellular ligands. Due to the transition metal character of this ion, the above results were refined with a more accurate methodology, such as the Diffusion Quantum Monte Carlo method, which is better than the Density Functional Theory method at estimating electronic energies of transition metal complexes such as the one under study. The goals of this work were threefold: i) to check using different high level theoretical methods whether the experimental assumption that Mn-protein complexes are high spin species is correct. ii) To see the behavior of B3LYP method in comparison to the more sophisticated DMC. iii) To check to which extent do the nature of the coordinated ligands unbalance the relative stability of high-spin complexes.

Multiplicity of Mn^{2+} in a protein metal binding site.

Our calculations strongly support that the sextuplet high-spin state of this ion is the lowest-energy spin state in protein environments. In most theoretical works, it had been always assumed, but never checked. However, the relative energies of the various spin-states (sextuplet, quadruplet and doublet), are highly sensitive to the ligand field provided by amino acid sidechains. In particular, there is a significant stabilization of doublet and quadruplet spin states with respect to sextuplet upon complexation with a negatively charged acetate ligand, which can represent the glutamate or aspartate sidechains.

Diffusion Monte Carlo versus Density Functional Theory.

From a methodology perspective, it is remarkable that B3LYP with a proper basis set is able to reproduce the qualitative trend of energies among the various spin states and their trends upon ligand-field stabilization, without the need of including second-shell ligands or counterions as suggested by other authors [121]. Nevertheless, it is also observed that B3LYP energy values for the bare Mn ion and for the complex with six neutral ligands are too low with respect to the experimental and DMC results.

Affinity dependence of the spin state.

Lower binding affinities were found for the higher spin states. In other words, high-spin Mn^{2+} complexes are more stable than low spin complexes, but they lose affinity for ligand. However, it is unlikely that Mn^{+2} undergoes a spin change to optimize the ligand affinity, since our results suggest that the energy cost of electron-pairing is higher than the gain by an enhanced ligand-binding affinity .

Chapter 5. Locating Hotspots: LFA-1/ICAM-1 and LFA-1/Inhibitor

The goal of the present part of the thesis was to establish a relationship between specific residues or hotspots in the interaction of the low, intermediate and high affinity structures of the αL I-domain with ICAM-1 and the experimentally measured affinity constants [60], i.e., their increasing affinity for the natural ligand ICAM-1 (Kd=1500μM, 3μM and 0.15μM, respectively). With this goal in mind, the low and high affinity complexes were created following different docking protocols and Molecular Dynamics Simulations were run with the three complexes under study. To locate the interaction hotspots a Free Energy Decomposition was applied to the generated geometries. Binding free energies were calculated and decomposed on a per-residue basis, systematically analyzing the magnitude and the nature of the contribution of all the amino acids in the interaction between the low, medium and high affinity conformations of the I-domain of integrin LFA-1 with its counterpart ICAM-1. The same analysis was on the complex formed by the high affinity αL I-domain and the LFA-1 antagonist HC0303.

LFA-1/ICAM-1 complex

The low affinity αL I-domain as it is in the Protein Data Bank (PDB ID=1ZOP) needs to undergo structural rearrangements in order to form a complex with ICAM-1. Therefore, additional structural changes are required prior to ICAM-1 binding. It can also be inferred from the results that ICAM-1 is not able to provoke these structural rearrangements in the isolated I-domain. This result may represent an additional support for the theory of inside-out signaling in integrins.

The high affinity αL I-domain as it is the Protein Data Bank (PDB ID=1MQ9) can form an stable complex with ICAM-1. Upon complex generation, it is observed that the position of the C-terminal α7 helix varies dramatically with respect to its position in the pseudo-liganded crystal structure. This rearrangement consists of the helix flipping away from the central core of the protein.

The Free Energy Decomposition analysis revealed that the three complexes are stabilized by interactions of different nature. While the low affinity complex is mainly vdW in nature, the high affinity complex is mainly electrostatic. The contribution of favorable vdW interactions to the total binding free energy decreases with respect to the weight of the total electrostatic interactions (interactions + desolvation), in the path from the L.A. to the M.A. and H.A. complex.

The increased affinity of the H.A. αL I-domain for ICAM-1 with respect to the M.A. and H.A. complexes results from a decrease in the electrostatic desolvation penalty to be paid.

Not only do the three-dimensional structures of the M.A. and the H.A. I-domains display high homology, but the estimated free energies of binding to ICAM-1 are quantitatively similar. These results suggest that both structures may not represent and intermediate and last position in the activation path but, on the contrary, they are highly equivalent structures.

Hotspots have been identified in each complex, and the most important interactions described. In addition, it was found that the number of residues making significant contacts does not determine the relative affinity of one protein for the other. In fact, it was observed that an increase in affinity is accompanied by a decrease in the number of favorable hotspots in general, and a decrease in the number of unfavorable hotspots and an increases in the efficiency of the interactions, with respect to the medium affinity and low affinity complexes.

LFA-1/Inhibitor complex

The MD and FED analysis performed in this work revealed that compound HC0303 and the H.A. I-domain form an stable complex, were important electrostatic contacts are created with the metal in the MIDAS motif. In addition, and confirming results in Chapter 3, there is an important stabilization of the complex due to the desolvation of the hydrophobic groups in the different position of the central pyrrolidine ring.

The MD simulation and the FED analysis also revealed that the binding of the LFA-1 antagonist HC0303 could be further optimized by modifying the spatial distribution of the negatively charged carboxylic and the positively charged amine groups present in the inhibitor. Two new families of inhibitors were designed and synthesized. At least one compound in each family were shown in *in vitro* and *in vivo* tests to have improved inhibitory capacities.

Chapter 6. NMR experiments to locate the binding pocket

Nuclear Magnetic Resonance and Surface Plasmon Resonance techniques were used to locate the binding site, estimate the binding constant and verify the inhibitory capacity of our compounds, both to the wild type and active mutant αL I-domains. The results and conclusions are shown below.

