Universidad de Alcalá

Programa de Doctorado en Señalización Celular

Computational approaches to study mechanisms of regulation and inhibition of enzymes involved in phospho-transfer reactions

Tesis Doctoral presentada por

Marta Camacho Artacho

Directores:
Dr. Daniel Lietha
Dr. Federico Gago

Alcalá de Henares, 2015

Universidad de Alcalá

DPTO. CIENCIAS BIOMÉDICAS
Facultad de Medicina
Universidad de Alcalá
Ctra. Barcelona, Km. 33,600
28871 Alcalá de Henares (Madrid)
Telf.: 91 885 45 49 - Fax: 91 885 45 91
e-mail: dpto.biomedicas@uah.es

D. DANIEL LIETHA, INVESTIGADOR PRINCIPAL EN EL LABORATORIO DE ADHESIÓN Y SEÑALIZACIÓN CELULAR EN EL DEPARTAMENTO DE BIOLOGÍA ESTRUCTURAL Y BIOCOMPUTACIÓN DEL CENTRO NACIONAL DE INVESTIGACIONES ONCOLÓGICAS, Y D. FEDERICO GAGO BADENAS, CATEDRÁDICO DE FARMACOLOGÍA DEL DEPARTAMENTO DE CIENCIAS BIOMÉDICAS DE LA UNIVERSIDAD DE ALCALÁ,

CERTIFICAN: que el trabajo titulado: "Computational approaches to study mechanisms of regulation and inhibition of enzymes involved in phospho-transfer reactions", ha sido realizado por Dª. Marta Camacho Artacho, bajo nuestra dirección y que cumple todos los requisitos para su defensa pública como Tesis Doctoral.

Alcalá de Henares, 10 de Noviembre dos mil quince.

Fdo: Daniel Lietha

Fdo: Federico Gago Badenas

UNIVERSIDAD DE ALCALÁ, PATRIMONIO DE LA HUMANIDAD

Universidad de Alcalá

DPTO. CIENCIAS BIOMÉDICAS
Facultad de Medicina
Universidad de Alcalá
Ctra. Barcelona, Km. 33,600
28871 Alcalá de Henares (Madrid)
Telf.: 91 885 45 49 - Fax: 91 885 45 91
e-mail: dpto.biomedicas@uah.es

D. FRANCISCO ZARAGOZÁ GARCÍA, CATEDRÁTICO DE FARMACOLOGÍA Y DIRECTOR DEL DEPARTAMENTO DE CIENCIAS BIOMÉDICAS DE LA UNIVERSIDAD DE ALCALÁ,

CERTIFICA: que el trabajo titulado: "Computational approaches to study mechanisms of regulation and inhibition of enzymes involved in phospho-transfer reactions", ha sido realizado por Dª. Marta Camacho Artacho, bajo la dirección de los Dres. Daniel Lietha, Investigador Principal del Grupo de Adhesión y Señalización Celular en el Departamento de Biología Estructural y Biocomputación del Centro Nacional de Investigaciones Oncológicas y Federico Gago Badenas, Catedrático de Farmacología de la Universidad de Alcalá, y cumple todos los requisitos para su defensa pública como Tesis Doctoral.

Alcalá de Henares, 9 de noviembre de dos mil quince.

UNIVERSIDAD DE ALCALÁ · DEPARTAMENTO DE CIENCIAS BIOMÉDICAS

UNIVERSIDAD DE ALCALÁ, PATRIMONIO DE LA HUMANIDAD

A Mamá y Papá,
Ter y Ro,
A Mi Pa.

Well, nobody's perfect.
Some like it hot (1959)

Carpe diem. Seize the day, boys. Make your lives extraordinary.
Dead Poets Society (1989)

Porque este viaje empezó hace algún tiempo, más de esos cuatro años que constan oficialmente, aquí va mi pequeño y particular homenaje a todas esas personas que habéis compartido conmigo esta apasionante aventura.

Todo empezó hace bastantes años en Nicaragua. Entonces sólo era una estudiante de farmacia que creía que se podía cambiar el mundo (y aún lo sigo creyendo, hay cosas que no cambian). Me fui de voluntaria creyendo que podría ayudar y cual fue mi sorpresa cuando descubrí que allí podían enseñarme mucho más que yo a ellos. Gracias Félix, por enseñarme que desde un laboratorio con pocos recursos, sin luz, sin agua, sin internet, pero con muchas ganas y dedicación, se pueden cambiar las cosas. Por abrirme los ojos a la investigación y a los demás. Gracias mi Xu, por estar desde el primer momento conmigo en esta aventura. Por creer que aunque no fuéramos a durar, aquí seguimos y por muchos años más. Por animarme y creer y en mi, hasta cuando me costaba verlo a miles de km y me llamabas y venías a verme allá donde estuviera para recordármelo.

Poco después tuve la suerte, porque en esta vida también hay que tener suerte para que las cosas ocurran, de caer en el laboratorio de farmacología, en mi último año de carrera. Gracias Federico, por abrirme las puertas y por ayudarme siempre. Por enseñarme el poder de las máquinas, de la importancia de los detalles, del rigor científico, por dejarme entrar en tu laboratorio y darme la oportunidad de ir a Nueva York. Porque aquí conocí a mi hacker, Juancho, eres un crack y siempre estaré aquí. A Claire, a Antonio, a Pedro y Álvaro. Por estar siempre disponibles, para un café, un email, unas palabras, por ayudarme siempre, gracias. A todos los "chicos" del CBM, Antonio, Alberto, David, Hugo, Rubén, Javi, Alfonso, y a Almu. Por esas cenas, risas, y encuentros.

A mi An. Porque juntas empezamos y juntas continuamos. Por ser mi hermana, asesora, compañera y amiga. Por enseñarme todo. Por compartir mesa, casa, ciudades, risas y llantos. Por ayudarme a crecer. Por tener siempre las palabras que necesitaba. Por ser mi alma gemela. Por guiarme cuando me perdía y animarme cuando encontré mi camino. Gracias se queda corto pero es la única palabra que se acerca. Gracias una y mil veces.

Nueva York, que ciudad, que vida, qué oportunidad. Gracias Marta por darme LA oportunidad de mi vida. Por ayudarme a superarme, por esos nespressos en tu despacho para que no estuviera nerviosa. Por tu paciencia y por enseñarme que hay que pensar antes de escribir. Porque, como más de una persona me ha dicho, tuve la oportunidad de hacer el postdoc antes del doctorado. Gracias. Hao, Zeng, Neville y Juan Carlos. Gracias a todos y sobre todo, gracias Jen y Davide, por acogerme y cuidarme. Por nuestros retreats, nuestros mexicanos, por formar parte de mi vida, siempre os tendré presente, lo sabéis.

Gracias a Alfonso y Francesco, por darme la oportunidad de volver a casa, Madrid. Gracias a mis italianos, a todos; a Ludo por ayudarme a entender; a las niñas, Ila y Silvi.

Por estar ahí siempre, por "partirnos de las risas", por todas esas post cards que me mandabas Ila, por las notitas, las comidas, los cafés, las confidencias, por todo. Mi Debo, mi guapi, no hay palabras. Por todos esos rodillas, cenas, teatros, cafés, Italia, Priego, Córdoba, mensajes, porque eres alucinante. Gracias por cada palabra de apoyo y por cada abrazo, gracias con todas mis fuerzas. A Simo, por todas esas sigarretas, esos descansos, por entenderme y aguantarme. Nicole, porque si hay algo que todos sabemos es que eras LA madre antes incluso de ser madre. Nos ayudabas a todos, y lo sigues haciendo. Por corregirme, aconsejarme, pedirme opinión, acompañarme y por dejarme conocer tu otra casa y familia, Göttingen. Gracias a todos los chicos del Max Planck, en especial Béla y Tim, por estar siempre pendientes y ayudarme en todo. Fuisteis mi familia allí. Gracias a mis compis de máster, Chechu, gracias por todo, por enseñarme el principio de indeterminación (entre mil cosas), siempre podré estar aquí y allí.

Gracias a mi grupo adoptivo, si, vosotros. Los bioinformáticos. Por adoptarme siempre: a la hora de la comida, los descansos, los seminarios, para las risas y … sólo hay risas, nunca hay malos rollos, sois geniales. Juan, gracias amigo. Mi corderito Mari, eres maravillosa, no cambies nunca. Jon mi cuidador, el traficante de chocolates que sabe como alegrarme el día y nunca se olvida de avisarme para ir a comer, aunque siempre esté ocupada. No dejes de hacerlo ☺. Fede, Miguel, Michael, Dani, David, Victor, Tirso, Jose, Enrique, Felipe, Vera, Simone, Martin, Osvaldo, Andrés, Jose Mª, Elena, Gonzalo, Miriam, David P., gracias a todos por vuestros saludos, felicitaciones y momentos compartidos. Por ayudarme siempre que os lo pido, y sobretodo, cuando no os lo pido. A mi Reich, por ser sencillamente genial, se te echa de menos todo el rato. A Belén y a Gema, porque siempre me ayudáis y con una sonrisa, gracias. A mi Ángel de la guarda. Porque aunque "eres mi ángel" te lo habrán dicho hasta la saciedad, rubia no hay más que una. Está tesis existe gracias a ti. Por todo tu apoyo, por ser el mejor en tu trabajo pero más por ser quien eres, mi amigo. Excepcional. Sin ti, el CNIO se hunde. A Mir por venir siempre a por mi, taconcitos, gracias.

Gracias, Daniel. Por darme la oportunidad de continuar y terminar mi doctorado trabajando contigo y aprendiendo de ti, es un auténtico honor y privilegio. Gracias por apoyarme siempre, darme proyectos que me encantan, preguntarme siempre lo que no sé, ser un jefe insuperable y mejor persona. Y sobre todo, gracias por animarme cuando (muy de vez en cuando) lo hago bien. Jo, Pilar, Chevi, gracias compis, es un placer trabajar con vosotros. Mi Marti pequeña, eres la brisa espontanea. Gracias por contagiar tu ilusión y tus ganas, por compartirlas conmigo y emocionarte conmigo, me encanta. Paquillo, paisano, gracias por cebarme como a un pavo. Gracias a todos los estructurales, Santiago, Maria, Ara, Rafa, Jaska, Igor, Ramón y Clara, por todas vuestras palabras amables y hacerme sentir una más. A la serbia más guapa del mundo por alegrarme los días y compartir conmigo tantos momentos. A mi sevillana, Alba. Gracias amiga por todo y más. Aquí estaré siempre para ti, cuando y donde tu quieras.

Gracias a mis niñas, Clara, María, y Rocío. Porque esta tesis es un trocito vuestra. Por aguantarme, obligarme a desconectar, hacerme ver más allá, quererme y cuidarme.

Gracias porque voy a trabajar feliz, amo lo que hago y el ambiente, gracias a todas las personas que acabo de mencionar y a las que seguro se me habrán olvidado. Miro hacia atrás y me considero realmente afortunada y puedo hacer un balance más que positivo de todos estos años.

Gracias al CNIO, porque pese a los momentos difíciles que ha vivido estos años, me ha dado la oportunidad de trabajar, aprender, y desarrollarme científica y profesionalmente en las mejores condiciones. Por hacerme sentir una privilegiada. Gracias a la Obra Social La Caixa, porque cuando el resto de precarios recursos de este país no me ofrecían futuro alguno, ellos me hicieron participe de una de las becas más prestigiosas. Gracias, por haberme dado la tranquilidad estos cuatro años para poder dedicarme única y exclusivamente a lo que realmente quería, la investigación.

Por último (siempre lo mejor para el final), quiero dar las gracias a quien va dedicada esta tesis. A los que han hecho posible que hoy esté donde esté y sea quien soy. Cada paso que he dado, cada palabra, cada escalón superado, cada momento, es gracias a ellos. Mamá, Papá, gracias por todo vuestro esfuerzo, sacrificio y dedicación. Por ayudarme cada día con los deberes. Por apagarme la tele – Marta, a estudiar. Por llevarme al británico aunque llorase, a Irlanda. Por ser mitad farmacéutica mitad físico-informática. Gracias Papá por enseñarme toda la farmacología que sé (y la botánica) y enseñarme cómo debe ser el mejor profesional sanitario. Gracias Mamá por sentarte conmigo y explicarme siempre, incluso conseguiste que aprendiese a integrar, ¡hasta diferenciales! Gracias por creer en mi siempre, por ser mis fans números uno. Por no dejar que me hundiera jamás. Por invertir hasta vuestra última gota de esfuerzo y ganas en que haya sido y sea quien soy, alguien feliz. Por hacerme sentir una hija orgullosa y presumir cada día de vosotros. Porque de los dos he aprendido todo aquello que necesito en la vida. Por quererme incondicionalmente.

Ter, Ro. Gracias hermanas y mejores amigas. Porque tengo la suerte de tener dos pilares como vosotras. Por no separaros nunca de mi, por sentiros a mi lado aunque físicamente no estuviéramos juntas. Por haberme corregido cada mail, propuesta, escrito. Por dejarme enseñaros cada presentación y discurso. Por escucharme. Por leerme la mente, incluso antes de pensarlo. Por ser mis medias langostas. Por acabar cada frase de película o serie que digo. Porque no sé que haría sin vosotras y porque cada éxito mío es y será siempre un éxito vuestro. Os quiero.

A mi Pa. Gracias a mi mejor amigo, compañero, confidente, novio y marido. Por estar conmigo en los buenos y en los menos buenos momentos. Por darme siempre palabras de apoyo y abrazos. Por hacerme ser una luchadora hasta el final del combate. Por no dejar que me rinda. Porque este camino empezó a tu lado y no has dejado de animarme ni un solo momento. Por mirarme con orgullo y por mimarme. Porque te admiro y creo que conseguirás todo aquello que te propongas y más. Por hacerme feliz. Gracias por venir a buscarme siempre, daba igual la hora y el lugar: a casa, a Alcalá, al CNIO, pero sobre todo, gracias por venir a buscarme a NYC. Te quiero.

Summary

Protein kinases are the enzymes in the cell that catalyze phosphorylation reactions. They are essential for almost all cellular processes and many of them are considered promising pharmaceutical targets since they are involved in a large number of tumorigenic functions such as proliferation, immune evasion, anti-apoptosis, metastasis and angiogenesis. The progress in high-resolution structure determination techniques has contributed enormously to a better understanding of the structural basis of kinase regulation and the associated structural plasticity. However, because of the high sequence and structural conservation across the kinome, new efforts are required that combine a variety of methodologies, which in particular exploit the differential dynamical behaviour of kinases.

In the following doctoral thesis different computational methodologies are employed to study three topics related to phosphorylation:

1. Understanding the reaction mechanism of phosphorylation and dephosphorylation:
   - Using PKA and GSK3β as model kinases we perform molecular dynamics simulations and carry out hybrid quantum mechanics/molecular mechanics (QM/MM) calculations on the evolution of the Michaelis complexes formed between these kinases and their bona fide substrates towards the respective phosphorylated products and characterize each step of the phosphorylation reactions in atomic detail paying particular attention to the roles and fates of the catalytic metal ions .
   - We analyse the dephosphorylation reaction catalyzed by the SHIP2 inositol phosphatase. Models of the two substrates, PI(4,5)P3 and IP4, in complex with SHIP2 phosphatase are built to understand the reaction mechanism in atomic detail . In addition, Principal Component Analysis and molecular dynamics simulations are used to study the allosteric role of the C2 domain and to propose and test different mutants with a view to confirming or rejecting our hypothesis.

2. Analysis of conformational changes involved in the activation of two prototypical kinases:
   - Free energy calculations using umbrella sampling and metadynamics are applied to validate the energetic profiles of the opening and closing of the activation loop in non-receptor Abelson tyrosine kinase (Abl) codificated in the protooncogene *ABL1* and to characterize the differences between the phosphorylated and the unphosphorylated forms of this pharmacologically important enzyme.
   - Molecular dynamics simulations and normal mode analysis are performed on focal adhesion kinase (FAK), another non-receptor tyrosine kinase involved in cancer, in the presence or absence of ATP/Mg^{2+} in order to understand the allosteric effect of ATP on the conformational and dynamic properties of the enzyme.

3. Computational search for specific protein kinase inhibitors:
   - We perform extensive molecular dynamics simulations of the apo enzymes to identify transient and potentially targetable allosteric pockets.
   - We calculate molecular interaction fields and putative hotspots on the active site and regulatory domains of these kinases to characterize the potential ligand-binding sites.
   - We make use of a variety of docking tools to identify new potential hits present in chemical libraries and/or fragment databases (large-scale virtual screening).