ICAM-1 dependent closed-to-open conformational change of the MIDAS motif

In this thesis work, experimental evidences have been found for the first time that indicate that the active mutant αL I-domain in solution does not have an *open* MIDAS motif. These results, support the idea that ICAM-1 may provoke the closed-to-open conformational change and generate the active-open structure observed in the pseudo-liganded (PDB ID=1MQ9) and ICAM-3 liganded (PDB ID=1TOP) αL I-domain structures published in the Protein Data Bank. The disulfide bridges experimentally introduced to lock the C-terminal α7 helix in the low, medium or high affinity conformation, may only prime the I-domain to bind extracellular ligands, but do not provoke the conformational change in the MIDAS motif.

Compound binding to the allosteric site

LFA-1 antagonists, VLA-4 antagonists and compounds non-related to integrin inhibition were observed to bind to the hydrophobic pocket underneath the C-terminal α7 helix of the wild-type αL I-domain, instead of to the expected MIDAS motif at the top of the domain. The same is true for compound binding to the high affinity αL I-domain. It is proposed for the first time that the IDAS site in the isolated αL I-domain is artifactually sticky, which may not happen in the context of the intact heterodimer. The latter is supported by the fact that previously reported *in vivo* and *in vitro* results with our compounds, revealed the existence of compound specificity by VLA-4 and LFA-1.

Compound binding to the IDAS site in the isolated wild type I-domain was found to be in the low micromolar range (7-10 μM), and in the high micromolar range (80 μM aprox.) for binding to the high affinity mutant αL I-domain.

Inhibition of ICAM-1 binding

Surface Plasmon Resonance experiments revealed that compound binding to the allosteric site of the high affinity αL I-domain does not inhibit binding to ICAM-1. On the contrary, *in vivo* and cell adhesion *in vitro* experiments revealed that the inhibition of the LFA-1/ICAM-1 interaction by our compounds was significant and promising. There is therefore inconsistency between both type of experimental results. The above summarized results, together with a deep analysis of the literature, support the hypothesis that the isolated αL I-domain may not be an appropriate biological unit to study the binding of small molecular weight compounds. On the contrary, a proper characterization of compound binding requires the use of the intact αLβ2 heterodimer.

Bibliography

[1] Alberts, B.; Johnson, A.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell.* Garland Science, New York, 2002.

[2] Zamir, E. *Cell Adhesion.* Oxford University Press, Israel, 2002.

[3] Silverstein, J.; Silverstein, R. L. *Cancer Invest.*, **1998**, *16(3)*, 176.

[4] Seymour, G. B.; Tucker, G.; Leach, L. *Biotechnol. Genet. Eng. Rev.*, **2004**, *21*, 123.

[5] Jones, Y. *Curr. Op. Cell Biol.*, **1996**, *8(5)*, 602.

[6] Lukas, Z.; Dvorak, K. *Acta Vet. Brno.*, **2004**, *73*, 93.

[7] Ranscht, B. *Curr. Op. Cell Biol.*, **1994**, *6*, 740.

[8] Nose, A.; Nagafuchi, A.; Takeichi, M. *Cell*, **1988**, *54*, 993.

[9] Takeichi, M. *Annu. Rev. Biochem.*, **1990**, *59*, 237.

[10] Gumbiner, B. M. *Cell*, **1996**, *84*, 345.

[11] Nose, A.; Tsuji, K.; Takeichi, M. *Cell*, **1990**, *61*, 147.

[12] Nagar, B.; Overduin, M.; Ikura, M.; Rini, J. M. *Nature*, **1996**, *380*, 360.

[13] Huang, Z.; Korngold, R. *Biopolymers*, **1997**, *43*, 367.

[14] Hynes, R. O. *Cell*, **1992**, *69*, 11.

[15] Wickham, T. J.; Mathias, P.; Cheresh, D. A.; Nemerow, G. R. *Cell*, **1993**, *73*, 309.

[16] Schlaepfer, D. D.; Hunter, T. *Trends Cell Biol.*, **1993**, *8*, 151.

[17] Clark, E. A.; Brugge, J. S. *Sciene*, **1995**, *268*, 233.

[18] Diamond, M. S.; Springer, T. A. *Curr. Biol.*, **1994**, *4*, 506.

[19] Hynes, R. O. *Cell*, **2002**, *110*, 673.

[20] Tedder, T. F.; Steeber, D. A.; Chen, A.; Engel, P. *Faseb J.*, **1995**, *9*, 866.

[21] Lasky, L. A. *Science*, **1992**, *258*, 964.

[22] Bevilacqua, M. P.; Nelson, R. M. *J. Clin. Invest.*, **1993**, *91*, 379.

[23] Kholova, I.; Ludvikova, M.; Ryska, A.; Cap, J. *Cs. Patol.*, **2002**, *38*, 11.

[24] Tanaka, T.; Umeki, K.; Yamamoto, I.; Aratake, Y.; Sakamoto, F.; Noguchi, S.; Ohtaki, S. *Int. J. Cancer*, **1995**, *64*, 326.

[25] Chang, H. J.; Jee, C. D.; Kim, W. H. *J. Am. Surg. Pathol.*, **2002**, *26*, 758.

[26] Zeev, A. B. *Curr. Opin. Cell Biol.*, **1997**, *9*, 99.

[27] Beavon, I. R. G. *J. Clin. Pathol.*, **1999**, *52*, 179.

[28] Bartek, J.; Lukas, J.; Bartkova, J. *J. Pathol.*, **1999**, *187*, 95.

[29] Mueller, S. C.; Ghersi, G.; Akiyama, S. K.; Sang, Q. X. A.; Howard, L.; Pineiro-Sanchez, M.; Nakahara, H.; Yeh, Y.; Chen, W. T. *J. Biol. Chem.*, **1999**, *35*, 24949.

[30] Saito, T.; Yamamoto, Y.; Shinno, S. *No To Hattatsu*, **2002**, *34(4)*, 326.

[31] Turner, G. D. H.; Van, C. L.; Nguyen, T. H. M.; Tran, T. H. C.; Nguyen, H. P.; Bethell, D.; Wyllie, S.; Louwrier, K.; Fox, S. B.; Gatter, K. C.; Day, N. P.; Tran, T. H.; White, N. J.; Berendt, A. R. *The Am. J. of Pathol.*, **1998**, *152*, 1477.