Abbreviations

Abl Abelson Kinase

AMBER Assisted Model Building with Energy Refinement

AMP Adenosine monophosphate

AMP-PNP Non-hydrolysable ATP analogue adenylyl imidodiphosphate

AP endonucleases Apurinic/Apyrimidinic endonucleases

ATP Adenosine triphosphate

A-loop Activation loop

Bcr-Abl fusion gene: “Breakpoint Cluster Region" gene and Abl1 gene

BE Bias exchange

BLA Biologic license application

CHARMM Chemistry at HARvard Macromolecular Mechanics

CGenFF CHARMM Generate Force Field

CML Chronic Myeloid Leukemia

CV Collective variable

C-lobe C-terminal lobe

DFG Asp-Phe-Gly

EMEA European MEdicines Agengy

Ev Eigenvector

EGFR Epidermal growth factor

FAK Focal adhesion kinase

FAT Focal adhesion targeting domain

FBLD Fragment-based lead discovery

FDA Food and Drug Administration

FERM F for 4.1 proteins, Ezrin, Radixin, Moesin

FGFR-2 Fibroblast growth factor receptor 2

^{19}F-NMR Fluorine-19 nuclear magnetic resonance

GROMACS

GSK3β Glycogen synthase kinase 3β

GTP Guanosine triphosphate

HTS High-throughput screening

IGFR-l Insulin-like growth factor receptor 1

INPP5B Inositol-1,4,5-trisphosphate 5-phosphatase

IPP5C Inositol polyphosphate 5-phosphatase

IP_2 Ins-1,4-P2/inositol 1,4-bisphosphate

IP_3 Ins-1,4,5-P3/inositol 1,4,5-trisphosphate

IP_4 Ins-1,3,4,5-P4/Inositol 1,3,4,5-tetrakisphosphate

IR Insuline receptor kinase

KKB Human Kinome Knowledge

MD Molecular dynamics

metaD Metadynamics

MM Molecular Mechanics

NCE New Chemical Entity

NCI National Cancer Instiute

NMA Normal mode analysis

NME New molecules entities

NMR Nuclear magnetic resonance

N-lobe N-terminal lobe

OCRL Lowe sculocerebrorenal syndrom

PBC Periodic boundery conditions

PCA Principal Component Analysis

PDB Protein Data Bank

PI3K phosphoinositide 3-kinase

PH Pleckstrin homology domain

PH-R Pleckstrin homology related domain

PIP5Kiγ phosphatidylinositol 4-phosphate 5 kinase γ

PIP_2 PI(4,5)P2/PtdIns-4,5-P2/phosphatidylinositol (4,5)-biphosphate

PIP_3 PI(3,4,5)P3/PtdIns- 3,4,5-P3/phosphatidylinositol (3,4,5)-trisphosphate

PKA cAMP-dependent protein kinase/protein kinase A

PKI Protein kinase inhibitor

PMF Potential of Mean Force

PTEN

PTK2 Protein Kinase Tyrosine 2

PtmetaD Parallel tempering metadynamics

PX Phox homology domain

PYK2 Prolyne-rich tyrosine kinase

PyMOL Python-enhanced molecular graphics tool

QM Quantum mechanics

QM/MM Hybrid quantum mechanics/molecular mechanics

RMSD Root Mean Square Deviation

RMSF Root Mean Square Fluctuation

SAM Steril Alpha Motif domain

SHIP2 SH2 domain-containing inositol 5-phosphatase

SH2 Src Homology 2 domain

SPR Surface Plasmon Resonancy

Src Sarcoma-family kinases

TPX2 Targeting protein for Xklp2

US Umbrella Sampling

VS Virtual Screening

WHAM Weighted histogram analysis method

WTE-metaD Well-tempered metadynamics

X-ray Crystallography

5-Ptase 5-phosphatase

Index

Chapter 1

1. Introduction

1.1 Protein Kinases and phosphatases

1.1.1 PROTEIN PHOSPHORYLATION AND THE TRANSDUCTION OF EXTRACELLULAR SIGNALS

"... In the preceding paper [1] *it was found that phosphorylase, as extracted from resting muscle, is predominantly in the b form, i.e. requiring AMP. It has been determined that this enzyme can be converted readily to phosphorylase a in the cell-free extracts. The requirements for the reaction of phosphorylase b include a divalent metal ion and, under certain conditions, ATP. Whether this implies that during conversion there is a direct phosphorylation of the enzyme or the formation of an "active" intermediate cannot be stated at this time* [2]*"*.

In the late 1950s Krebs and Fisher discovered reversible protein phosphorylation as a biological regulatory mechanism through their elegant work on the interconversion reactions of the two forms of glycogen phosphorylase and phosphorylase kinase. Subsequently, the discovery of cAMP-dependent protein kinase (PKA)[3], which was the first protein kinase to have had its X-ray crystallographic structure elucidated and served as a prototype for studies on protein kinases in general for many years. Then Src kinase, and others gave the first indications about the diversity of the protein kinase family that we recognize nowadays. In fact, protein kinases make up one of the largest gene families encoded by eukaryotic genomes [4].

Today, the protein phosphorylation reaction has been thoroughly studied and it is the most common post-translational modification used in signal transduction [5], [6]. Protein kinases are the family of enzymes that catalyse the transfer of the γ-phosphate from ATP to specific amino acids in proteins, namely Ser, Thr and Tyr, in eukaryotes [7]. These enzymes, in fact, serve as essential switches that regulate biological

functions. Although most eukaryotic protein kinase domains are structurally similar and have similar modes of regulation [8] they differ in terms of the charge and hydrophobicity of surface residues, and these differences are important for substrate specificity (Figure 1.1-a). If the protein kinases are in charge of phosphorylating targeted proteins, the phosphatases dephosphorylate them (Figure 1.1-b). This reversible modification provides an OFF/ON switch that controls many diverse cellular processes including metabolic pathways, signaling cascades, intracellular membrane traffic, gene transcription, and movement.

The importance, then, of having one (or more) phosphate(s) group(s) added to or removed from a protein makes the difference between it being in an active or an inactive state. Moreover, kinases not only add phosphates to other proteins but also can, sometimes, phosphorylate themselves (*cis* or *trans* autophosphorylation) [9] or be phosphorylated by another kinase.

In the last 50 years we have learnt that a number of diseases result from disregulation of kinases which can arise, most often, through activating or inactivating mutations [10]. For this reason, kinases are currently considered as important drug targets. Here we study the mechanistic details of both the phosphorylation reaction, carried out by kinases, and the dephosphorylation reaction, which is catalyzed by phosphatases. We also attempt to gain further insight into the mechanisms of activation and inhibition of several protein kinases, including Ser/Thr kinases (PKA, GSK3) and Tyr kinases (Abl, FAK) [11].

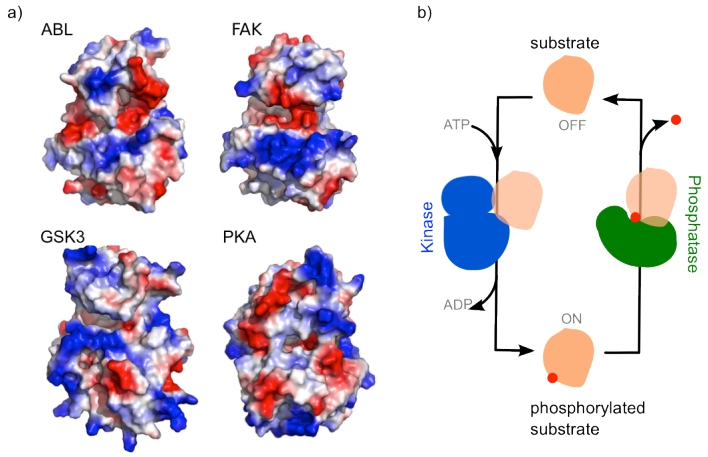


Figure 1.1. a) Representation of the molecular electrostatic potential on the solvent-accessible surface of four different protein kinases. b) Example of an "on/off" activation/inactivation mechanism mediated by a phosphatase/kinase pair.

1.1.2 A COMMON AND CONSERVED FEATURE: THE PROTEIN KINASE DOMAIN

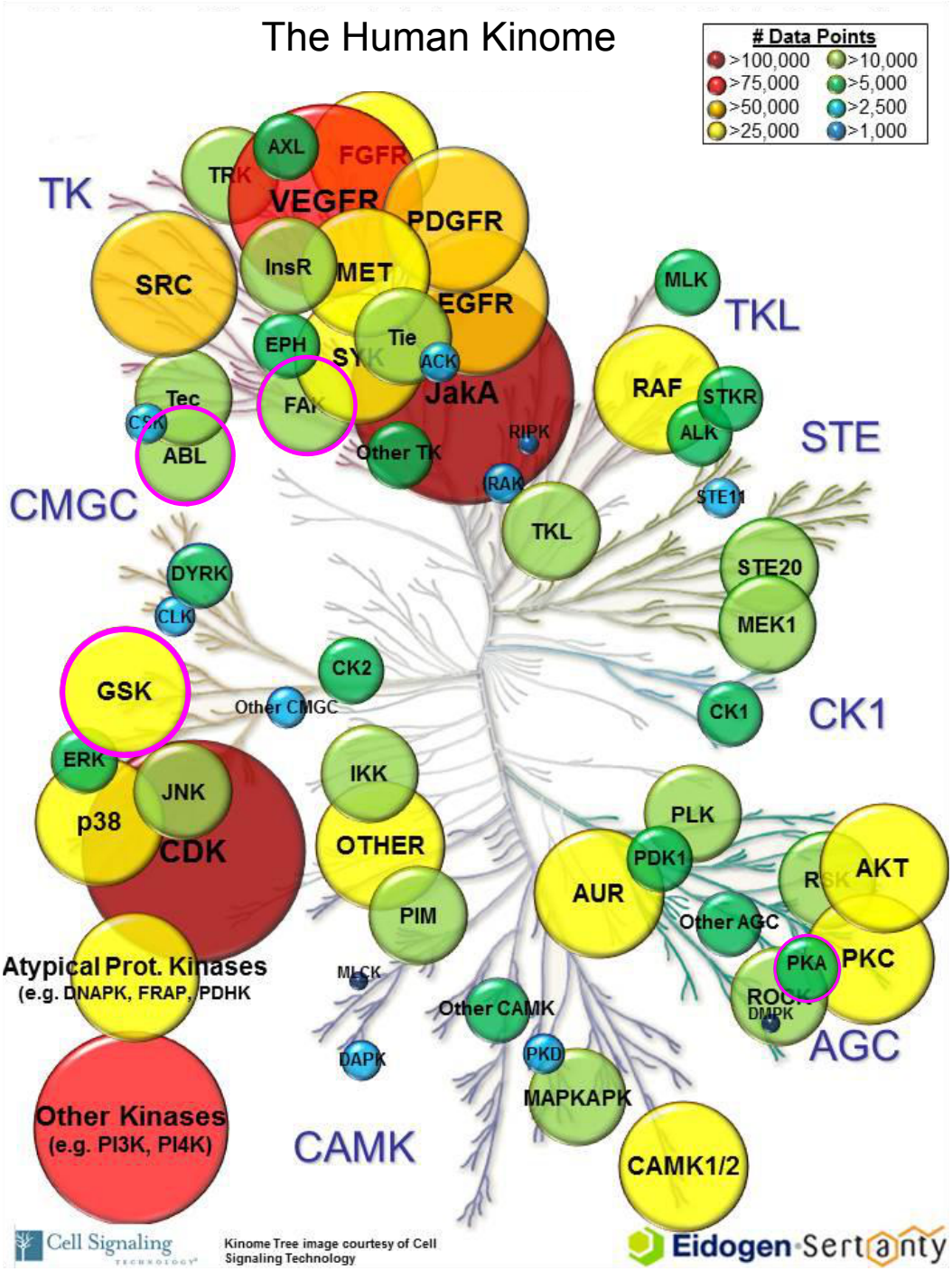


Figure 1.2. Distribution of kinase knowledgebase (KKB) bio-activity data points by family - Q3 2014 Release [211]. The KKB is an Eidogen-Sertanty's database of kinase structure-activity and chemical synthesis data, where the most relevant data such as in vitro enzyme activity data and cell-based activity and toxicity data are selected for the development of advanced activity models (eADME and eTox) designed to support medicinal chemists during all project stages of drug discovery. Marked with a coloured magenta circle are the kinases studied in this thesis: GSK3, Abl, FAK and PKA.

Sequencing of whole eurkaryotic genomes allowed the definition of complete kinomes and revealed the full magnitude and diversity of the family. The human kinome (Figure 1.2), for example, encodes at least 518 protein kinases, many more if one considers splice variants. This corresponds to nearly 2% of the genome and attests to the extraordinary importance of this family in regulating biological events [12].

While the number of protein sequences deposited in public databases continues to expand [12], [13], progress in assigning a function to all of the encoded proteins remains slow. To accelerate this process, a better understanding of the relation between the sequence and the structure makes the third vertex of the triangle, the function, more accessible.

The kinase core was classified into 12 subdomains by Hanks and Hunter, when they discovered a bunch of conserved sequence motifs [14], and these subdomains could be mapped when the first protein kinase was crystallized and its structure solved [15]. A prototypical protein kinase consists of two structurally and functionally distinct lobes, a smaller N-terminal (N)-lobe mainly formed by a five-stranded β-sheet coupled to a helix, the conserved α-helix C, and the larger C-terminal (C)-lobe, consisting mostly of α-helices plus a β-sheet (Figure 1.3). A binding site for ATP and the metal cations coordinating the ATP phosphates is located in a large groove between the two lobes. As a template for the kinase core structure and display of the key structural features present in the protein kinase domain we will use cAMP-dependent protein kinase (PKA) (Figure 1.3) [16].

1.1.2.1 Active site cleft

This cleft is located between the N-terminal lobe (N-lobe) and the large C-terminal lobe (C-lobe). It is the location where the substrate protein is recognized by interacting with the activation segment or the activation loop (A-loop). Most of the residues in this cleft are conserved and contribute either directly or indirectly to ATP binding and/or catalysis. Hydrophobic residues sandwich the adenosine ring whereas polar residues interact with the phosphates of ATP and coordinate the divalent cation(s). Hydrogen bonds are formed between the adenosine ring and the hinge region of the kinase domain, which connects the two lobes. In addition, a number of conserved residues in the C-lobe act as the floor of this active site stabilizing the pocket and facilitating the phosphoryl transfer reaction.

1.1.2.2 Glycine-rich loop or P-loop

This small loop located at the beginning of the N-lobe is critical for the correct positioning of the ATP. It has been shown to be highly flexible and one of the most mobile parts of the molecule, as we will see later. It contains a glycine-rich motif (GxGxxG) where the glycine backbone amides interact with the ATP phosphate groups.

1.1.2.3 α-helix C

The α-helix C is recognized as one of the central features in controlling kinase activity and can switch between an inactive and an active conformation. It contains a conserved glutamate that can twist from an inward position in the active form interacting with a conserved lysine in the active site, to an inactive outward position facing the solvent.

1.1.2.4 Activation loop

This large loop of 20-25 amino acids [17] is also considered as one of the critical features regulating the kinase activity. It can assume a large number of conformations, but for full activity it needs to adopt a fully open conformation. In most kinases the activation loop is stabilized in the active conformation by the phosphorylation of one residue in this loop.

1.1.2.5 DFG-motif

This motif, formed by three amino acids: Asp-Phe-Gly, is highly conserved at the beginning of the activation loop and is structurally coupled to the activation loop, the α-helix C and ATP thereby playing a key role in kinase regulation. The aspartate is responsible for coordinating the divalent cation (Mg^{2+}/Mn^{2+}), which places the phosphates of ATP correctly for phosphotransfer in the so-called DFG-in conformation. On the other hand, the phenylalanine is packed below the α-helix C for the correct positioning of the helix and the activation loop. In some inactive kinase conformations the backbone torsional angle, ϕ, of the DFG aspartic acid turns almost 180º into a so-called DFG-out conformation, where the phenylalanine and the aspartate switch positions in such a way that ATP cannot fit into the active site.

1.1.2.6 R-spine

The regulatory (R)-spine is another key motif in kinase regulation that forms a hydrophobic core that bridges the N- and C-lobes. It is found in all active kinases but is

missing in the inactive ones. The R-spine is formed by a well conserved spatial motif of four residues: two leucines located in the N-lobe, a tyrosine (or a histidine in some cases), and a phenylalanine from the C-lobe. The disruption of this arrangement leads to an inactive conformation [4].

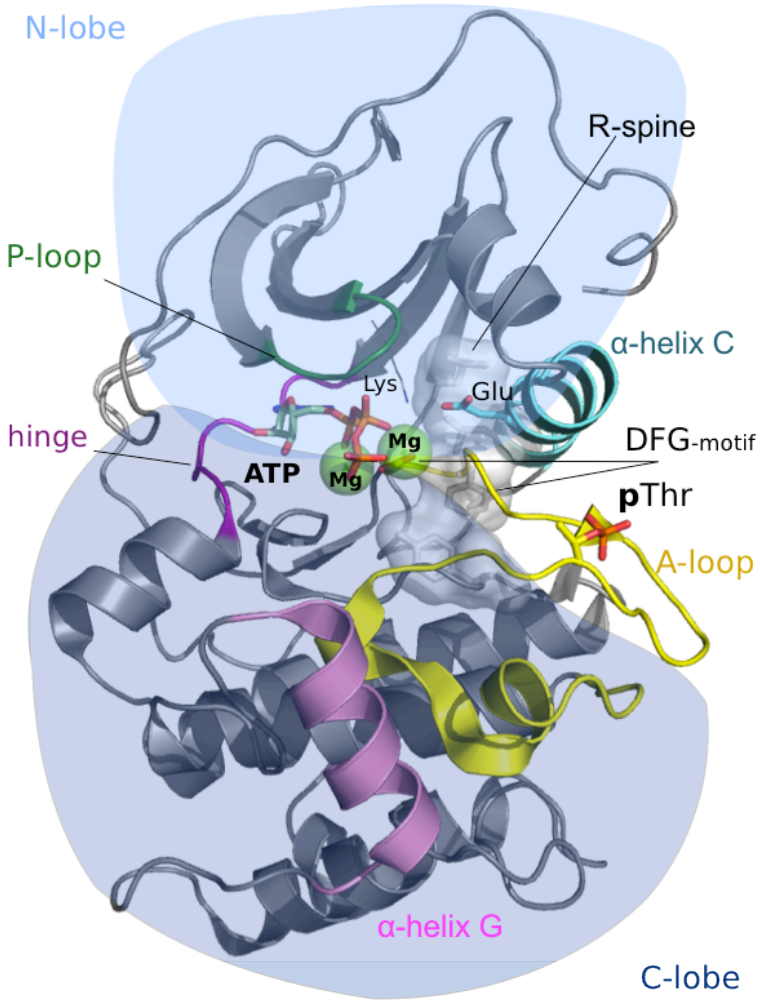


Figure 1.3. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase (PKA) in complex with ATP and two Mg^{2+} ions (PDB code: 4HPU). Important parts are highlithed such as: R-spine in grey surface and sticks representation, α-helix C in cyan, P-loop in green, ATP and Mg^{2+}s in sticks and spheres respectively, the hinge in purple, the α-helix G in pink and the A-loop in yellow with the phosphorylated **p**Thr in sticks.