[32] Tamkun, J. W.; DeSimone, D. W.; Fonda, D.; Patel, R. S.; Buck, C.; Horwitz, A. F.; Hynes, R. O. *Cell*, **1986**, *46*, 271.

[33] Hynes, R. O. *Cell*, **1987**, *48*, 549.

[34] Humphries, M. J. *Biochem. Soc. Trans*, **2000**, *28*, 311.

[35] Springer, T. A. *Proc. Nat. Acad. Sci.*, **1997**, *94*, 65.

[36] Larson, R. S.; Corbi, A. L.; Berman, L.; Springer, T. *J. Cell Biol.*, **1989**, (73), 703.

[37] Collombati, A.; Bonaldo, P. *Blood*, **1991**, (77), 2305.

[38] Chothia, C.; Jones, E. Y. *Annu. Rev. Biochem.*, **1997**, *66*, 823.

[39] Baldwin, E. T.; Sarver, R. W.; Bryant, G. L. J.; Curry, K. A.; Fairbanks, M. B. *Structure*, **1998**, *6*, 923.

[40] Lee, J.-O.; Bankston, L. A.; Arnaout, M. A.; Liddington, R. C. *Structure*, **1995**, *3*, 1333.

[41] Lee, J.-O.; Rieu, P.; Arnaout, M. A.; Liddington, R. C. *Cell*, **1995**, *80*, 631.

[42] Kallen, J.; Welzenbach, K.; Ramage, P.; Geyl, D.; Kriwacki, R.; Legge, G.; Cottens, S.; Weitz-Schmidt, G.; Hommel, U. *J. Mol. Biol.*, **1999**, *292*, 1.

[43] Legge, G. B.; Kriwacki, R. W.; Chung, J.; Hommel, U.; Ramage, P. *J. Mol. Biol.*, **2000**, *295*, 1251.

[44] Qu, A.; Lehaly, D. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 10277.

[45] Qu, A.; Leahy, D. J. *Structure*, **1996**, *4*, 931.

[46] Emsley, J.; King, S. L.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. *Cell*, **2000**, *101*, 47.

[47] Emsley, J.; King, S. L.; Bergelson, J. M.; Liddington, R. C. *J. Biol. Chem.*, **1997**, *272*, 28512.

[48] Rich, R. L.; Deivanayagam, C. C.; Owens, R. T.; Carson, M.; Hook, A. *J. Biol. Chem.*, **1999**, *274*, 24906.

[49] Nolte, M.; Pepinsky, R. B.; Venyaminov, S. Y.; Koteliansky, V.; Gotwals, P. J.; Karpusas, M. *FEBS Lett.*, **1999**, *452*, 379.

[50] Huang, C.; Lu, C.; Springer, T. A. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*, 3156.

[51] Shimaoka, M.; Takagi, J.; Springer, T. A. *Annu. Rev. Biophys. Biomol. Struct.*, **2002**, *31*, 485.

[52] Gahmberg, C. G.; Tolvanen, M.; Kotovuori, P. *Eur. J. Biochem.*, **1997**, *245*, 215.

[53] Shattil, S. J.; Kashiwagi, H.; Pampori, N. *Blood*, **1998**, *91*, 2645.

[54] Lauffenburger, D. A.; Horwitz, A. F. *Cell*, **1996**, *84*, 359.

[55] Stewart, M.; Hogg, N. *J. Cell Biochem.*, **1996**, *61*, 554.

[56] Lub, M.; van Kooyk, Y.; Fidgor, C. G. *Immunol. Today*, **1995**, *16*, 479.

[57] Humphries, M. J. *Curr. Op. Cell Biol.*, **1996**, *8*, 632.

[58] Mould, A. *Biochem. J.*, **1998**, *333*, 821.

[59] Humphries, M. J.; P.Newham. *Trends Cell Biol.*, **1998**, *8*, 78.

[60] Shimaoka, M.; Xiao, T.; Liu, J. H.; Yang, Y.; Dong, Y.; Jun, C. D.; McCormack, A.; Zhang, R.; Joachimiak, A.; Takagi, J.; Wang, J.-H.; Springer, T. A. *Cell*, **2003**, *112*, 99.

[61] Labadia, M. E.; Jeanfavre, D. D.; Caviness, G. O.; Morelock, M. M. *J. Immunol.*, **1998**, *161*, 836.

[62] Lu, C.; Shimaoka, M.; Znag, Q.; Takagi, J.; Springer, T. A. *Proc. Nat. Ac. Sci.*, **2001**, *98*, 2393.

[63] Dransfield, I.; Cabanas, C.; Craig, A.; Hogg, N. *J. Cell Biol.*, **1992**, *116*, 219.

[64] Griggs, D. W.; Schmidt, C. M.; Carron, C. P. *J. Biol. Chem.*, **1998**, *273*, 22113.

[65] Chateau, M.; Chen, S.; Salas, A.; Springer, T. A. *Biochemistry*, **2001**, *40*, 13972.

[66] Novick, S. G.; Fodfrey, J. C.; Pollack, R. L.; Wilder, H. R. *Med. Hypotheses*, **1997**, *49*, 347.

[67] Sebastian, E. S.; Mercero, J. M.; Stote, R. H.; Dejaegere, A.; Cossio, F. P.; Lopez, X. *J. Am. Chem. Soc.*, **2006**, *128 (11)*, 3554.

[68] Sebastian, E. S.; Matxain, J.; Eriksson, L.; Stote, R. H.; Cossio, A. D. F.; Lopez, X. *J. Phys. Chem. B.*, **2007**, *17608410*.

[69] Amzel, L. M.; Poljak, R. J. *Ann. Rev. Biochem.*, **1979**, *48*, 961.

[70] Williams, A. F.; Barclay, A. N. *Ann. Rev. Immunol.*, **1988**, *6*, 381.

[71] Huang, Z.; Li, S.; Korngold, R. *Med. Chem. Res.*, **1997**, *7*, 137.

[72] White, J. M.; Littman, D. R. *Cell*, **1989**, *56*, 725.

[73] Taylor, P.; Bilsland, M.; Walkinshaw, M. *Acta Crystallogr.*, **2001**, *57*, 1579.

[74] der Merwe, P. V.; McNamee, P. N.; Davies, E. A.; Barclay, A. N.; Davis, S. J. *Curt. Biol.*, **1995**, *5*, 74.

[75] Jones, E. Y.; Davis, S. J.; Williams, A. F.; Harlos, K.; Stuart, D. I. *Nature*, **1992**, *360*, 232.

[76] Bodian, D. L.; Jones, E. Y.; Harlos, K.; Stuart, D. I.; Davis, S. J. *Structure*, **1994**, *2*, 755.

[77] Binnerts, M. E.; van Kooyk Y.; Simmons, D. L.; Fidgor, C. G. *Eur. J. Immunol.*, **1994**, *24*, 2155.

[78] Song, G.; Yuting, Y.; Liu, J.-H.; Casasnovas, J. M.; Shimaoka, M.; Springer, T. A.; Wang, J.-H. *Proc. Nat. Acad. Sci. USA*, **2005**, *102*, 3366.

[79] Yednock, T. A.; Cannon, C.; Fritz, L. C.; Sanchez-Madrid, F.; Steinman, L.; Karin, N. *Nature*, **1992**, *356*, 63.

[80] von Adrian, U.; Engelhardt, B. *N. Engl. J. Med.*, **2003**, *68*, 68.

[81] Miller, D. H.; Khan, O. A.; Sheremata, W. A.; Blumhardt, L. D.; GP, G. P. R.; Libonati, M. A.; Willmer-Hulme, A. J.; Dalton, C. M.; Miszkiel, K. A.; O'Connor, P. W. *N. Engl. J. Med.*, **2003**, *348*, 15.

[82] Lou, J.; Chofflon, M.; Juillard, C.; Donati, Y.; Mili, N.; Siegrist, C. A.; Grau., G. E. *Neuroreport*, **1997**, *8*, 629.

[83] Nishimura, M.; Obayashi, H.; Maruya, E.; Ohta, M.; Tegoshi, H.; andG. Hasegawa, M. F.; Shigeta, H.; Kitagawa, Y.; Nakano, K.; Saji, H.; Nakamura, N. *Hum. Immunol.*, **2000**, *61*, 507.

[84] Bloemen, P. G.; van den Tweel, M. C.; Henricks, P. A.; Engels, F.; Wagenaar, S. S.; Rutten, A. A.; Nijkamp, F. P. *Am. J. Respir. Cell Mol. Biol.*, **1993**, *9*, 586.

[85] Grnberg, K.; Sterk, P. J. *Clin. Exp. Allergy*, **1999**, *29*, 65.

[86] Lima, C.; Franco-Penteado, C.; Canalli, A.; Conran, N.; Lorand-Metze, I.; Costa, F.; Ferreira, H. *Leukemia Res.*, **2007**, *31*, 695.

[87] Gulubova, M.; Manolova, I.; Vlaykova, T.; Prodanova, M.; Jovchev, J. *Int. J. Col. Dis.*, **2007**, *22*, 581.

[88] Yusuf-Makagiansar, H.; Anderson, M. E.; Yakovleva, T. V.; Murray, J. S.; Siahaan, T. J. *Med. Res. Rev.*, **2002**, *22*, 146.

[89] Quian, F.; Vaux, D. L.; Weissman, I. K. *Cell*, **1994**, *77*, 335.

[90] Schadendorf, D.; Heidel, J.; Gawlik, C.; Suter, L.; Czarnetzki, B. M. *J. Natl. Cancer Inst.*, **1995**, *87*, 366.

[91] Langley, R. R.; Carlisle, R.; Ma, L.; Specian, R. D.; Gerritsen, M. E.; Granger, D. N. *Microcirculation*, **2991**, *8*, 335.

[92] Tomita, Y.; Saito, T.; Saito, K.; Oite, T.; Shimizu, S.; Sato, S. *Int. J. Cancer*, **1995**, *60*, 753.

[93] Matsumura, N.; Puzon-McLaughlin, W.; Irie, A.; Morikawa, Y.; Kakudo, K.; Takada, Y. *Am. J. Pathol*, **1996**, *148*, 55.

[94] Yasoshima, T.; Denno, R.; Kawaguchi, S.; Sato, N.; Okada, Y.; Ura, H.; Kikuchi, K.; Hirata, K. *Jpn. Cancer Res.*, **1996**, *87*, 153.

[95] Papadimitriou, M. N.; Menter, D. G.; Konstantinopoulos, K.; Nicolson, G. L.; McIntire, L. V. *Clin. Exp. Metastasis*, **1999**, *17*, 669.

[96] Yasoshima, T.; Denno, R.; Kawaguchi, S.; Sato, N.; Okada, Y.; Ura, H.; Kikuchi, K.; Hirata, K. *Jpn. J. Cancer*, **1996**, *87*, 153.

[97] Gulubova, M. V. *Histochem. J.*, **2002**, *34*, 67.

[98] Miele, M. E.; Bennett, C. F.; Miller, B. E.; Welch, D. R. *Exp. Cell Res.*, **1994**, *214*, 231.

[99] Wang, H. H.; McIntosh, A. R.; Hasinoff, B. B.; MacNeil, B.; Rector, E.; Nance, D. M.; Orr, F. W. *Eur. J. Cancer*, **2002**, *38*, 1261.

[100] Hakansson, A. J.; Gustafsson, B.; Krysander, L.; Hjelmqvist, B.; Rettrup, B.; Hakanson, L. *J. Interf. Cytok. Res*, **1999**, *19*, 171.

[101] Soede, R. D.; Zeelenberg, I. S.; Wijnands, Y. M.; Kamp, M.; Roos, E. J. *J. Immunol.*, **2001**, *166*, 4293.

[102] St-Pierre, Y.; Aoudjit, F.; Lalancette, M.; Potworowski, E. F. *Leuk. Lymphoma*, **1999**, *34*, 53.

[103] Harning, R.; Myers, C.; Merluzzi, V. J. *J. Clin. Exp. Metastasis*, **1993**, *11*, 337.

[104] Driessens, M. H.; van Hulten, P.; Zuurbier, A.; Riviere, G. L.; Roos, E. *J. Leukoc. Biol.*, **1996**, *60*, 758.