1.1.2.7 The gatekeeper

The so-called "gatekeeper" refers to a single residue located at the back of the ATP binding clelft (Figure 1.6). Depending on the kinase, some of them have a large residue such as Leu, Met or Phe, and others a smaller residue (Thr, Val). This difference might

play an important role in the design of new kinase inhibitors. It has been shown that mutations of the gatekeeper residue are responsible for drug resistance [18].
Once all the features are correctly placed, i.e. the R-spine assembled, the α-helix C well oriented so that the Lys-Glu salt bridge can be formed, the A-loop outside the active site, and the DFG-motif correctly positioned, the ATP and the Mg^{2+} ions bind and catalysis can proceed.

1.1.3 DYNAMIC FEATURES OF KINASE ACTIVATION

Most of the dynamic features playing a role in catalysis have already been described in the previous section and many of them have been related to the active or inactive forms. While the active conformations are indeed very similar, inactive states of kinases have varied structures because they are not constrained by the need to catalyze the phospho-transfer reaction. However, each kinase has to be treated as a special case in the light of all the available structural information although common features could be shared. For example, in the inactive forms, it is usual to find the α-helix C including the glutamate within the helix, pointing outside the ATP binding site and the A-loop blocking the binding site in order to avoid catalysis. Another common inactive conformation is the DFG-motif in an "out" position. If the phenylalanine is located in the ATP site, there will be no possibility of ATP binding or Mg^{2+} coordination, whereas for the DFG-in conformation other features, such as the R-spine or the A-loop, will define active versus inactive states. Indeed, for activated kinase conformations the common features are more restricted and only a couple of requisites must be present: (i) the P-loop or glycine-rich loop must adopt a position such that ATP is able to orient correctly and especially place the γ-phosphate in a position ready for catalysis, and (ii) the α-helix C has to be pointing inwards with the conserved Glu establishing a salt bridge with the Lys in the active site. In addition to the kinase core domain it is important to underline that in most cases other domains or motifs, linkers and tails are also involved in shifting the equilibrium from active to inactive states, or vice versa. For example, the SH2 and SH3 domains in many non-receptor tyrosine kinases, the PH, PX and C2 domains involved in lipid signaling, a linker or FERM domain in FAK and other domains associated with calcium or mediating interactions with other proteins (SH3, SAM…) play important regulatory roles as well [12], [19], [20], [21].

1.1.4 ALLOSTERIC CONTROL OF KINASE ACTIVITY

Allosteric regulation is a type of control mechanism in which binding of a regulatory molecule at one site on a protein causes a change in its conformation in such a way that the activity at another site on the protein is altered (Figure 1.4). The concept of allostery was first articulated more than 50 years ago [22] when the "action at a distance" phenomenon was difficult to interpret in the absence of structural information or site-specific biochemical data [23]. Even nowadays describing allostery in quantitative terms is essential to thoroughly understand processes such as cellular signaling and disease. In many cases it remains a biophysical enigma. With respect to protein kinases, they rarely act alone but rather form part of a regulated cascade, in many cases modulated by a network of allosteric events. These events include: (i) dimerization, as in epidermal growth factor receptor (EGFR) forming homodimers [24], (ii) change in motif conformations as in Abl, c-Src or PKA (as typical examples for α-helix C specific rotations and DFG-motif orientations [25]), (iii) additonal domains and linker interactions as it happens in FAK, PYK2 or Abl kinases [19], [26], and (iv) binding of other proteins, as illustrated by the interaction of Aurora kinases A and B with a microtubule-associated protein, TPX2 or that between cyclins and cyclin-dependent kinases [27]. Although many details of these complex processes are still unclear, recent technological advances and also long-timescale molecular simulations have helped to provide a more comprehensive dynamic picture of these events.

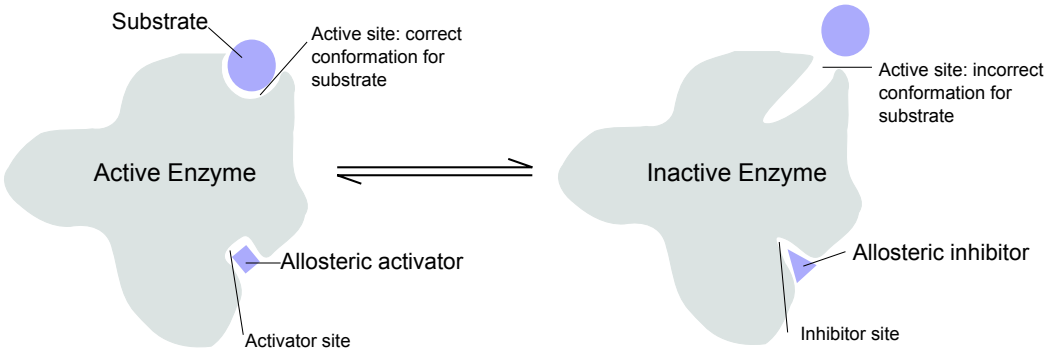


Figure 1.4. Allosteric inhibition scheme: on the left, a substance binding outside the orthosteric site inducing the correct conformation of the active site activating the enzyme. On the right side, an inhibitor is bound at a specific site and modifies the active site preventing the binding of the substrate.

1.1.5 PROTEIN KINASES AND CANCER

"A term for diseases in which abnormal cells divide without control and can invade nearby tissues." Definition of *cancer* by the National Cancer Instiute (NCI).

As mentioned above, the most common signaling process in a human cell is phosphorylation. If we consider that this phosphorylation is turned on and off by protein kinases and protein phosphatases, respectively, abnormal activity of these enzymes can produce either active kinases without control or phosphatases unable to inactive them or both. These uncontrolled processes may end up in the deregulation of many regulatory pathways involved in cancer, such as those affecting growth, survival, neovascularization, metastasis and invasion [28], [29]. Besides the well-known fact that phosphorylation regulates most aspects of cell life, a compilation of cancer genes indicates that mutations in more than 1% of genes contribute to human cancer. Indeed, the most common domain encoded by cancer genes is the protein kinase domain [30]. Therefore, taking into account both informations regarding phosphorylation and gene mutations, protein kinases have become the most intensively pursued anticancer drug targets [31].

In 1946 Goodman and Gilman published a landmark study in cancer therapy. Nitrogen mustard agents that were discovered in World Ward II to induce drastic lymphoid cell depletion were shown accidentally to produce remarkable responses in human tumors [32]. Two years later, Farber followed a more rational target-based strategy and discovered the anti-folates (e.g. methotrexate) as anti-leukemia drugs [33]. Then natural products derived from plants (e.g. vincristine) and microorganisms (e.g. actinomycin, mitomycin) came out. In recent years there has been an evolution from this type of cytotoxic drugs, which non-specifically kill proliferating cells, towards more specific target-oriented agents that have completely changed the cancer treatment landscape. The class of kinase inhibitors, in particular, has progressed astonishingly in the past decade and a number of specific agents are currently approved for clinical use (Table 1.1).

Kinase Inhibitors	YEAR of FDA approval	Known targets	Indications
Afatinib (Gilotrif)	2013	EGFR, ERBB2 (HER2)	Non-small cell lung cancer
Axitinib (Inlyta)	2012	VEGFR-1, 2 and 3, PDGFR-β, c-Kit	Renal cell carcinoma
Bosutinib (Bosulif)	2012	Bcr-Abl, SRC	Chronic myelogenous leukemia
Crizotinib	2011	ALK,Met	NSCLC with Alk mutation
Vemurafenib (Zelboraf)	2011	BRAF	Melanoma
Ruxolitinib (Jakafi)	2011	JAK	myelofibrosis
Pazopanib (Votrient)	2009	VEGFR-1, 2 y 3; PDGFR-α y β, FGFR1 y 3, c-Kit, Itk, Lck, c-Fms	Soft tissue sarcoma Renal cell carcinoma
Nilotinib (Tasigna)	2008	Bcr-Abl, c-Abl, PDGFR, c-KIT, DDR1	Chronic Mieloid Leukimia (CML) imatinib resistant
Lapatinib (Tykerb)	2007	ERBB2 (HER2)	Breast cancer
Sunitinib (Sutent)	2006	c-KIT, PDGFR, VEGFR	Renal cell carcinoma Gastrointestinal stromal tumor (GIST)
Dasatinib (Sprycel)	2006	Bcr-Abl, familia SRC, fam. TEC	CML imatinib resistant
Sorafenib (Nexavar)	2005	c-KIT, PDGFR, VEGFR (b-raf)	Renal cell carcinoma
Erlotinib (Tarceva)	2004	EGFR, ERBB2 (HER2)	Non-small cell lung cancer (NSCLC), pancreatic cancer
Gefitinib (Iressa)	2003	EGFR, ERBB2 (HER2), HER4	NSCLC
Imatinib (Gleevec)	2001	Bcr-Abl, c-Abl, PDGFR, c-KIT, DDR1	CML, GIST

Table 1.1 Drugs targeting protein kinases approved by the FDA since 2001. *Type I* inhibitors are highlighted in green; *Type II* inhibitors are highlighted in pink (see below).

Kinase inhibitor design has benefited immensely from advances in protein crystallography and computational methodologies. However, a number of significant challenges remain. First, selectivity is often poor for this kind of ligands. Most currently known kinase inhibitors target the ATP binding site, which at the same time is the most conserved part of the kinase core. This makes the achievement of selectivity and the avoidance of toxicity very difficult. Therefore, understanding the structural basis of kinase selectivity and developing selective inhibitors to target every member of the kinome is still a major challenge in the field. Second, understanding the mechanistic basis of unexpected side effects observed during preclinical and clinical trials would

provide information on which inhibition should be avoided in order to prevent repeated mistakes. And third, to be able to apply efficient combination therapies it is crucial to develop more accurate models of how the different kinase signaling cascades are interconnected and how they get reprogrammed in response to oncogenic events or drug resistance.

Dealing successfully with all of these challenges requires a highly interdisciplinary collaboration between clinical investigation, pharmacology, chemistry, biology, protein crystallography and computation. This Ph.D. thesis involves several of these disciplines: pharmacology, chemistry, crystallography, biology and computation, with the main objectives being not only to investigate new selective inhibitors or figure out what are the regulatory mechanisms of these proteins, but also to show that an interdisciplinary collaboration enriches research in general and provides new insights, particularly in cancer, which otherwise would be very difficult to obtain.

1.1.5.1 Kinase inhibitors and binding sites

The different types of kinase inhibitors can be characterized by looking at the pocket and state of the targeted protein, active or inactive [34]. Even though the vast majority of the kinase inhibitors discovered to date are ATP-competitive *(I and II)* we can distinguish between three different types of inhibitors (Figure 1.5):

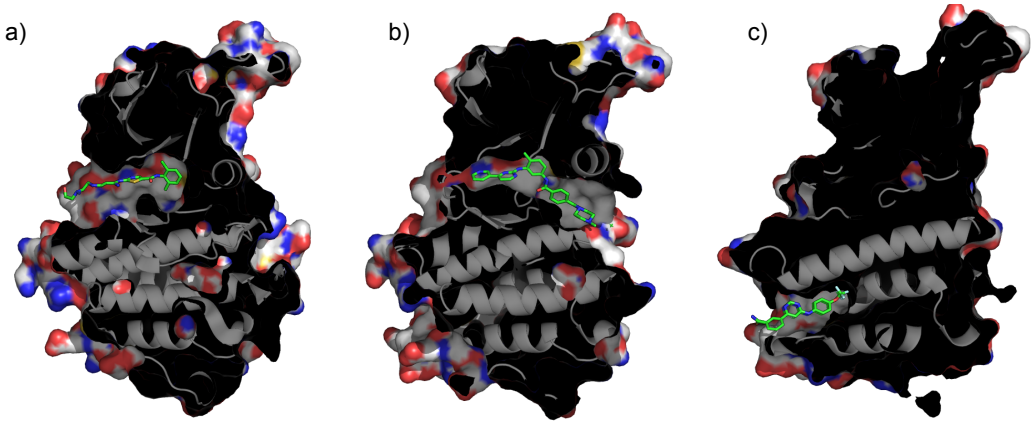


Figure 1.5. X-ray structures of Abl in complex with the different types of inhibitors. a) *Type I*, represented by dasatinib (PDB code: 2GQG), b) *Type II*, with bound imatinib (PDB code: 1OPJ), and c) *Type III*, as compound GNF-2 (PDB code: 3K5V). Each protein structure is represented as a cartoon and the enveloping surface has been sliced to enable appreciation of the different pockets. The inhibitors are shown in stick representation with C atoms coloured in green.

- Type I inhibitors: are ATP-competitive inhibitors that normally target the protein in its active form, with the activation loop outside the binding site and the highly conserved DFG-motif in the so-called 'in' state [35]. In this conformation the phenylalanine or the DFG occupies the allosteric pocket beside the ATP-binding site and below the α-helix C (Figure 1.5-a). These inhibitors usually mimic the purine ring of the adenine moiety of ATP and establish hydrogen bonds with "the hinge region" of the kinase. There are several Type I inhibitors approved by the FDA [36] and the EMEA (Table 1.1). Although some Type I inhibitors are highly selective (e.g. erlotinib), the usual tendency is low kinase selectivity (e.g. dasatinib [37]) because the ATP pocket is conserved throughout the kinome .

- Type II inhibitors are still ATP-competitive inhibitors but they recognize an inactive kinase conformation in which the DFG motif is in an "out" position. Here the phenylalanine is positioned looking towards the catalytic site and stabilizing an inactive unphosphorylated form. This conformation leads to opening of a hydrophobic pocket at the back of the ATP binding cleft, located between the DFG motif and the α-helix C and very proximal to "the gatekeeper" residue that "guards" entry into the hydrophobic "back-pocket". This extends the ATP-binding site, which occupies almost the whole cleft between the "N-lobe" and the "C-lobe" (Figure 1.5-b) [36], [38]. Since this back-pocket is less conserved because it is not required for catalytic activity, it has better prospects as a target site for designing selective inhibitors. A very good example is the first tyrosine kinase inhibitor, imatinib. Designed in the early 1990s, and considered to be a "magic bullet" against chronic myeloid leukemia, it was one of the first inhibitors obtained following a rational drug design strategy. By establishing six hydrogen bonds, several hydrophobic and π-stacking interactions, imatinib inhibits the Bcr-Abl kinase (and some other kinases) very effectively.

- ***Type III inhibitors*** are allosteric ligands that bind outside the ATP-binding site, which means they are non-competitive with this nucleotide. Inhibitors belonging to this category have the highest degree of kinase selectivity because they can bind to pockets that are unique to particular kinases [39]. An example of allosteric ligand is the GNF-2 compound that binds to the myristoyl-binding pocket in the C-terminal lobe of Abl (Figure 1.5-c).

1.1.5.2 Resistance and selectivity

As the majority of kinase inhibitors have cytotoxic effects, cells are subjected to a strong selective pressure for acquiring resistance through mutations in order to avoid drug binding to the targeted kinase. One of the best characterised and most common examples is the resistance conferred by mutations of the gatekeeper residue (Figure 1.6). This amino acid is located in a key site because it allows access to a hydrophobic pocket where the Type II inhibitors are positioned [40]. When a mutation occurs, these interactions between the hydrophobic pocket and the ligand are lost and the ligand interaction is no longer stable.

However, these mutations do not affect ATP binding so the catalytic activity is not affected. The gatekeeper mutation was shown to be a common occurrence in the clinic, although the jury is still out as to whether it is pre-existing in a small number of cells in a heterogeneous tumour or acquired following inhibitor treatment. Regardless of this debate, several strategies are being investigated to solve this resistance problem [38]. One is designing inhibitors that allow different amino acids at the gatekeeper position; another is targeting alternative binding sites, such as allosteric sites to which *Type III* inhibitors bind. On the other hand, co-targeting other pathways that are used as "escape pathways" may prevent resistance.