[105] Fujisaki, T.; Tanaka, Y.; Fujii, K.; Mine, S.; Saito, K.; Yamada, S.; Yamashita, U.; Irimura, T.; Eto, S. *Cancer Res*, **1999**, *59*, 4427.

[106] Zubia, A. *Formación de enlaces Carbono-Carbono mediante catálisis de sales metálicas y de dendrímeros. Aplicación a la síntesis de inhibidores de interacctiones proteína-proteína. UPV/EHU, Donostia 2003.*

[107] Vivanco, S. *Estudios sobre el Mecanismo de la Cicloadición [3+2] entre Iluros de Azometino y Nitroalquenos. Aplicación a la Síntesis de Desintegrinas no Naturales. UPV/EHU, Donostia 2002.*

[108] Ayerbe, M. *Nuevos estudios sobre reacciones secuenciales y concertadas de alquenos pi-deficientes; UPV/EHU, Donostia 2003.*

[109] Aldaba, E. *Nuevas Aplicaciones de las cicloadiciones [3+2] en la preparación de compuestos de interés en Biomedicina y en Ciencias de Materiales. UPV/EHU, Donostia 2007.*

[110] Hohenberg, P.; Kohn, W. *Phys. Rev.*, **1964**, *136*.

[111] S. J. Vosko, L. W.; Nusair, M. *Can. J. Phys.*, **1980**, *58*, 1200.

[112] Becke, A. D. *Phys. Rev*, **1988**, *38*, 3098.

[113] Lee, C.; Yan, W.; Parr, R. G. *Phys. Rev. B.*, **1988**, *37*, 785.

[114] Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.*, **1992**, *70*, 612.

[115] Schlegel, H. B. *J. Comp. Chem.*, **1982**, *3*, 214.

[116] McKee, M. L.; Page, M. *K.B. Lipkowitz, D. B. Boyd, Eds.*, **1993**, *Reviews in Computational Chemistry, Volume IV VCH Publishers*.

[117] Hehre, W. J.; Radom, L.; v. R. Schleyer, P.; Pople, J. A. *John Wiley and Sons, New York*, **1986**.

[118] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models. John Wiley and Sons, New York*, **2002**.

[119] Trzaskowski, B.; Les, A.; Adamowicz, L. *Int. J. Mol. Sci*, **2003**, *4*, 503.

[120] de Visser, S.; Ogliaro, F.; Gross, Z.; Shaik, S. *Chem. Eur. J*, **2001**, *7*, 4954.

[121] Wagner, L.; Mitas, L. *Chem. Phys. Lett.*, **2003**, *370*, 412.

[122] Hartree, D. R. *Proc. Cambridge Phil. Soc.*, **1928**, *24*, 328.

[123] Fock, V. A. *Z. Phys.*, **1930**, *15*, 126.

[124] Drummond, N. D.; Towler, M. D.; Needs, R. J. *Phys. Rev. B*, **2004**, *70*, 235119.

[125] Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. *Rev. Mod. Phys.*, **2001**, *73*, 33.

[126] Mercero, J.; Matxain, J.; Lopez, X.; York, D.; Largo, A.; Eriksson, L. A.; Ugalde, J. M. *Int. J. Mass Spectrometry*, **2005**, *240*, 37.

[127] Anderson, J. B. *J. Chem. Phys.*, **1975**, *63*, 1499.

[128] Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.*, **1980**, *45*, 566.

[129] Anderson, J. B. *J. Chem. Phys.*, **1976**, *65*, 4121.

[130] P. J. Reynolds, D. M. C.; Alder, B. J.; Jr., W. A. L. *J. Chem. Phys.*, **1982**, *77*, 5593.

[131] Moskowitz, J. W.; Schmidt, K. E.; Lee, M. A.; Kalos, M. H. *J. Chem. Phys.*, **1982**, *77*, 349.

[132] Hammond, B. L.; Lester, W. A.; Reynolds, P. J. World Scientific, Singapore, 1994.

[133] Mitas, L.; Shirley, E. L.; Ceperley, D. M. *J. Chem. Phys.*, **1991**, *95*, 3467.

[134] Brooks, B. R.; Bruccoleri, B. D.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.*, **1983**, (4), 187.

[135] Leach, A. R. *Molecular Modelling-Principles and Applications, 2nd Edition.* Prentice Hall, 2001.

[136] Frenkel, D.; Smit, B. *Understanding Molecular Simulation. Acad. Press*, **1996**.

[137] Steinbach, P. J.; Brooks, B. R. *J. Comp. Chem.*, **1993**, *15*, 667.

[138] Ewald, P. *Ann. Phys.*, **1921**, *253*.

[139] Kittel, C. *Introduction to Solid State Physics. John Wiley and Sons, New York*, **1971**.

[140] Tosi, M. *Solid State Physics. Academic Press, New York*, **1964**.

[141] York, D. M.; Darden, T. A. *J. Chem. Phys.*, **1993**, *99*, 8345.

[142] Stote, R. H.; States, D. J. *J. Chem. Phys.*, **1991**, *88*, 2419.

[143] Greengard, L.; Rokhlin, V. *J. Comp. Phys.*, **1987**, *73*, 325.

[144] Shimada, J. H.; Kaneko, H. *J. Comp. Chem.*, **1994**, *15*, 28.

[145] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comp. Phys.*, **2002**, *23*, 1515.

[146] MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack, Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B*, **1998**, *102*(18), 3586.

[147] Case, D. A.; T.E. Cheatham, I.; Darden, T.; Gohlke, H.; Luo, R.; Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods., R. *J. Computat. Chem.*, **2005**, *26*, 1668.

[148] Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers*, **1984**, *23*, 1.

[149] Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comp. Chem.*, **1997**, *18*, 1955.

[150] Halgren, T. A. *J. Am. Chem. Soc.*, **1992**, *114*, 7827.

[151] Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.*, **1995**, *117*, 1955.

[152] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Kollman, J. W. C. P. A. *J. Am. Chem. Soc.*, **1995**, *117*, 5179.

[153] Cauchy, A. *Compte-Rendu de l'Acadmie des Sciences, Paris*, **1847**, *27*, 536.