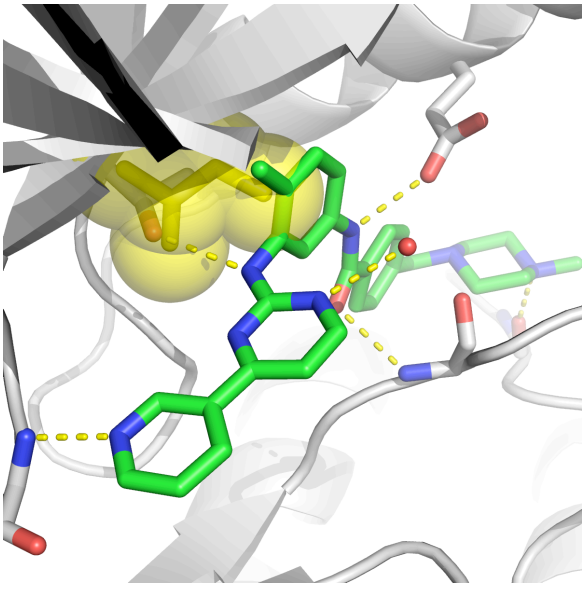


Figure 1.6. Abl kinase (grey cartoon) in complex with imatinib (sticks with C atoms in green). All the polar interactions between the inhibitor and the protein are shown as yellow dashes. The Ile replacing Thr315 is displayed as yellow spheres that show the steric clashes with imatinib.

Another pressing challenge is achieving selectivity. At present, several companies offer to screen for off-target binding in the kinome. This information is crucial to understand the different mechanisms for potential toxicity and to figure out which tumors may respond to a particular inhibitor. We will see some examples in the following section.

1.1.6 PROTEIN KINASES STUDIED IN THIS THESIS

1.1.6.1 Abelson Kinase (Abl) and BCR-Abl

The cellular form of the Abl kinase (c-Abl) is a nuclear and cytoplasmic nonreceptor tyrosine kinase involved in a variety of cell growth and differentiation processes. c-Abl is normally in a state of low catalytic activity and contains no phosphorylated tyrosine residues. In contrast, BCR-Abl is a constitutively activated form of c-Abl that arises by leukemiogenic fusion [41] following a characteristic reciprocal translocation between human chromosomes 9 and 22 that gives rise to the so-called Philadelphia chromosome which contains a fusion of c-ABL and BCR genes,. The resulting BCR-Abl fusion protein elevates Abl tyrosine kinase activity and leads to chronic activation of signaling pathways that are critical for transformation of hematopoietic cells resulting in chronic myeloid leukemia (CML) [42]. This type of myeloproliferative disease represents 15-20% of all leukemias diagnosed [43].

c-Abl is regulated by intramolecular interactions that keep the kinase domain in an inactive conformation. The inactive c-Abl conformation involves several intramolecular interactions including the SH3 domain and the proline-rich sequence present in the SH2-kinase linker. Also the SH2 domain interacts with the C-terminal lobe of the kinase domain. All these interactions participate in the kinase auto-inhibition. Opening the conformation has been predicted to induce catalytic activation [44]. In addition, phosphorylation of two key residues, namely Tyr 245 in the linker and Tyr 412 in the activation loop, is required for c-Abl activation by stabilizing Abl in an open and active conformation [45].

Since a very high-affinity and specific inhibitor for Abl, known as imatinib (Gleevec® or Glivec®), was discovered for CML treatment and FDA-approved in 2001, the statistics on this cancer have changed enormously. Before kinase inhibitors, the medium survival time for CML patients was 2 to 3 years from time of diagnosis. Nowadays, more than 90% of patients have a survival rate higher than 8 years. 65-75% of patients stop CML progression by using imatinib, although the treatment has to be continued indefinitely. As a result, many CML patients have a normal life expectancy [46].

Although these are very promising statistics and the prognosis is indeed one of the best for cancer treatments, drug resistance is a major problem, as previously mentioned [47]. For instance, 30% of patients on imatinib treatment acquire drug resistance mutations in the kinase domain of Bcr-Abl or intolerance to imatinib in the first five years [48].

In this context, our studies on Abl receptor provides a new conformational space that could be used for specific rational drug design.

1.1.6.2 Focal Adhesion Kinase (FAK)

This nonreceptor tyrosine kinase encoded in the *PTK2* human gene plays an important role in intracellular signal transduction pathways that promote the turnover of cell contacts with the extracellular matrix, a process that is crucial in the control of cell adhesion and migration. Different studies have found FAK overexpression in most types of solid tumors (thyroid, prostate, cervix, colon, rectum and ovary) [49]. Since FAK overexpression highly correlates with tumor invasiveness and poor prognosis, this kinase is considered an important anticancer drug target [49], [50]. However, it is not clear yet what is the most effective strategy to target FAK. Potential routes are inhibition of the catalytic domain or disruption of crucial protein-protein interactions. Several FAK inhibitors are currently being tested in pre-clinical and clinical settings. The majority of the inhibitors are designed to target the ATP-binding site so as to inhibit the enzymatic activity very efficiently. The problem, as mentioned before, is their selectivity. TAE-226, developed by Novartis, effectively inhibits FAK activity; unfortunately, it also inhibits the insulin receptor kinase (IR) and insulin-like growth factor receptor 1 (IGFR-l) [51]. Recently, another FAK inhibitor, VS-6063, has reached Phase II clinical trials [52] and supports FAK as a critical regulator of the survival of cancer stem cells.

Human FAK consists of an N-terminal FERM domain, a central catalytic domain, and a C-terminal focal adhesion targeting (FAT) domain. The inactive crystal structure reveals a mechanism of inhibition in which the N- terminal FERM domain directly binds the kinase domain, blocking access to the catalytic cleft and protecting the FAK activation loop from Src phosphorylation [53]. Whereas the kinase domain in the autoinhibited structure adopts a conformation similar to that of the isolated kinase domain in the unphosphorylated state, the active phosphorylated FAK kinase rearranges

the activation loop to adopt a conformation that is resilient to FERM inhibition. In a recent study published by our group [21] (*Annex I*), the FAK activation mechanism has been proposed as a sequence of events: first, cell adhesion via integrin receptors to the extracellular matrix results in integrin clustering and the recruitment of focal adhesion proteins at the cytosolic side of the integrins (including FAK and PIP5KIγ) to form adhesion structures that link integrins to the actin cytoskeleton. Then, recruitment of PIP5KIγ results in a local increase of $PI(4,5)P_2$ levels. This phosphoinositide binds the FERM domain of FAK, which results in a relaxed FERM/kinase conformation that promotes FAK clustering at the cell membrane. Finally, this conformational change allows efficient autophosphorylation and recruitment of the tyrosine-protein kinase encoded in the proto-oncogene *SRC*. Src phosphorylates the activation loop of the FAK kinase domain, which results in full activation and release of the kinase from the membrane-clustered FERM domain.

To target specific stages of this activation mechanism, in this thesis we describe the use of a new drug design strategy that combines long-timescale molecular dynamics simulations with large-scale virtual screening of chemical libraries with the aim of discovering new selective and allosteric compounds that can inhibit FAK signalling.

1.1.6.3 Glycogen Synthase Kinase-3β (GSK3β)

Our last kinase example is a serine/threonine protein kinase known as glycogen synthase kinase 3β (GSK3β) that plays central roles in a diverse range of signalling pathways. Dysregulation of GSK3β activity, particularly hyperactivation, is associated with various pathological conditions such as diabetes mellitus, Alzheimer's disease, neurological disorders and cancer, hence the potential therapeutic value of selective GSK3β inhibitors [53, 54, [56]. GSK3β activity is controlled by phosphorylation, and the precise mechanism of phosphoryl transfer and the detailed structural validation of the individual steps involved in catalysis are still not fully understood. However, within the kinase superfamily, protein kinase A (PKA), which is the most widely studied member and that for which the largest amount of structural [16] and biochemical data are available (both for the wild-type enzyme and for selected variants obtained through site-directed mutagenesis), has attracted attention over the years as a suitable model on which to study the general mechanism of phosphoryl transfer from both experimental and theoretical standpoints [57]. This is so because the nature and arrangement of the

amino acids making up the active site are very well conserved across the whole kinase family. Indeed, the phosphoryl transfer reaction was recently characterized in atomic detail using X-ray crystallography and computational analyses [58], as well as other techniques using PKA as a test case. These studies have shown the unique environment provided by the enzyme for the Mg^{2+}-aided binding of both ATP and substrate in suitable geometries for Michaelis complex formation. An important role was demonstrated for the Mg^{2+} ions, together with the P-loop alanine and glycine residues, in promoting the required three-dimensional arrangement of reactive groups to enforce an L-shape in the triphosphate chain of ATP such that the γ-phosphate group is placed in a suitable position for attack. Also important is the lysine residue that acts as a bridge to connect this γ-phosphate of ATP to the phosphorylatable Ser/Thr residue of the substrate. The role of this Lys has been discussed earlier but its amino group has been described both as a possible proton donor for the phosphoryl group or just as a binding aid to fix the Ser/Thr hydroxyl and phosphate groups in place for the reaction to proceed [57]. No previous published work, to the best of our knowledge, has addressed this ambiguity. Understanding the exact phosphorylation mechanism of GSK3β and PKA can help us to understand how to design a better kinase inhibitor.

Although more than fifty GSK3β inhibitors have been reported, there are still many open questions about how some of these inhibitors act, specially lithium. This light monocation, which has been used since 1870 to treat mania patients [59] and was approved by the FDA in 1970 for the treatment of depressive disorders, was one of the first GSK3β inhibitors described. It has been shown that Li^{+} competes with Mg^{2+}, which means that it might bind in the ATP site by replacing one of the Mg^{2+} ions [60]. Although GSK3β is involved in so many signalling pathways, it is intriguing how lithium does not have direct consequences that will lead to activation/suppression of various of these pathways. Moreover, the importance of GSK3β in many apoptotic conditions is further supported by evidence that selective small-molecule inhibitors of GSK3β, in addition to lithium, provide considerable protection from apoptotic cell death [61].

Understanding the exact activation mechanism of GSK3β and inhibitory function of lithium cation may lead to improved therapies for certain diseases in the future [62].

1.1.7 SHIP2 PHOSPHATASE

Kinases and phosphatases are enzymes that catalyse the addition or removal of phosphate groups from specific substrates. While kinases catalyse the transfer of γ-phosphate from ATP (or GTP) to their substrates, the phosphatases catalyse the transfer of a substrate-bound phosphate to a water molecule [63]. In particular, we are interested in the SH2 domain-containing inositol 5-phosphatase (SHIP2), a lipid phosphatase from the large family of inositol polyphosphate-5-phosphatases that belongs to a Mg^{2+}-dependent family of phosphoesterases endowed with the ability to remove the 5-position phosphate from the inositol ring of both inositol phosphates and phosphoinositides [64].

Inositol and phosphatidylinositol phosphates are key modulators of cellular processes such as signal transduction, cell proliferation, apoptosis, vesicle trafficking, cell motility, cytoskeletal organization and transcription [65]. Phosphoinositides are ubiquitous phospholipid constituents of the cell membrane that are made up of a fatty acid backbone, which anchors the lipid in the membrane, attached to a six-membered inositol ring, which can be phosphorylated at the 3, 4 or 5 position to generate a list of signalling molecules. Depending on which site(s) on the inositol ring is/are phosphorylated, the activity of the inositol molecules varies. In the case of 5-phosphatases (5-ptases), PtdIns(3,4,5)P_3 is rapidly degraded at the 5-position forming PtdIns(3,4)P_2 (Figure 1.7).

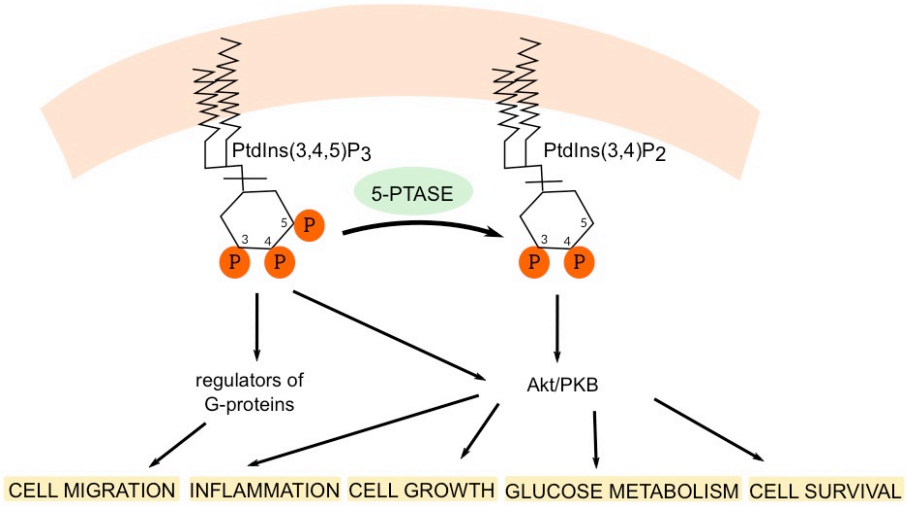


Figure 1.7. Different roles of phosphoinositides PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2 involved in 5-ptase dephosphorylation.

Given the diverse signalling functions of the polyphosphoinositides, the 5-ptases play critical roles in the regulation of many cellular events, in particular membrane trafficking and cell growth [66] and therefore in the pathogenesis of human diseases [67], [68]. Especially, they are important downregulators of the PI3K/Akt pathway (i.e. PTEN) and its dysregulation can promote the development of various diseases, in particular of cancer, reducing apoptosis and allowing proliferation, obesity and type 2 diabetes. For instance, small-molecule inhibitors of SHIP2 were recently shown to lower plasma glucose levels and improve glucose intolerance in diabetic mice [69]. Additionally, there is ample evidence that disregulation of SHIP2 and its close homolog SHIP1 can result in tumorigenesis and specific inhibitors of this enzyme show promising effects on breast cancer and multiple myeloma models [70]. Moreover, 5-ptases have been associated with certain genetic deficiences such as the Lowe sculocerebrorenal syndrome (OCRL), where a 5-ptase encoded by the *OCRL1* gene is deficient in OCRL patients [71].

To date, 10 human 5-phosphatases have been identified (Figure 1.8) showing specificity for discrete substrates that include inositol phosphates Ins(1,4,5)P_3 and Ins(1,3,4)P_4 and phosphoinositide messenger molecules PtdIns(4,5)P_2, PtdIns(3,4,5)P_3 and PtdIns(3,5)P_2 [72] [73] [71].

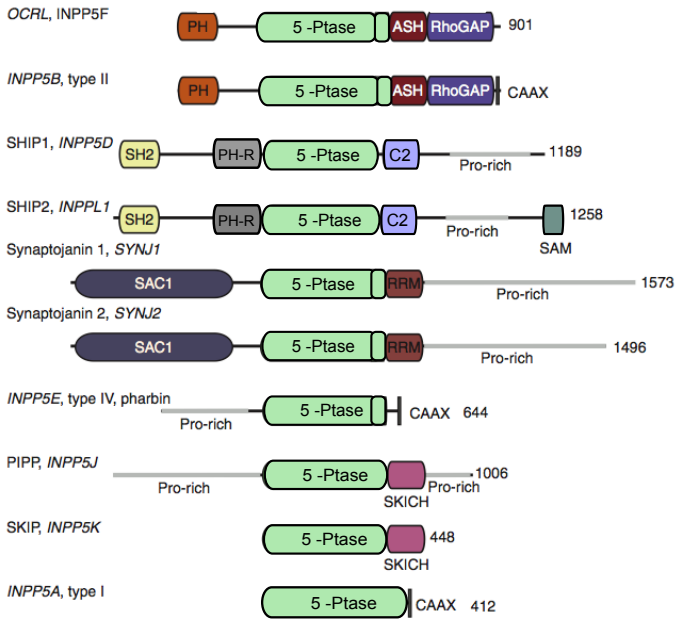


Figure 1.8. Domain organization of the inositol 5-ptases. The gene names (as annotated by the HUGO gene nomenclature committee) are indicated in italics, and alternative names are also given [71].

All 5-phosphatases contain a conserved 350-amino acid central region designated as the "5-Ptase domain" (domain colored in green in Figure 1.8 and Figure 1.9) which demonstrates approximately 30% amino acid identity between family members apart from the 43-kDa 5-phosphatase (5-phosphatase I), which is more extensively diverged.

Moreover, these enzymes contain additional domains partnered with the catalytic domain (Figure 1.8 and 1.9-d). The SHIP enzymes consist of an N-terminal SH2 domain followed by a pleckstrin homology-related (PH-R) domain, the central 5-phosphatase (5-Ptase) catalytic domain, followed by a C2 domain and a proline-rich region (Figure 1.8). In addition, SHIP2 has a SAM domain located at its C-terminus. Upon stimulation, SHIP enzymes are recruited to growth factor and cytokine receptors. Here, we shed some light on the role of the C2 domain and also on the SHIP2 substrate preferences, which have not been discussed previously.

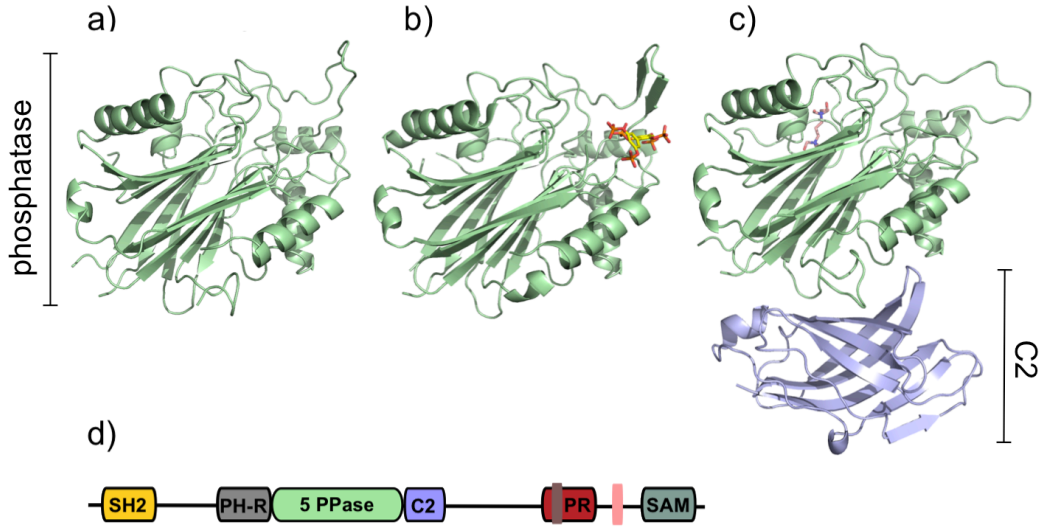


Figure 1.9. Crystal structures of the human SHIP2 ptase: a) SHIP2 with no ligand bound (apo form, PDB code: 3NR8); b) Human SHIP2 in complex with biphenyl 2,3',4,5',6-pentakisphosphate (PDB code: 4A9C); and c) 5-ptase and C2 domains of SHIP2. The domain organization of SHIP2 is shown in different colours in d).

1.2 *In silico* methodologies

1.2.1 MOLECULAR MODELLING

Three-dimensional visualization of molecular structures, as provided by molecular models, has become an essential part of the study of organic and inorganic chemistry as well as biochemistry.

Sharing Dorothy Crowfoot Hodgkin inquisitiveness's at the middle of the past century: *"... I became captivated by the edifices chemists had raised through experiment and imagination—but still I had a lurking question. Would it not be better if one could really "see" whether molecules as complicated as the sterols, or strychnine were just as experiment suggested?"* [74]. Indeed, this is what molecular models and molecular graphics programs have addressed, the possibility of "seeing" molecules.

Molecular modelling has been present since the early 50's, when the first satisfactory model of a large molecule, a peptide α-helix, was constructed by Linus Pauling based on X-ray diffraction data and using the general principles of molecular structure [75]. Then in 1958 the first protein model, myoglobin, was built by Kendrew and co-workers [76]. Over time we have witnessed an exponential growth of biomolecular structure models with atomic resolution, based on experimental data (Figure 1.10) thanks to X-ray crystallography, cryoelectron microscopy and NMR spectroscopy.

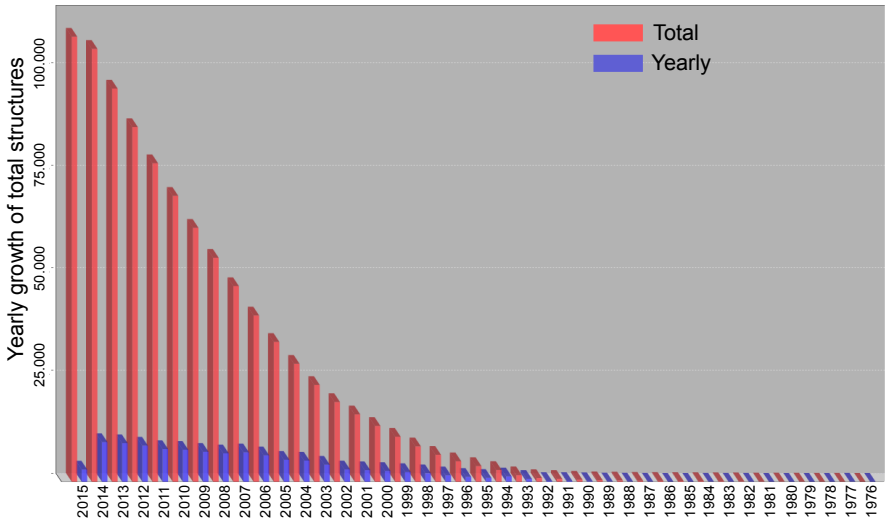


Figure 1.10. Graph displaying the number of searchable structures per year in the PDB databank (http://www.rcsb.org/pdb/home/home.do).

In addition to this structural knowledge, the way of visualizing 3D molecules has also undergone a complete revolution from the "ball and spokes" models (Kendrew), or the "wire models" (Rubin and Richardson [77]), to the computer-aided visualization programs that we use today: RasMol (http://www.umass.edu/microbio/rasmol/), Chime and Protein Explorer [78] are some of the first Molecular Visualization Freewares, as well as JMol (http://jmol.sourceforge.net/), VMD (http://www.ks.uiuc.edu/Research/vmd/) or PyMOL (https://www.pymol.org). Some of these programs allow us not only to visualize protein structures, but also to edit them and model-build structures of closely related proteins. Taking advantage of these tools we can generate graphical models in order to understand protein function mechanisms. E.g. the atomic description of a synaptic vesicle, the first model of a full organelle by Takamori et al (Figure 1.11) [79].

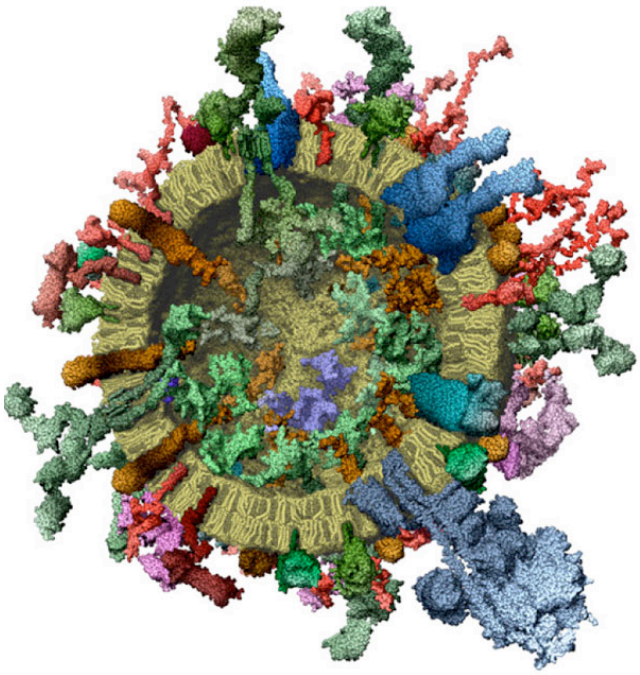


Figure 1.11. Cross-section of the molecular model of an average synaptic vesicle.

However, static molecular models limit our understanding of protein behaviour. To this aim we have to resort to other computational methodologies.

In the following section a brief overview of the various computational methods applied in this doctoral thesis for studying the behaviour of the proteins with our interactive graphics models is presented. For a more detailed discussion on the available methodologies the reader should refer to the corresponding references.

1.2.2 FORCE FIELD

A force field is the functional form and parameter sets used to calculate the potential energy of a system of atoms in molecular mechanics and molecular dynamics simulations. Parameters can be obtained from either experimental data (X-ray, NMR and IR spectroscopy) or *ab initio* and semi-empirical quantum mechanical calculations. In the 1970s, Norman L. Allinger developed the MM2 force field [80] primarily for conformational analysis of hydrocarbons and other small organic molecules. It was designed to reproduce the equilibrium covalent geometry of molecules as precisely as possible, [81].

Several force fields have been developed since the 70s [82]–[84] and, in general, the basic functional form of potential energy includes bonded terms for interactions of atoms that are linked by covalent bonds (*stretching*, *angles* and *dihedrals*), and nonbonded (also called "noncovalent") terms that describe the long-range electrostatic and van der Waals forces.

1.2.3 QUANTUM MECHANICS (QM)

Chemical processes in molecules are of quantum mechanical nature, involving transitions of electrons. Therefore, in theoretical studies of those processes molecules are necessarily considered as collections of nuclei and electrons [85].

All QM calculations are based on finding approximate solutions to the Schrödinger Equation:

$$\hat{H}\,\Psi = E\,\Psi \tag{2.3}$$

We know that quantum mechanics ‘‘works’’ in the sense that if we want to calculate just about anything concerning molecular structure we can do so, in principle, to any desired degree of accuracy. However, the computational resources required to obtain sufficiently exact solutions on a conventional computer generally increase exponentially with the number of atoms involved. This rendered such calculations intractable for all but the smallest of systems until recently [86]. Even with today’s methods and computational resources, simulations of large biomolecules are still outside the scope of QM methods.

1.2.4 MOLECULAR MECHANICS (MM)

Another way of studying molecular structure and dynamics is applying classical mechanics, commonly called molecular mechanics (MM). This method is purely empiric and is based on classical mechanics or Newton mechanics in order to predict molecular structures and/or molecular properties. For such a classical treatment of (bio)chemical systems, the key components of atoms, electrons and nuclei, are decoupled, assuming that electrons are distributed optimally around the nuclei, and attention is focussed on the location of the nuclei. Thus, a molecule is represented as an ensemble of spherical masses (atoms) bonded by harmonic or elastic forces.

The fundamental advantages of this particular theoretical level used in MM, with respect to QM, is the considerably reduced complexity of the biological system, allowing the study of systems with thousands of atoms such as proteins and nucleic acids.

Using the force field potential and energy minimization algorithms such as conjugate gradients [87] or steepest descent [88] it is possible to optimize the geometry of a molecular system to be compatible with the parameters of the force field reaching the closest minimum on the potential energy surface. This process is known as ***energy minimization***.

1.2.5 HYBRID QM/MM methods

The general idea of a hybrid quantum mechanics/molecular mechanics (QM/MM) method is to divide a large chemical system into at least two regions, a reacting part and a surrounding part. The atoms of the reacting part are treated at a quantum level, while in the outer region a less computationally demanding MM method is applied. The total energy of the system is calculated according to the general formula (equation 2.4)

$$E_{TOT}\,(TOT) = E_{QM}\,(QM) + E_{MM}\,(MM) + E_{QM/MM}\,(QM/MM), \qquad (2.4)$$

where the indices refer to the computational method and the labels in parentheses to the region where it is applied. $E_{QM/MM}$ (QM/MM) describes the interaction between the two

regions and can be separated into its QM component, E_{QM} (QM/MM), and its MM component, E_{MM} (QM/MM).

1.2.6 MOLECULAR DYNAMICS (MD)

An extraordinary revolution regarding computational power and methodologies has occurred in the last 60 years.

At the end of the 50s, Alder and Wainwright [89] introduced a new method to study interactions among rigid spheres, termed Molecular Dynamics (MD).

MD simulations are able to predict the projection of the system of interest into the future starting from an initial spatial configuration. New states are generated by applying classical laws from molecular mechanics, in particular Newton's equations of motion (equation 2.5)

$$\frac{d^2x_i}{dt^2} = \frac{F_i}{m_i} \quad , \tag{2.5}$$

where the motion of particle i with mass m_i is described along the coordinate x_i, being F_i the force applied and t the time.

In 1964 the first simulation was performed using liquid argon by Rahman [90], and ten years later, in 1974, the first MD simulation of a biologically relevant system, liquid water [91]. In 1977 the first protein simulation was published, using bovine pancreatic trypsin inhibitor [92]. Although the simulation lasted less than 10 ps and was *in vacuo*, it was a huge step to stop watching proteins as rigid structures and transforming the view to the dynamic entities that we all now think of. Lately, custom-made massively parallel supercomputers have been built, e.g. ANTON [93], [94], that can extend the length of this type of simulations into the millisecond range and therefore allow the study of process such as protein folding, ligand binding and behaviour of large macromolecular assemblies. This time scale, however, is still far from that in which many biological process take place (Figure 1.12) [95].

MD is one of the most powerful, useful and important computational tools to study macromolecules, since it allows to address dynamic processes, such as allosteric effects, activation mechanisms or drug binding.

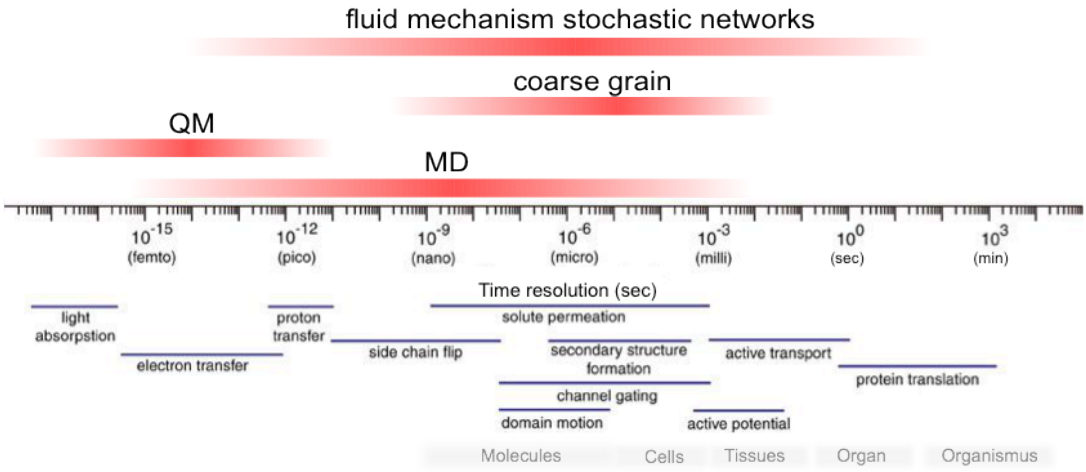


Figure 1.12. Amplitude and time scales of various computational techniques and the corresponding biological processes.

This deterministic method calculates the behaviour of a given molecular system along the time axis generating a sequence of snapshots as a function time (the trajectory); these points in space belong to an ensemble, and they correspond to the different conformations of the system and their respective momenta. Once the positions and the velocities of each atom are known, the state of the system can be predicted at any time in the future (or, less often, in the past).

The problems that can be investigated with MD simulations range from: structural insights, detailed information about the dynamic behaviour such as conformational changes or fluctuations, stability, folding, protein-protein interactions, protein-ligand interaction, membrane inclusion, ion transport, etc.

Several conditions and parameters must be taken into account when an MD simulation is being set up: initial *velocities*, fixed number of *particles*, *volume*, *temperature* of the simulation [96], [97], *pressure*, use of *periodic boundary conditions* (PBC), *box size,* etc. In the end, we will perform the steps summarized in Figure 1.13.

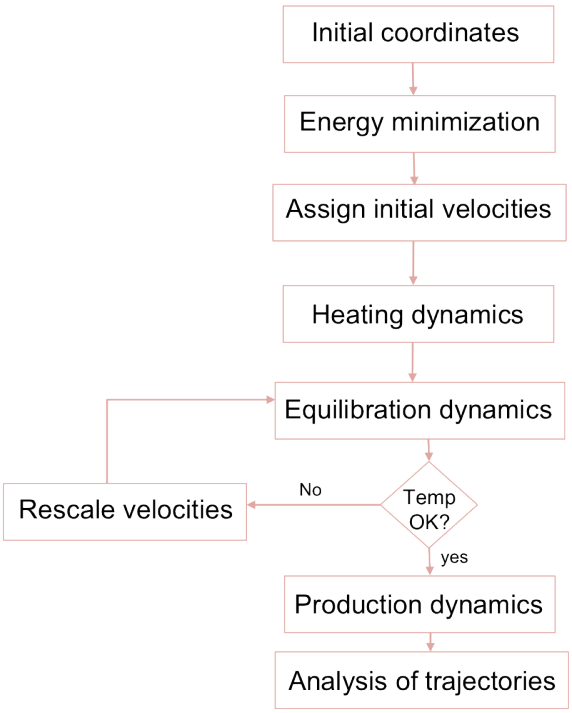


Figure 1.13. Scheme depicting how to set up and run an MD simulation: first we need the initial 3D coordinates, then we minimize the structure to avoid artifacts emanating from steric clashes. After this, initial velocities are assigned and the heating process starts, either with a temperature ramp or all at once (depending on the program used). Then different steps of equilibration can be performed. After the system is equilibrated, the dynamics production can start and the trajectories generated are ready to be analyzed.

One of the main shortcomings of MD are the large trajectory files that are produced, which have to be stored using gigabytes of disk space when large systems such as proteins are simulated for tens or hundreds of nanoseconds. Principal Component Analysis (PCA) is one of the most used techniques to reduce the amount of data and capture the essential movements of a macromolecular system.

1.2.6.1 Principal Component Analysis (PCA)

PCA is a variable reduction procedure appropriate once measurements on a number of observed variables have been obtained and one wishes to derive a smaller number of variables, called principal components (PC). We will focus on this variable reduction procedure [98] for analyzing trajectories generated by MD. Commonly used MD timesteps are on the order of 1 fs while the simulation time may range from few to hundreds or even thousands of nanoseconds. Then a single trajectory can easily contain a huge amount of data with large amounts of different motions and conformations obtained, being some of them more important than others from a biological relevance point of view (activation, inhibition…). In order to be able to correlate those motions and discard those not considered important, a PCA generally provides a stable and statistically reliable definition of the essential subspaces [99].