[154] Fletcher, R.; Reeves, C. M. *Comp. Journal*, **1964**, *7*, 149.

[155] Luenberger, D. *Addison-Wesley: Menlo Park, CA.*, **1984**.

[156] Metropolis, N.; Rosembluth, A.; Rosembluth, M.; Teller, A. *J. Chem. Phys.*, **1953**, *21*, 1087.

[157] Allen, M. P.; Tildesley, D. J. *Oxford University Press*, **1989**.

[158] Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.*, **1992**, *4*, 1082.

[159] Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.*, **1991**, *62*, 187.

[160] Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comp. Chem.*, **1988**, *9*, 327.

[161] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.*, **1990**, *112*, 6127.

[162] Atkins, P. W. *Freeman*, **1990**, *New York*.

[163] Warwicker, J.; Watson, H. C. *J. Mol. Biol.*, **1982**, *157*, 671.

[164] Schaefer, M.; Sommer, M.; Karplus, M. *J. Phys. Chem. B*, **1997**, *101*, 1663.

[165] Schaefer, M.; v. Vlijmen, H. W.; Karplus, M. *Adv Protein Chem*, **1998**, *51*, 1.

[166] Warshel, A.; Papzyan, A. *Curr. Op. Struc. Biol.*, **1998**, *8*, 211.

[167] van Vlijmen, H. W.; Schaefer, M.; Karplus, M. *Proteins Struc. Func. Genet.*, **1998**, *33*, 145.

[168] Nielsen, J. E.; McCammon, J. A. *Prot. Sci.*, **2003**, *12*, 1894.

[169] Kyte, J. Garland Publishing Inc., New York, 1995.

[170] Honig, B.; Nicholls, A. *Science*, **1995**, *268*, 1144.

[171] Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T. *Acc. Chem. Res.*, **2000**, *33*, 889.

[172] Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.*, **2003**, *330*, 891.

[173] Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins*, **1998**, *32*, 339.

[174] Gohlke, H.; Case, D. A. *J. Comp. Chem.*, **2004**, *25*, 238.

[175] Zoete, V.; Meuwly, M.; Karplus, M. *Proteins*, **2005**, *61*, 79.

[176] Hendsch, S.; Tidor, B. *Protein Science*, **1999**, *8*, 1381.

[177] Hutter, E.; Fendler, J. *SPARKY 3, University of California, San Francisco*, **2004**, *16*, 1685.

[178] Huth, J. R.; Olejniczak, E. T.; Mendoza, R.; Liang, H.; Harris, E. A. S.; M. L. Lupher, J.; Wilson, A. M.; Fesik, S. W.; Stauton, D. E. *Proc. Nat. Acad. Sci.*, **2000**, *97*, 5231.

[179] Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR*, **1995**, *6*, 277.

[180] Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR*, **1994**, *4*, 603.

[181] Goddard, T. D.; Kneller, D. G. *SPARKY 3, University of California, San Francisco.*

[182] Aplin, A. E.; Howe, A. K.; Juliano, R. L. *Curr. Op. Cell Biol.*, **1999**, (11), 737.

[183] Plow, E. F.; Haas, T. A.; Zhang, L.; Loftus, J.; Smith, J. W. *J. Biol. Chem.*, **2000**, (275), 21785.

[184] Gottschalk, K.-E.; Kessler, H. *Angew. Chem. Int. Ed.*, **2002**, (20), 3767.

[185] Larson, R. S.; Springer, T. A. *Immunol. Rev.*, **1990**, (114), 181.

[186] Huang, C.; Springer, T. A. *J. Biol. Chem.*, **1995**, *270*, 19008.

[187] Stanley, P.; Hogg, N. *J. Biol. Chem.*, **1998**, *273*, 3358.

[188] Dudev, T.; Lim, C. *Chem. Rev.*, **2003**, *103*, 773.

[189] Lu, C.; Shimaoka, M.; Oxvig, M.; Takagi, J.; Springer, T. A. *Proc. Nat. Ac. Sci.*, **2001**, *98*, 2387.

[190] Ajroud, K.; Sugimori, T.; Goldmann, W. H.; Fathallah, D. M.; Xiong, J.-P.; naout, M. A. A. *J. Biol. Chem.*, **2004**, *279*, 25483.

[191] Zubia, A.; Mendoza, L.; Vivanco, S.; Albada, E.; Carrascal, T.; Lecea, B.; Arrieta, A.; Zimmerman, T.; Vidal-Vanaclocha, F.; Cossio, F. P. *Angew. Chem. Int. Ed.*, **2003**, *44*, 2903.

[192] Gadek, T. R.; Burdick, D. J.; McDowell, R. S.; Stanley, M. S.; Jr., J. C. M.; Paris, K. J.; Oare, D. A.; Reynolds, M. E.; Ladner, C.; Zioncheck, K. A.; Lee, W. P.; Gribling, P.; Dennis, M. S.; Skelton, N. J.; Tumas, D. B.; Clark, K. R.; Keating, S. M.; Beresini, M. H.; Tilley, J. W.; Presta, L. G.; Bodary, S. C. *Science*, **2002**, *295*, 1086.

[193] Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nuc. Ac. Res.*, **2000**, *28*, 235.

[194] Brunger, A. T.; Karplus, M. *Proteins*, **1988**, *4*, 148.

[195] Mercero, J. M.; Matxain, J. M.; Lopez, X.; DM, D. M. Y.; A, A. L.; Eriksson, L. A.; Ugalde, J. M. *Int. J. Mass Spec.*, **2005**, *240*, 37.

[196] Becke, A. D. *J. Chem. Phys.*, **1993**, *98*, 5648.

[197] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.*, **1988**, *37*, 785.

[198] Labanowsky, J.; Andelzelm, J. *Density Functional Methods in Chemistry.* Springer-Verlag, New York, 1991.

[199] M. Alcamí and O. Mo and M. Yañez. *Mass. Spectr. Rev.*, **2001**, *20*, 195.

[200] Garmer, D. R.; Gresh, N. *J. Am. Chem. Soc.*, **1994**, *116*, 3556.