Technically, a PC can be defined as a *linear combination of optimally weighted observed variables*. To construct this linear transformation, it is first necessary to construct the covariance matrix and diagonilize it. Then the eigenvectors and eigenvalues, with the information about the variables and the information represented by each PC (measure of the variance), respectively, are extracted and projected. Therefore, the amount of information captured by each PC decreases by number: the first PC represents more information than the second one and so on.

Another limitation from MD simulations is the calculation of thermodynamic properties, i.e. free energies. Although we can never calculate absolute free energies (since we do not have an appropriate reference state), relative free energies can be estimated using several computational techniques. However, it is very difficult for MD to obtain these parameters due to the insufficient sampling in a finite length and time scale simulation. In order to estimate free energies, alternative methods must be applied.

1.2.6.2 Free Energy Calculations

Extracting absolute free energies from MD simulations involving thousands of atoms is challenging since it requires the calculation of Z_n (partition function), which is typically prohibitive in a high dimensional phase space. However, for most practical applications it suffices to determine the relative free energy between two states, A and B. For example, as we will see in the results section, the free energy difference between one protein kinase conformation in its active state, A, and in its inactive state, B.

In transitions between states that occur infinitely slowly, the work W performed by the system is exactly the free energy difference ($W = \Delta F$). Since the free energy is a state function, several methods use conveniently chosen reaction coordinates to drive the system from state A to state B. By performing this transition infinitely slowly and recording either the mean force or the probability distribution, ΔF can be calculated.
Potential of Mean Force (PMF), $W(\xi)$, along some coordinate ξ, first introduced by Kirwood in 1935 [100], is a key concept in statistical mechanics [101]. PMF calculations are also used to determine free energy differences between two states that are conformationally different.

The methods used in this thesis for free energy calculations are the following:

- Umbrella Sampling (US): US is a widely used approach to deal with rare events, such as an induced conformational transition, that forces the system to sample all the positions along a previously defined reaction coordinate, ξ, which is discretized in a number of bins with reference positions at ξ_i (Figure 1.14). The US method, originally proposed by Torrie and Valleau [102], uses an external bias potential, w_i, to confine the molecule around a preestablished position, ξ_i. The relative probabilities within the sampled regions are collected, and a correction is included to account for the external bias prior to obtaining $W(\xi)$. A common form of the biased potential w_i is a harmonic potential of strength k (equation 2.6) which restrains the molecule of interest in the vicinity of ξ_i.

$$w_i(\xi) = \frac{1}{2}k(\xi - \xi_i)^2 \tag{2.6}$$

The choice of k, the strength of the bias, is the most critical decision [103]. The value of k has to be large enough to drive the system over the barrier, but not too large because it would give rise to very narrow distributions and the configurations with high energies would be overrepresented.

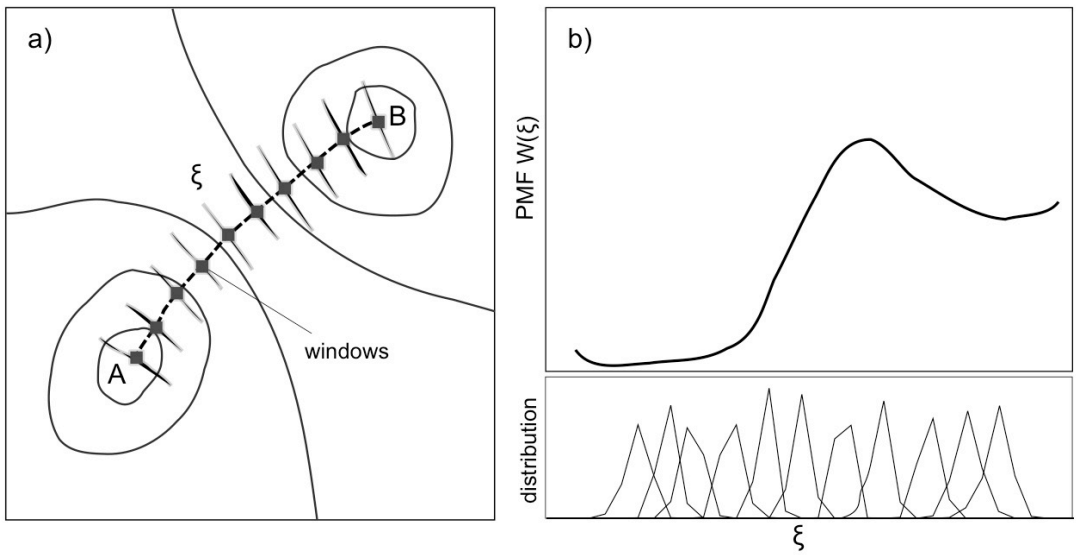


Figure 1.14. a) Representation of the reaction coordinate (dashes line) between two states, A and B, represented by two minima on the potential energy surface. The system is sampled along the reaction coordinate ξ in each window. b) At the top, the PMF is represented (thick solid curve), and at the bottom, the biased distributions obtained from the simulation are shown (thin solid curves).

After the simulations, the free energy curves are combined with either the weighted histogram analysis method (WHAM) [104], [105] or by means of umbrella integration [106].

- Metadynamics (metaD): this method, introduced for the first time by Laio and Parrinello in 2002 [107], is based on a dimensional reduction that allows the system to escape from the free energy minima and to reconstruct the free energy surface. To achieve this aim, metaD requires the preliminary identification of a set of collective variables (CVs) which are assumed to be able to describe the process of interest [108]. The dynamics in the space of the chosen CVs is enhanced by a history-dependent potential constructed as a sum of Gaussians centered along the trajectory followed by the CVs (Figure 1.15). Then, the sum of Gaussians is exploited to reconstruct the free energy surface.

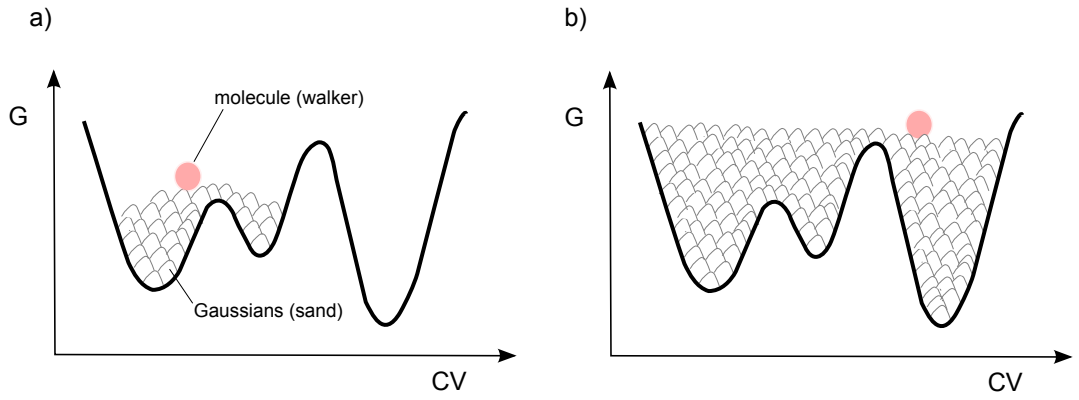


Figure 1.15. Schematic representation of the metadynamics technique. A particle (or a walker in the example, pink dot) explores the unknown free energy surface (pool), with Gaussians help (sand). a) At the beginning of the search, b) when all the surface has been explored.

The idea of this method can be informally described by a simple example [108]: imagine that a person who is walking during the night falls in an empty swimming pool, without any light that can show something. After a while, the walker realizes that the walls are too steep for climbing and because of the complete darkness, it is very difficult to find the shallowest pocket to try to jump and escape. In metaD, there would be a large source of sand that the walker can use it to deposit below him. Although at the beginning he will explore more the deepest part of the pool, after a while the sand will be filling the whole pool and he will be able to explore higher regions. Now the key

part: if the walker is capable to remember every position he placed the sand (Gaussians) in, he will be able to reconstruct the negative image of the whole surface of the pool (free energy surface).

Although several CVs can be used to describe the transitions of interest, around two or three, or even more, the method only works if all the relevant variables are included, and sometimes this is many times more than two or three. In these cases we cannot apply metaD. However, some extensions of the methodology have been developed to solve this limitation.

(i) Multiple walkers: in order to speed up the free energy calculations multiple replicas of the system can be run simultaneously [109].

(ii) Parallel tempering metadynamics (PTmetaD): as in the previous extension, it is based on running multiple replicas of metaD but at different temperatures. It allows us to explore low probability regions and overcoming large barriers along a multidimensional space at different temperatures [110].

(iii) Bias exchange (BE): another technique based on the combined use of replica exchange and metaD, it consists of running in parallel, at the same temperature, a larger number of MD simulations, biasing each replica with a metaD potential acting only on one or two CVs.

(iv) Well-tempered metadynamics (WTE metaD): this variant of metaD uses the potential energy as a CV. The WTE can be designed so as to have approximately the same average energy as the canonical ensemble but much larger fluctuations [111]. A biased ensemble that enhance sampling speed, especially when it is combined with parallel tempering, observing transitions between states that otherwise would be very difficult to study.

- Special cases:

(i) "Steered", "biased", or "activated" molecular dynamics: the addition of external forces reduces the energy barriers and increases the probability of unlikely configurations in a regular simulation. This procedure is suitable for studying processes that are intrinsically fast but constitute rare events (average frequency $<< 10^{11}$ s^{-1}) because they are limited by one or more energy barriers. As an example, we have used them to promote product release from one of our kinases of interest, namely GSK3β.

(ii) Targeted molecular dynamics: an additional term is added to the potential energy function based on the mass-weighted root-mean-square deviation (rmsd) of a set

of atoms in the current structure compared to a reference structure. At each MD step, the algorithm performs a best-fit of the reference structure to the simulation structure and calculates the rmsd for the selected atoms. This method can be optimal for simulating large conformational transitions that happen on a timescale of milliseconds, as in the case of the Abl kinase.

1.2.8 NORMAL MODE ANALYSIS

Normal mode analysis (NMA) is one of the simplest and most inexpensive simulation techniques to probe the large-scale, shape-changing motions in biological molecules [112], [113]. NMA is a harmonic analysis and it is as accurate as the force fields that are used in MD simulations. One of its main applications has been to predict functional motions in proteins or other biological molecules. Functional motions are those related to function and it is always assumed that the normal modes with the largest fluctuation (lowest frequency modes) are the ones that are functionally relevant. This is so, because like function, they exist by evolutionary design rather than by chance [114].

A standard NMA requires a set of coordinates, a force field, and software to perform the required calculations. Although it is included in packages like GROMACS, there are also easier and interactive ways for any user (e.g. NOMAD: http://lorentz.immstr.pasteur.fr/nomad-ref.php; elNémo: http://www.sciences.univ-nantes.fr/elnemo/; FlexServ: http://mmb.irbbarcelona.org/FlexServ/; iMod: http://chaconlab.org/methods/multiscale-simulations/imod). The performance of an NMA in Cartesian coordinate space requires three main calculation steps: (i) minimization of the conformational potential energy as a function of the atomic Cartesian coordinates; (ii) the calculation of the so-called “Hessian” matrix, which is the matrix of second derivatives of the potential energy with respect to the mass-weighted atomic coordinates; and (iii) the diagonalization of the Hessian matrix. The eigenvectors obtained from this matrix are the “normal modes”, and the eigenvalues are the square of the associated frequencies. The first six modes obtained correspond to translations and rotations, hence it is only from the seventh mode onwards that we can extract the interesting motion information [115], [116].

Furthermore, using NM analysis combined with coupling analysis with the programme ProDy [117], which allows to analyse dynamical properties of individual residues, we can indicate regions allosterically coupled. This method is commonly known as Allosteric Coupling Analysis (ACA).

1.3 Ligand – Receptor Complexes

1.3.1 LIGAND BINDING

The interaction between molecules (proteins, nucleic acids, small molecules or ligands…) constitutes the basic language of biological systems. Many diseases are related to the malfunctioning of these interactions [118], and therefore it is crucial to know how they occur. This knowledge would allow us to interfere and manipulate them through the search and design of new ligands to decrease, or even abolish, the effects associated with a given illness provided we undertand it well in molecular terms.

In the design of new ligands two main considerations should be taken into account: the binding *affinity* and the *selectivity*.

Since the end of the nineteen-century, different models representing ligand binding have been proposed (Figure 1.16). In 1894 the first model was postulated by Emil Fisher: the famous "lock and key" model. In this model only the correct key (the substrate with the specific shape that is complementary to the active site) can fit in its lock on the enzyme [119]. This is a very rigid approximation and it does not consider the mutual adaptation between both ligand and receptor. The next working model was proposed by Daniel Koshland in 1958 [120] and is known as the induced-fit model. It considers flexibility as an intrinsic property of the target, and the target rearranges to accommodate the ligand. The conformational selection model was proposed in 1999 [121] and postulates that the ligand selects, among a conformational ensemble, the most appropriate conformer for binding. This latter model was proposed in the 1960s for allosteric systems by Monod, Wyman and Changeux [22].

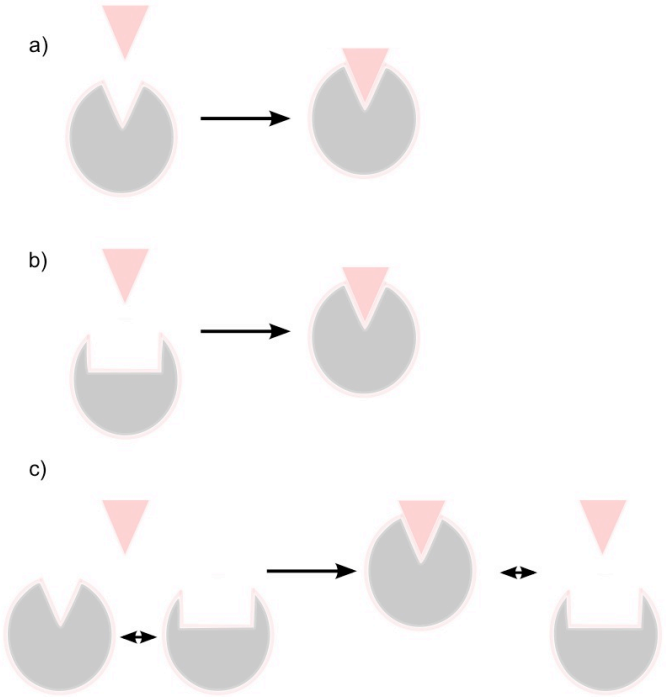


Figure 1.16. Models of molecular recognition where the ligand is shown in pink and the target in grey. a) Lock and key. b) Induced fit. c) Conformational selection. The unbound state exists as an equilibrium between different conformations. The ligand selectively binds the matching target while the unbound ligands re-establish the equilibrium.

Although there are some pieces of evidence showing that conformational selection is the most appropriate model describing many biological systems behaviour [122], a recent review suggests a combination between both mechanisms, conformational selection and induced fit, in order to better describe several systems [123].

1.3.2 LIGAND BINDING ENERGY

"The general aim… has been to determine the extent to which the effects produced by drugs on cells can be interpreted as processes following known laws of physical chemistry" (A.J. Clark). In 1937, Alfred J. Clark was the first to apply mathematical approaches used in enzyme kinetics [124] to the study of drug action (reaction 3.1).

$$\text{Ligand} + \text{Receptor} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Ligand-Receptor} \qquad (3.1)$$

The main magnitude that determines the ligand-receptor binding complex is the free energy of binding (ΔG_{bin}, equation. 3.2).

$$\Delta G_{binding} = -RT \ln(K_d) \tag{3.2}$$

where R is the gas constant, T is the temperature and K_d is the dissociation constant. This constant, as we saw above, measures the propensity of a complex to break down into its components at equilibrium (equation 3.3).

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[\text{Ligand}][\text{Receptor}]}{[\text{Ligand}-\text{Receptor}]} \tag{3.3}$$

Taking into account the last expressions, the more negative the ΔG_{bin} the higher the shift of the equilibrium shown in 3.1 towards complex formation and therefore the higher the ligand's affinity for the receptor.

1.3.2.1 Enthalpic and entropic contribution.

At the same time, the binding free energy depends on both the binding enthalpy (ΔH) and the binding entropy (ΔS) (equation 3.4). The extremely high affinities of some ligands require that both terms contribute favourably to complex formation.

$$\Delta G_{binding} = \Delta H_{binding} - T\Delta S_{binding} \tag{3.4}$$

ΔH is much more difficult to calculate than ΔS. ΔH is mainly contributed to by specific interactions between the ligand and the target and is defined by different types of forces, such as van der Waals, electrostatic and hydrogen bond interactions. A favourable enthalpy will mean that the interactions established between the ligand and the target are enough to compensate for the unfavourable enthalpy associated with desolvation. But because both forces, have an opposite effect, it is very difficult to compute the enthalpy. On the other hand, the main contributors to the entropy are the desolvation, which is favourable, and the change in rigidity, which is usually unfavourable because when the ligand binds into a cavity both target and ligand lose degrees of freedom. The enthalpic contribution is easier to compute as we can constrain the ligand poses in the free form and within the complex.

In order to design potential ligands, it is essential to understand the type of interactions that occur between ligand and receptor, as will be discussed below.

1.3.2.2 Hydrogen bonding. This is a highly directional interaction established between a hydrogen atom attached to an electronegative atom, so-called hydrogen bond donor, and another electronegative atom, the so-called hydrogen bond acceptor. The strength of the bond depends on how close the three atoms involved are to the optimal distances and angles for hydrogen bonds [125]. This bond formation is essential for understanding the energetic basis of protein folding and for obtaining structure-based predictions of the energies of ligand binding for the purpose of drug design [125].

1.3.2.3 Solvent effects. All relevant interactions in biology take place in an aqueous environment. When molecules are isolated in solution, they are completely surrounded by water molecules. However, when the process of binding occurs, many of those water molecules are displaced. Such a displacement bears an enthalpic cost that should be counterbalanced by the new interactions being formed. In addition, a gain in entropy from the released solvent molecules also takes place. From a computational perspective there are two extreme models that are used to account for solvent effects: (i) explicit models, where each solvent molecule is represented in atomic detail, and (ii) implicit models, where a mathematical function is built to mimic the behaviour of the bulk solvent. It is also possible to consider mixed models in which some of the relevant solvent molecules are treated explicitly whereas the rest are considered implicitly.

1.3.2.4 The hydrophobic "interaction": some amino acids have hydrophobic side chains (leucine, valine, proline…) and many ligands also possess moieties that repel water. If two hydrophobic moieties get into contact, the surrounding water molecules are released and this interaction gives rise to a "hydrophobic effect" that makes a positive contribution to the overall stabilization energy.

1.3.2.5 The entropic contribution: The concept of entropy is always related to the idea of order. An isolated molecule is free to translate, rotate and vibrate. When an intermolecular complex is formed, some of these motions are lost. As a consequence, more order is introduced into the system and this entails a decrease in its

entropy. Apart from the solvent entropy, the solute entropy is usually divided into conformational and vibrational parts. The conformational part has to do with the reduction in the number of energy troughs that the ligand and the target can visit once the binding has occurred whereas the vibrational part refers to motions within a single trough. Estimating entropy is a daunting task and the associated error bars are usually quite large [126].

1.3.2.6 Importance of solvation/desolvation in the binding event:

Several detailed binding free energy studies have suggested that differences in solvation may play an important role in differences in binding free energy between relatively similar compounds [127], [128], [129]. Two molecules might have similar interactions with a protein, similar strain energies, etc., but have different solvation properties in water, leading to solvation-driven differences in binding free energies. The importance of solvation and desolvation is supported by an emerging trend toward including approximate estimates of solvation/desolvation energies in approximate docking methods for scoring protein-ligand binding. Including such estimates appears to result in improved scoring [130], [131]. Without these contributions, charged ligands can wrongly appear to bind better than polar ligands in a polar binding site. A charged ligand may make favourable electrostatic interactions in a polar binding site, but it also costs a huge amount of energy to remove it from water [127], [131].

1.3.2.7 Other contributions:

These include formation of covalent bonds, non conventional hydrogen bond (-CH··· acceptor) and interactions with metals or halogen atoms. These are also important interactions that are usually implicitly included in the previously discussed terms but that occasionally need a special treatment for correctly predicting ligand binding.

1.3.3 PREDICTION OF LIGAND BINDING MODES

Drug discovery, development and marketing are clearly crucial milestones in an ambitious task that requires tremendous efforts in research, development and money investment (on average, 15 years and around >1000 million dollars per approved

molecule [132]). In the last 30 years the use of combinatorial chemistry [133] and high-throughput screening (HTS) [134] has greatly advanced drug discovery efforts, However, the ratio between the number of new drugs and the resources invested is well below the initial expectations. Computational techniques, on the other hand, attempt to rationalize and accelerate the initial steps in the drug design cycle.

Computer-aided modelling and simulation methods [135], if correctly applied, allow the medicinal chemist to pinpoint the more promising candidates (*hits*) out of pools of thousands or even millions of molecules (*chemical libraries*). This reduced set can then be subjected to experimental analysis and any promising compound can be subsequently optimized to attain the desired pharmacokinetic and pharmacodynamic profile in order to become a *lead*.

There are many tools available to perform these computational tasks [135] . When the structure of the target protein is known, receptor-based computational methods can be employed. We can find different scenarios depending on which data are available. The most favourable scenario is when we have the crystallographic structure of the target in complex with one or several bound ligands; this information can be used to predict similar ligand binding modes. The second scenario is when we have the target receptor structure and know that a particular ligand binds to a particular site on the receptor, but we do not exactly what the binding mode is. Finally, another two less favourable receptor-based scenarios exist, one when we do not know which ligand could bind but we still know the binding site, and another when we ignore both the ligand and the binding site.

Depending on the scenario different methodologies described below could be applied. But independently of the scenario, it is crucial to previously characterize the pocket, when we know it, or to search for and characterize all possible binding sites if we are to predict the correct binding mode.

1.3.3.1 Binding site characterization: Taking advantage of several computational tools (e.g., cGRILL, GRID, PP_SITE from LigBuilder, DogSiteServer) [136], makes it possible to predict which are the key residues involved in the interactions as well as the different chemical properties of the pocket and its

druggability when the binding site is already known. These methodologies allow the prediction of favourable energies between different parts of the molecule and the target.
favourable energies between different parts of the molecule and the target. In this thesis, we use the PyMOL plugin cGRILL, based on the classical GRID program [137]. This algorithm calculates different affinity maps, or molecular interaction fields (MIF), using probes, that are represented as three dimensional grids so as to display the different regions where the probe establishes favourable interactions with the target. These probes are lipophilic (CH_3), hydrogen bond acceptor (=O), hydrogen bond donor (NH_4^+), mixed hydrogen bond donor-acceptor (OH), and hydrophobic-like (hydrophobic). The cGRILL program also implements a way to coalesce these maps into interaction hotspots that provide accurate locations for specific functional groups on the putative ligand (Figure 1.17).

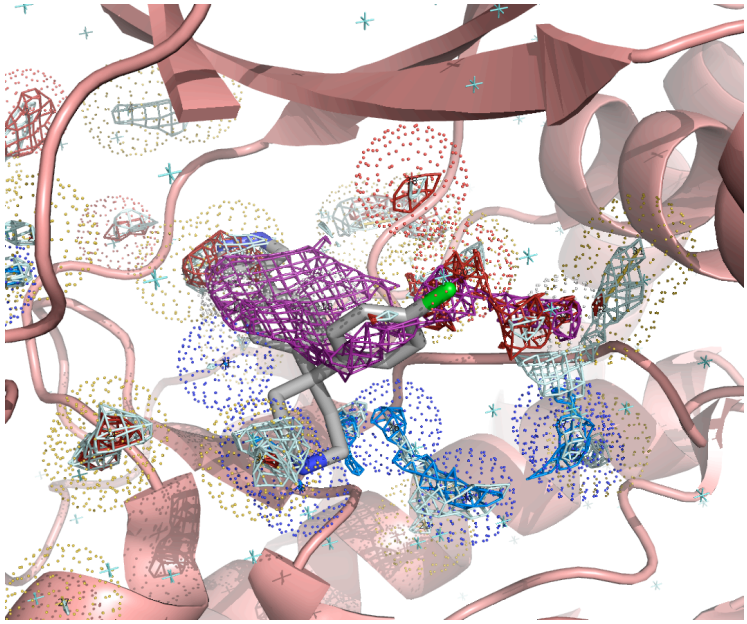


Figure 1.17. Binding site of PKA (in light pink cartoon representation) analyzed with cGRILL. The co-crystallized inhibitor is shown as sticks with C atoms coloured in grey. The different affinity maps are shown as grids of different colours (purple, red, dark and light blue, representing hydrophobic, hydrogen bond acceptor, hydrogen bond donor and hydrophilic affinity maps, respectively). The hotspots are displayed as dotted spheres.

Further, there are several on-line tools also available to identify potential binding pockets. For example, FTMap [138], [139] (http://ftmap.bu.edu/param) docks a panel of solvent molecules into the most suitable pockets and crevices of the target structure in

order to identify those that could represent high-affinity binding sites for drug molecules (Figure 1.18-a). Other programs like Firestar [140] (http://firedb.bioinfo.cnio.es/Php/FireStar.php) or Relibase [141] (http://relibase.ccdc.cam.ac.uk) also perform the same type of predictions. In addition, the SiteMap program [142] has been developed to identify binding pockets and also to predict their druggability (whether the chance to find a ligand for that protein is high or low) (Figure 1.18-b). In essence, this software considers several descriptors (size, degree of enclosure/exposure, tightness, van der Waals forces, hydrophobic/hydrophilic character, and hydrogen-bonding possibilities) to obtain a drugablility score (DScore) for each identified pocket. DScore can discriminate pockets with the high potency and druggability from pockets that just have high potency (e.g. if a protein catalytic site is highly potent with nM compounds exists, but the binders are highly charged and not druglike; DScore ranks down such sites).

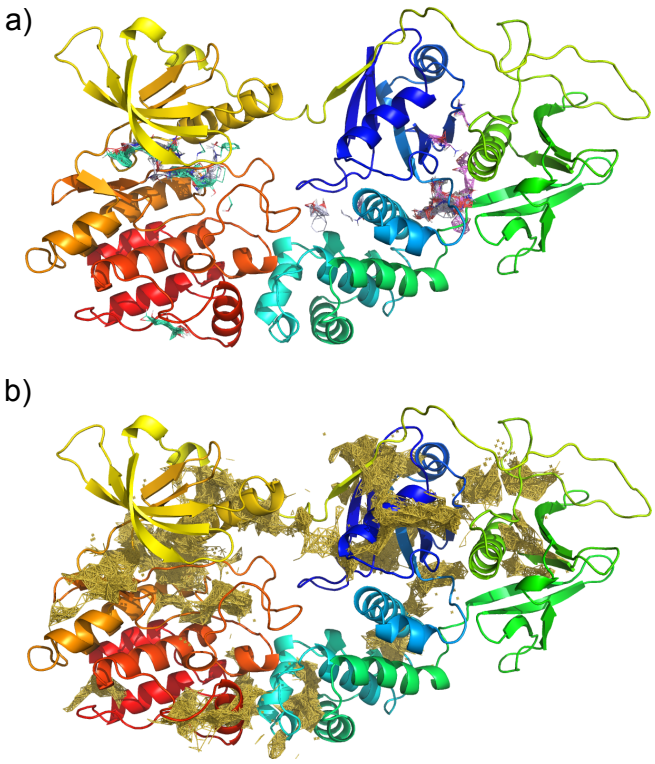


Figure 1.18. Potential binding sites on FAK shown as brown sticks and calculated with programs a) FTMAP and b) SiteMap. The former using a crystallographic structure, the latter using a long trajectory from a MD simulation.

1.3.3.2 Ligand docking:

Once we have defined the binding pocket, we can proceed with the docking process [135]. The docking process consists of exploring exhaustively and automatically every possible binding mode of a ligand onto a receptor. This search can be performed by different algorithms that will (i) characterize the binding site, (ii) place the ligand inside the pocket, and (iii) evaluate the goodness of the fit for each ligand pose. For the latter a *scoring function* that represents the thermodynamics of interaction of the protein–ligand system adequately as to distinguish the true binding modes from all the others explored, and to rank them accordingly, is required. To perform ligand docking there are two main problems that we need to solve:

(i) Explore the conformational space of the ligands that we intend to bind to the target receptors.

(ii) Score all the poses generated to rank them according to their estimated binding affinities.

***- Searching engines*:** The success of a docking algorithm in predicting a ligand binding pose is normally measured in terms of the root-mean-square deviation (RMSD) between the experimentally observed heavy-atom positions of the ligands and the one(s) predicted by the algorithm. The flexibility of the system is a major challenge in the search for the correct pose and there are different levels of approximation:

(i) Rigid body docking, when both the target and ligand are considered as static components, reducing the complexity of the system to just rotations and translation. The advantage of this option is the speed of the calculation; the major disadvantage is that it is too simplistic to achieve a high success rate. (ii) Rigid protein – flexible ligand docking, which is a good approach if it is assumed that the lock-and-key model is reasonable. The majority of the programs use this approximation. (E.g. DOCK [143], Glide, CRDOCK [144], GOLD [145], [146], Autodock [147], [148] …). (iii) Fully flexible docking, when the flexibilities of both the ligand and the protein are considered. Many of the programs mentioned before, such as GOLD, Glide or Autodock, have included different degrees of freedom to the protein.

In this work, we have made extensive use of Autodock and Glide following the second

approach: rigid protein – flexible ligand.

For the flexible-ligand search docking there are three general categories of algorithms: systematic methods; random or stochastic methods; and simulation methods. The former, systematic search, tries to explore all the degrees of freedom in a molecule, and can be further divided in three main types: (i) conformational search methods, when all rotatable bonds in the ligand are systematically rotated through 360° using a fixed increment, until all possible combinations have been generated and evaluated. The major disadvantage here is that the number of structures generated increases immensely with the number of rotatable bonds, a phenomenon known as the combinatorial explosion. Hence, the application of this type of methods is very limited. (ii) The fragmentation methods incrementally grow the ligands into the active site, either by docking the several fragments into the active-site and linking them covalently to recreate the initial ligand ("the place-and-join approach"), or by dividing the ligand into a rigid core-fragment that is docked in first place and flexible regions that are subsequently and successively added ("the incremental approach"). And the latter, the database method, arises to overcome the combinatorial explosion problem by using libraries of pregenerated conformations (conformational ensembles) to deal with the ligand flexibility issue.

The second type of algorithms, random search algorithms, sample the conformational space by performing random changes to a single ligand or a population of ligands. When all ligand conformations have been generated each one is compared with the earlier conformations, and rejected or accepted on the basis of a predefined probability function. Different algorithms have been developed to perform this searching process: Monte Carlo (MC) methods [149] based on Boltzmann probability function; Genetic Algorithm (GA) [150] applying ideas derived from genetics and the theory of biological evolution to docking and Tabu search methods, by imposing restrictions that prevent the search from revisiting already explored areas of the conformational space, promoting the analysis of new regions

The last category of algorithms, simulation methods, employs a rather different approach to the docking problem, and is based on the calculation of the solutions to Newton's equations of motion.

Other possibility to overcome the first docking problem is to discretize the search space by defining a three dimensional grid in a manner similar to Goodford's GRID algorithm [137]. These grids speed up the docking process several orders of magnitude. However, since all the grid values are pregenerated, there is no flexibility for the target, unless some parts of the target are sampled together with ligand, increasing the computational cost. After this, the ligand poses generation start. To be able to search for all the possible ligand conformations and orientations we need to consider all the structural combinations taking into account not only the inherent flexibility due to changes in torsional angles but also the six degrees of rotational and translational freedom.

Our common protocol to perform docking follow the last approach where: we defined a cubic box of 50 x 50 x 50 Å, applying a space grid of 0.375 Å. Subsequently, 100 cycles of Lamarckian genetic algorithm [151] are usually run and the best scored poses are selected attending to the lowest energy and most populated clusters.

Although usually the receptor is considered rigid, there are programs like Glide [152], [153] or FLEXX [154], [155], that already include the ligand-receptor docking with certain flexibility in the binding site. In addition, more than one conformation of the target can be considered [156], extracted for example from a normal mode analysis or from MD simulation, as later discussed in the Results section.

***- Scoring function*:** The second problem in docking, namely the *scoring function*, is probably the most critical step; how to score all the different poses generated and rank them in accordance to the estimated binding energy and how to choose the scoring function that will allow us to obtain the correct binding pose. Scoring functions are mathematical functions than estimate the binding affinity rather than calculate it. The form of a typical scoring function is (equation 3.5):

$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hyd\ bind.} + \Delta G_{solv.eff} + \Delta G_{tors} \quad (3.5)$$

Hydrophobic effects, van der Waals interactions, hydrogen bonding, steric and electrostatic interactions, and desolvation effects are among the factors that contribute to ligand bindi

ng evaluation (see section 1.3.2). In theory, a lower Gibbs energy (equation 3.5) of a complex indicates that the protein-ligand complex is more stable.

The scoring functions normally employed in protein– ligand docking are generally able to predict binding free energies and can be divided into three major classes [157]: force field-based, empirical, and knowledge-based scoring functions. The force field based scoring function generally quantifies the sum of two energies: the interaction energy between the receptor and the ligand, and the internal energy of the ligand. The empirical scoring functions are designed to reproduce experimental data and are based on the idea that binding energies can be approximated by a sum of several individual uncorrelated terms. Finally, knowledge-based scoring functions focus on following the rules and general principles statistically derived that aim to reproduce experimentally determined structures, instead of binding energies, trying to implicitly capture binding effects that are difficult to model explicitly.

Whichever docking program and scoring function are chosen by the user, a further validation is need. Apart from experimental validation (NMR, SPR, crystallization, etc.) of the top-scoring compounds, good alternatives prior to performing experiments are MD studies. It has been shown that MD simulations can provide validation of docking results as it takes into account the "motion" of the complex, not only the binding affinity and can determine whether the docked molecule is stable in the complex along time and not just within a "static picture" [158].