[201] Gresh, N.; Stevens, W. J.; Krauss, M. *J. Comp. Chem.*, **1995**, *16*, 843.

[202] Gresh, N.; Garmer, D. R. *J. Comp. Chem.*, **1996**, *17*, 1481.

[203] McQuarrie, D. A. *Statistical Mechanics*. Harper and Row, New York, 1976.

[204] Mercero, J. M.; Fowler, J. E.; Ugalde, J. M. *J. Phys. Chem. A*, **1998**, *102*(35), 7006.

[205] Mercero, J. M.; Fowler, J. E.; Ugalde, J. M. *J. Phys. Chem. A*, **2000**, *104*, 7053.

[206] Mercero, J. M.; Lopez, X.; Matxain, J. M.; Fowler, J. E.; Ugalde, J. M. *Int. J. Quantum Chem.*, **2002**, *90*, 859.

[207] Mercero, J. M.; Mujika, J. I.; Matxain, J. M.; Lopez, X.; Ugalde, J. M. *Chem. Phys.*, **2003**, *295*, 175.

[208] Mercero, J. M.; Matxain, J. M.; Rezabal, E.; Lopez, X.; Ugalde, J. M. *Int. J. Quantum Chem.*, **2004**, *98*, 409.

[209] Simonson, T. *J. Am. Chem. Soc.*, **1998**, *20*, 4875.

[210] Simonson, T.; Brooks, C. L. *J. Am. Chem. Soc.*, **1995**, *118*, 8452.

[211] Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.*, **1992**, *4*, 1082.

[212] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, a.11. Gaussian, Inc., Pittsburgh PA, 2001.

[213] Katz, A. K.; Glusker, J. P.; Beebe, S. A.; Bock, C. W. *J. Am. Chem. Soc.*, **1996**, *118*.

[214] Glusker, J. P.; Katz, A.; Bock, C. W. *The Rigaku Journal*, **1999**, *16*, 8.

[215] Marcus, Y. *Chem. Rev.*, **1988**, *88*, 1475.

[216] Alberts, I. L.; Nadassy, K.; Wodak, S. J. *Prot. Sci.*, **1998**, *7*, 1700.

[217] It is usual to assume that the Mn^{2+} complexes in proteins are high-spin molecules an such an assumption has been widely used in almost everywork concerning this type of complexes[235, 237]. Anyway, in this study, high (IV) and low (II) -spin Mn^{2+} complexes have been studied. We have found that as expected the Mn^{2+} structures with water ligands are high-spin complexes, the electronic energy difference between high-spin and low-spin states being 7.0 kcal/mol. Similarly, for the acetate-bound or acetate-unbound Mn^{2+} open MIDAS motifs, the most stable molecules are high-spin complexes, with an electronic energy difference of 13.1 and 8.0 kcal/mol, respectively. As a consequence, we decided to use the high-spin complexes for every reaction discussed in this work in which Mn^{2+} cation was involved.

[218] Because of the geometric and electronic features of the MIDAS, we have restricted our analysis to octahedral geometries for Ca^{+2}, although other configurations for this cation have been described[213]. Further work is in progress in order to study the possible effects in MIDAS of integrins by different coordination patterns.

[219] Ciancaglini, P.; Pizauro, J. M.; Curti, C.; Tedesco, A. C.; Leone, F. A. *Int. J. Biochem.*, **1990**, (22), 747.

[220] Lukat, G. S.; M., A.; Stock, J. B. *Biochemistry*, **1990**, (29), 5436.

[221] Sun, L.; Martin, D. C.; Kantrowitz, E. R. *Biochemistry*, **1999**, (38), 2842.

[222] Lee, L. V.; Russel; Poyner, R. R.; Vu, M. V.; Cleland, W. W. *Biochemistry*, **2000**, (39), 4821.

[223] Peschke, M.; Blades, A. T.; Kebarle, P. *J. Am. Chem. Soc.*, **2000**, *122*, 10440.

[224] Dudev, T.; Lim, C. *J. Phys. Chem. B.*, **2001**, *105*, 4446.

[225] Dudev, T.; Lim, C. *J. Am. Chem. Soc.*, **2000**, *122*, 11146.

[226] Dudev, T.; Lim, C. *J. Phys. Chem. B*, **2004**, *108*, 4546.

[227] Goldfarb, D.; D. Ariel, j. . A. v. . . p. . . y. . . a. . . title= Spin distribution and the location of protons in paramagnetic proteins.

[228] Messerschmidt, A.; Huber, A. R.; Poulos, T.; Wieghart, K. *New York: Wiley*, **2001**, p 1472.

[229] Borgstahl, G. E.; Parge, H. E.; Hickey, M. J.; Jr., W. F. B.; Hallewell, R. A.; Tainer, J. A. *Cell*, **1992**, *71*, 107.

[230] Sundaramoorthy, M.; Kishi, K.; Gold, M.; Poulos, T. *J. Biol. Chem.*, **1994**, *269*, 32759.

[231] Wilce, M. C. J.; Bond, C. S.; Dixonr, N. E.; Freeman, H. C.; Guss, J. M.; Lilley, P. E.; Wilce, J. A. *Biochemistry*, **1998**, *95*, 3472.

[232] Yousefi, S.; Higgins, E.; Daoling, Z.; Pollex-Kruger, A.; Hindsgaul, O.; Dennis, J. W. *J. Biol. Chem.*, **1991**, *266*, 1772.

[233] Beyer, T. A.; Hill, R. L. *J. Biol. Chem.*, **1980**, *255*, 5373.

[234] Qu, A.; Lehaly, D. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 10277.

[235] Waizumi, K.; Ohtaki, H.; Masuda, H.; Fukushima, N.; Watanabe, Y. *Chemm. Lett.*, **1992**, *8*, 1489.

[236] Trachtman, M.; Markham, G. D.; Glusker, J. P.; George, P.; Bock, C. W. *Inorg. Chem*, **1998**, *37*, 4421.

[237] Siegbahn, P. E. M. *Theor. Chem. Acc.*, **2001**, *105*, 197.

[238] Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. *Rev. Mod. Phys.*, **2001**, *73*, 33.

[239] Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.*, **1992**, *70*, 612.

[240] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.*, **1980**, *72*, 650.

[241] Mitas, L. *Phys. Rev. A*, **1994**, *49*, 4411.

[242] Matxain, J. M.; Mercero, J. M.; Irigoras, A.; Ugalde, J. M. *Mol. Phys.*, **2004**, *102*, 2635.

[243] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[244] Needs, R. J.; Towler, M. D.; Drummond, N. D.; Kent, P. R. C. Casino v1.7 user manual. University of Cambridge, Cambridge, 2004.

[245] Pasternak, A.; Goldschmidt, Z. B. *Phys. Rev. A*, **1974**, *9*, 1022.

[246] Garcia-Riquelme, O. *Opt. Pur. Appl.*, **1977**, *10*, 275.

[247] Dunbrack; Cohen. *Protein Science*, **1997**, *6*, 1661.

[248] Dunbrack; Karplus. *Nat. Struct. Biol.*, **1994**, *26*, 334.

[249] Dunbrack; Karplu. *J. Mol. Biol.*, **1993**, *230*, 543.

[250] Lafont, V.; Schaefer, M.; Stote, R.; Altschuh, D.; Dejaegere, A. *PROTEINS: Structure, Function, and Bioinformatics*, **2007**, *To appear.*

[251] Ryckaert, J. P.; Berendsen, G. C. H. J. C. *J. Comput. Phys.*, **1977**, *23*, 327.

[252] Lu, C.; Shimaoka, M.; Ferzly, M.; Oxvig, C.; Takagi, J.; Springer, T. A. *Proc. Nat. Ac. Sci.*, **1998**, *98*, 2387.

[253] Gaillard, T.; Martin, E.; Sebastian, E. S.; Cossio, F. P.; Lopez, X.; Dejaegere, A.; Stote, R. H. *J. Mol. Biology*, **2007**, *In press.*

[254] Hopkins, A. L.; Groom, C. R.; Alex., A. *Drug Discov. Today*, **2004**, *9*, 430.

[255] Andrews, P. R.; Craik, D. J.; Martin, J. L. *J. Med. Chem.*, **1984**, *27*, 1648.

[256] Lipinski, C. A. *J. Pharmacol Toxicol Methods*, **2000**, *44*, 235.

[257] Nicolls, M. R.; Coulombe, M.; Beilke, J.; Gelhaus, H. C.; Gill, R. G. *J. Immunol.*, **2002**, *169*, 4831.

[258] Poston, R. S.; Robbins, R. C.; Chan, B.; Simms, P.; Presta, L.; Jardieu, P.; Morris, R. E. *Transplantation*, **2000**, *69*, 2005.

[259] Kriwacki, R. W.; Legge, G. B.; Hommel, U.; Ramage, P.; Chung, J.; Tennant, L. L.; Wright, P. E.; Dysona, H. J. *J. Biol. NMR*, **2000**, *16*, 271.

[260] Shimaoka, M.; Springer, T. A. *Nature Rev.*, **2003**, *2*, 703.

[261] Last-Barney, K.; Davidson, W.; Cardozo, M.; Frye, L. L.; Grygon, C.; Hopkins, J. L.; Jeanfavre, D. D.; Pav, S.; Qian, C.; Stevenson, J. M. *J. Am. Chem. Soc.*, **2001**, *123*, 5643.

[262] Liu, G.; Huth, J. R.; Olejniczak, E. T.; Mendoza, R.; DeVries, P.; Leitza, S.; Reilly, E. B.; Okasinski, G. F.; Fesik, S. W.; von Geldern, T. W. *J. Med. Chem.*, **2001**, *44*, 1202.

[263] Felding-Habermann, B.; Cheresh, D. A. *Curr. Opin. Cell Biol.*, **1993**, *5*, 864.

[264] Crowe, D. T.; Chiu, H.; Fong, S.; Weissman, I. L. *J. Biol. Chem.*, **1994**, *269*, 14411.

[265] J. W. Smith, R. S. P.; Mathis, D. *J. Biol. Chem.*, **1994**, *269*, 960.

[266] Kelly, T. A.; Jeanfavre, D. D.; McNeil, D. W.; Woska, J. R.; Reilly, J. P. L.; Mainolfi, E. A.; Kishimoto, K. M.; Nabozny, G. H.; Zinter, R.; Bormann, B.-J.; Rothlein, R. *The J. Immunol.*, **1999**, *163*, 5153.

[267] Woska, J. R.; t Shih, D.; Taqueti, V. R.; Hogg, N.; Kelly, T. A.; Kishimoto, T. K. *J. Leuk. Biol.*, **2001**, *70*, 329.

[268] Weitz-Schmidt, G.; Welzenbach, K.; Brinkmann, V.; Kamata, T.; Kallen, J.; Bruns, C.; Cottens, S.; Takada, Y.; Hommel, U. *Nat. Med.*, **2001**, *7*, 687.

[269] Liu, G.; Link, J. T.; Pei, Z.; Reilly, E. B.; Leitza, S.; Nguyen, B.; Marsh, K. C.; Okasinski, G. F.; von Geldern, T. W.; Ormes, M.; Fowler, K.; Gallatin, M. *J. Med. Chem.*, **2000**, *43*, 4025.

[270] Fotouhi, N.; Gillespie, P.; Guthrie, R.; Pietranico-Cole, S.; Hoffmann, Y. W. *Patent*, **1999**.

[271] Fotouhi, N.; Gillespie, P.; Guthrie, R.; Pietranico-Cole, S.; Hoffmann, Y. W. *Patent*, **2001**.

[272] Shimaoka, M.; Salas, A.; Yang, W.; Weitz-Schmidt, G.; Springer, T. A. *Immunity*, **2002**, *19*, 391.

[273] Welzenbach, K.; Hommel, U.; Weitz-Schmidt, G. *J. Biol. Chem.*, **2002**, *2002*, 10590.

[274] Keating, S. M.; Clark, K. R.; Stefanich, L. D.; Arellano, F.; Edwards, C. P.; Bodary, S. C.; Spencer, S. A.; Gadek, T. R.; Jr., J. C. M.; Beresin, M. H. *Prot. Sci*, **2006**, *15*, 290.