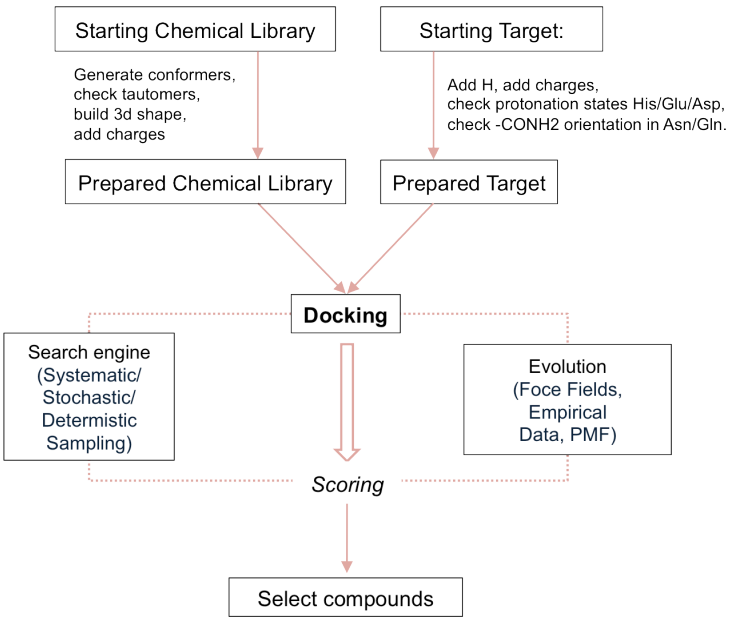


Figure 1.19. Scheme of the virtual screening process.

1.3.3.3 Virtual screening:

In addition to single ligand docking, we can automatically perform molecular docking of a large number of ligands (a virtual library) into the binding site of the target (i.e high-throughput docking, known as Virtual Screening or VS). VS consists of several steps (Figure 1.19): first, the preparation of the chemical library and the target. The former involves generating all the conformers, tautomers, 3D structures and atom point charges for the ligands. Regarding target preparation, once we have the 3D structure (either from a X-ray structure or from a model) the preparation of the binding site is dependent on the docking program being used. Some methods require hydrogens and others do not. Special attention must be paid to the histidines tautomers [159], crystal water molecules, ions, cofactors in the selected cavity to enhance the success rate. Then the docking method is selected and it predicts all the possible binding modes (*poses*) for each database compound [135], [160], [161]. Then this information will be used for calculating the goodness of the fit using the *scoring function*.

- Molecule databases (a.k.a. chemical libraries)

One of the essential parts of VS is to choose and prepare the database of compounds to be screened [160]. Currently there are many online databases from which this information can be directly obtained (e.g. the ZINC database at http://zinc.docking.org). It is important to carefully select which are the libraries of interest (specific target, commercially available, drug-like compounds, structures associated with toxicity or poor pharmacokinetic properties, etc) as these previous filters can save considerable computational resources and time.

- Fragment libraries

In the 1990s, the leading techniques for increasing drug discovery productivity were high-throughput technologies such as combinatorial chemistry and HTS [162] usually involving around a million of drug-like compounds. Although thanks to these techniques many hits have been found and developed, there are several limitations that are now into the spotlight [163]. Among them, the major limitation is to improve and optimize the hit rate of a drug-size molecule. Usually, when a compound has been found as a hit and it progresses into optimization, it reduces the affinity, the drug-likeness and finally, the development capability of the compound. To address this huge problem there is a screening alternative: fragment-based lead discovery (FBLD) [162].

The idea of FBLD is to target a protein with small molecules of low molecular weight, typically below 250 Da, in contrast to HTS that uses large whole-molecule libraries. By using such small ligands, the probability of having good complementarity with the target protein is much higher. Therefore, there is an inverse relationship between the complexity of a molecule and the probability of complementarity with the protein [164]. In addition, fragment screening allows the sampling of a greater portion of "chemical space" as the number of possible molecules rises exponentially as the molecular weight increases [165]. Even weak binding interactions of fragments to the protein target can be detected by applying high compound concentrations when some biophysical methods, such as NMR, surface plasmon resonance (SPR) or X-ray crystallography, are used.

However, there are two main obstacles with this approach: (i) the identification of suitable fragments that bind to neighbouring sites and (ii) the optimization of these fragments either by merging, linking, or growing to develop drug-like compounds [166]. In this thesis, we try to overcome these two hurdles by offering a computational protocol solution, including VS and MD simulations, for identification of hit fragments extracted from large public fragment-like databases (ZINC). We assess the validity of this protocol not only on one single structure, but on several conformations of the same so as to implicitly introduce flexibility in the protein. This type of protein conformational ensemble for docking has been regarded as one of the "most promising routes for future progress" [167]. Once the interaction within the cavity is experimentally validated we have the unique opportunity to design potent and efficacious drug-like compounds.

1.3.3.4 Incremental construction vs. fragment linking:

Once we have the starting fragment located in binding site, three different ways of growing the fragment can be performed. The first one, based on the first fragment, we can dock additional fragments and then link them to construct one larger molecule (Figure 1.20-a-b-c) satisfying all the requisites for the binding pocket, and taking advantages of the computational tools, like cGRILL and hotspot to select the best candidates. The second one, by an incremental construction involving the already known fragment as our starting point [166] (Figure 1.20-a,d). And the third and last one, creating analogues as we have done in section 3.3.3.1 increasing affinity of current hits by different computational approaches (affinity maps and hotspots generation).

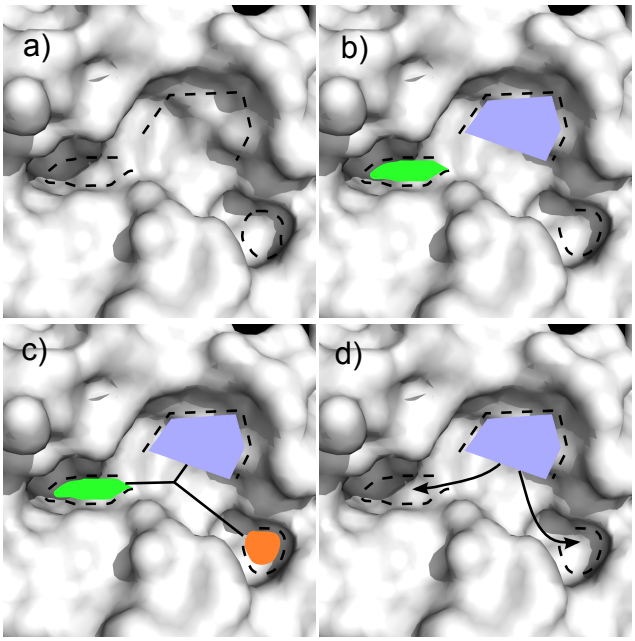


Figure 1.20. a-d. Ways of developing of a big molecule from a fragment hit [162].

1.3.3.5 Experimental validation:

To validate our virtual predictions we are going to use three different experimental methodologies:

- Surface Plasmon Resonance (SPR):

SPR is an optical technique utilized for detecting molecular interactions. The interaction experiment involves immobilizing binding partner on the sensor chip surface (the protein in our case) and injecting a series of concentrations of the ligands across the surface. Change in the index of refraction at the surface due to binding events are detected and recorded as response units (RU). Then, curves are generated from these RU and K_d values determined using the Biacore affinity analysis curve fitting algorithm.

To identify whether binding to the kinase domain occurs at a allosteric site or at the ATP binding site, we can use a competitive SPR assay. Compounds are tested in presence and absence of Cryzotinib, a known ATP site binding inhibitor, to determine whether there is an additive binding response, that would indicate non-competitive binding; i.e. allosteric binding.

- Co-crystallization vs. soaking:

Co-crystals are multi-component crystals where no covalent chemical modification of the constituents occurs as a result of the crystal formation. For crystallisation with compounds of limited supply, co-crystallisation is only sutiable in cases where crystals in the presence of the compound grow at the same contdition as apo-protein crystals, since in this case it is not necessary to perform compound consuming re-screening of co-crystallisation conditions. An alternative technique is soaking, where a ligand molecule is soaked into a pre-existing crystal. This technique is suitable if high-thoughput is required, e.g. for large number of compunds, or if crystallisation conditions would have to be rescreened for the complex. Co-crsytallisation is the method of choice if crystal latice contacts or small solvent channels prevent ligands to reach the binding site in a preexisting crystal. Other factors to consider are: the crystals may be fragile and suffer during compound soaking, the soaking time and inhibitor concentration need to be optimized or crystallisation conditions may need to be adapted to achieve effective ligand binding [168].

Louis, I think this is the beginning of a beautiful friendship.
Casablanca (1942)

Don't ask questions you don't want to know the answer to
Men in Black 3 (2012)

Chapter 2

2. Aims of this thesis

2.1 Motivation

This doctoral thesis will focus mainly on the analysis of protein kinases and phosphatases. These superfamilies' of proteins are an important subject of study for many reasons. The most important, is their implication in human diseases such as cancer. Kinases and phosphatases are involved in almost all essential signalling pathways in the cell, and their improper activation or dysregulation is, in majority of cases, the result of unchecked activity resulting in over-expression of proteins or downstream genes, and such a state of increased activity has often been positively correlated with many human diseases such as cancers, metabolic disorders, and infectious diseases.

Understanding the general concepts of kinase and phosphatase activation in general, and detailed mechanisms for each of them in particular, is one of the main goals of this thesis. Obtaining insight into the mechanisms by which kinases and phosphatases are activated will help us to better understand the weak points that we can be targeted to neutralize excessive activity in disease states.

I have studied these mechanisms using computational methodologies, including molecular modelling, molecular simulations, free energy calculations, and quantum mechanics in order to better understand the activation mechanism. Further I employed docking, ligand-receptor calculations and virtual screening in order to discover possible ligands as starting points for the development of highly specific cancer therapeutics.

Thus, one of our final aims will be identifying a strong and efficient approach to develop new cancer drugs, able to bind selectively and efficiently to specific targets in specific conformations. Although there has been an extraordinary effort from the scientific community to find new therapies against cancer, many challenges remain.

2.2 Specific Objectives

This thesis takes a computational approach to address the following five questions grouped together in three different topics:

Regarding the mechanism of phosphorylation and dephosphorylation reactions:

1. How do the GSK3β and PKA phosphorylation reaction occur?

   The objective is to study in depth, from the atomistic point of view, the phosphorylation reaction mechanism carried out by GSK3β. Similar studies are carried out on PKA, which has been extensively studied. We will use it as a template for our research, filling also some gaps in the current knowledge.

2. How is the SHIP2 phosphatase allosterically regulated by its C2 domain?

   The final goal is to understand the role of the C2 domain characteristic of SHIP phosphatases, and how it regulates the activity of SHIP2. For this aim, we will better characterize the dephosphorylation reaction mechanism carried out by SHIP, as it has been poorly studied and could help us understanding regulatory features.

Regarding the mechanism of activation of kinases:

3. How does the activation loop of Abl switch from an inactive to an active state? Are alternative intermediate states stabilized rather than the fully active or inactive states, like in Src? Does phosphorylation affect this stabilization?

All these questions aim to obtain an understanding of Abl activation and the role of activation-loop phosphorylation. We wonder the specific conformational influence of the activation loop which has been seen by crystal structures to adopt different conformations.

4. How does the presence of ATP/Mg^{2+} affect the FAK kinase domain?

   The main objective here is to elucidate the effect of ATP on the dynamic behaviour of the kinase within the context of the activation mechanism of FAK described in Goñi et al.

Regarding kinases inhibition:

5. Can we design a new efficient strategy combining computational and experimental work to find new scaffolds for specific inhibiton of protein signalling? As an example we will focus on FAK, however our approaches can be applied to any other drug target.

To infinity… and beyond.
Toy Story (1995)

Chapter 3

3. Results and discussion

3.1 Mechanism of phosphorylation

3.1.1 PHOSPHORYLATION REACTIONS IN cAMP-DEPENDENT PROTEIN KINASE (PKA) AND GLYCOGEN SYNTHASE KINASE-3β (GSK3β): FILLING THE GAPS WITH QM/MM SIMULATIONS

The aim of our work is to understand in detail the dynamics of the complete catalytic cycles of GSK3β and PKA, seamlessly linking all available crystallographic intermediate states through computation. In the following section, we show how our molecular dynamics/quantum mechanical simulations on PKA have allowed us to study the complete reaction mechanism (Figure 3.3), filling the gaps that have not been amenable to X-ray crystal structure determination. To assess the possible generality of the proposed mechanism we then extended the same methodology to GSK3β, another kinase which, unlike PKA, requires phosphate-primed substrates.

3.1.1.1 Phosphoryl transfer reaction

Our methods were first validated using the best studied PKA system as the test case to check that the results obtained were compatible with crystal structures available in the Protein Data Bank [1-3] and recent computational studies. Although the phosphoryl transfer reactions of PKA and GSK3β are highly homologous, we will be discussing both systems separately in order to better go in depth into their similarities and differences.

For PKA, the initial PKA-ATP-substrate complex was modelled using the reported complex of the enzyme with the so-called phosphorylated substrate SP20 peptide [170], which corresponds to residues 5 to 24 of a protein kinase inhibitor (PKI) variant

harbouring N20A and A21S substitutions (Figure 3.1). To build the peptide we used as template, PKA in complex with AMP-PNP and SP20, TTYADFIASGRTGRRASIHD (PDB code: 4HPU), where the underlined S is phosphorylated. The force field parameters for the phosphorylated residues were obtained from the AMBER parameter database [171].

We used the three-dimensional coordinates of the enzyme-peptide binary complex deposited in the Protein Data Bank with code 4HPT. The coordinates of the ATP cofactor were extracted from the best-fit superimposed ATP-bound PKA (PDB code 1ATP). Manganese ions were replaced by magnesium ions and the crystallographic water molecules around the active site were kept.

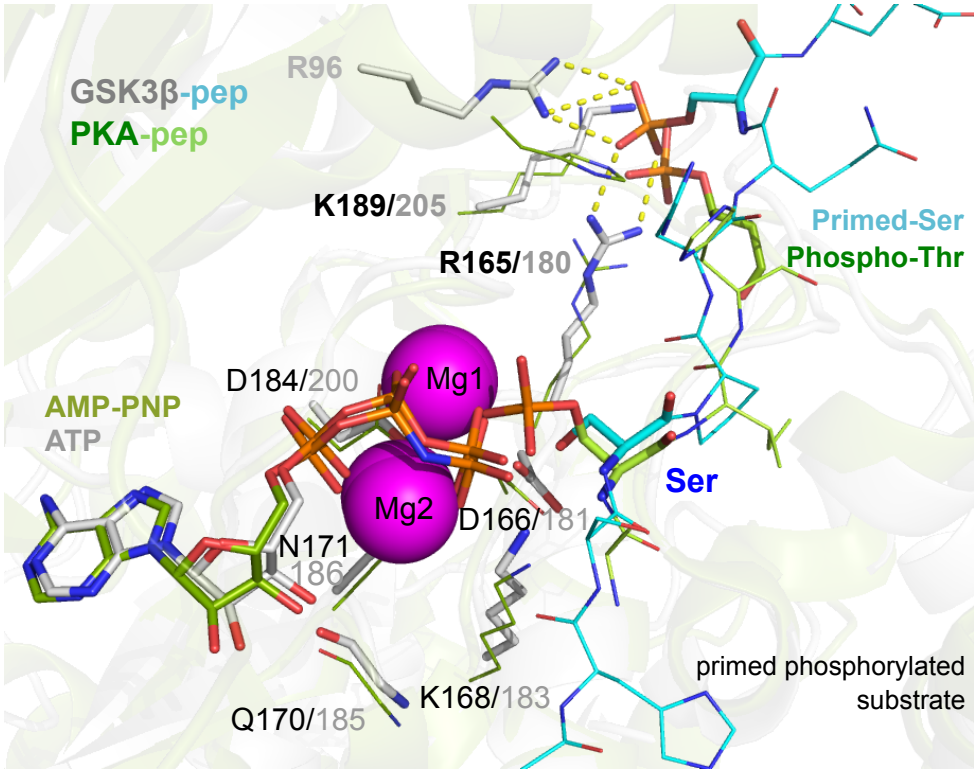


Figure 3.1. The structural superposition of the GSK3β-HSSPHQ^{P}SE and PKA-TTYADFIASGRTGRRASIHD^{P}T complexes shows the almost identical juxtaposition not only of the catalytic residues but also the primed ^{P}S in GSK3 with the phosphorylated Thr in PKA. This ^{P}S in the substrate is recognized, in the case of GSK3β, by the positively charged side chains of Arg96, Arg180 and Lys205. The numbering for PKA is in black, for GSK3β in grey. The protein and peptide colouring scheme is indicated at the top left.

Regarding GSK3β, the initial complex with the non-hydrolysable ATP analogue adenylyl imidodiphosphate (AMP-PNP), the two magnesium ions and the water molecules involved in metal coordination were extracted from the PDB entry 1PYX. The imidodiphosphoric acid nitrogen bridging the beta and gamma phosphates was replaced with an oxygen atom to yield ATP. To build the pGS (HSSPHQ^{P}SE) phosphoprimed peptide bound to GSK3β we used the GSK3 peptide GRPRTTSFAE in complex with AKT (PDB code: 1O6L) as a model for side chain replacement. The underlined S is the serine that undergoes the phosphorylation reaction (Figure 3.1) [172]. Each of the systems was first energy minimized and then equilibrated in order to find the most likely Michaelis complex.

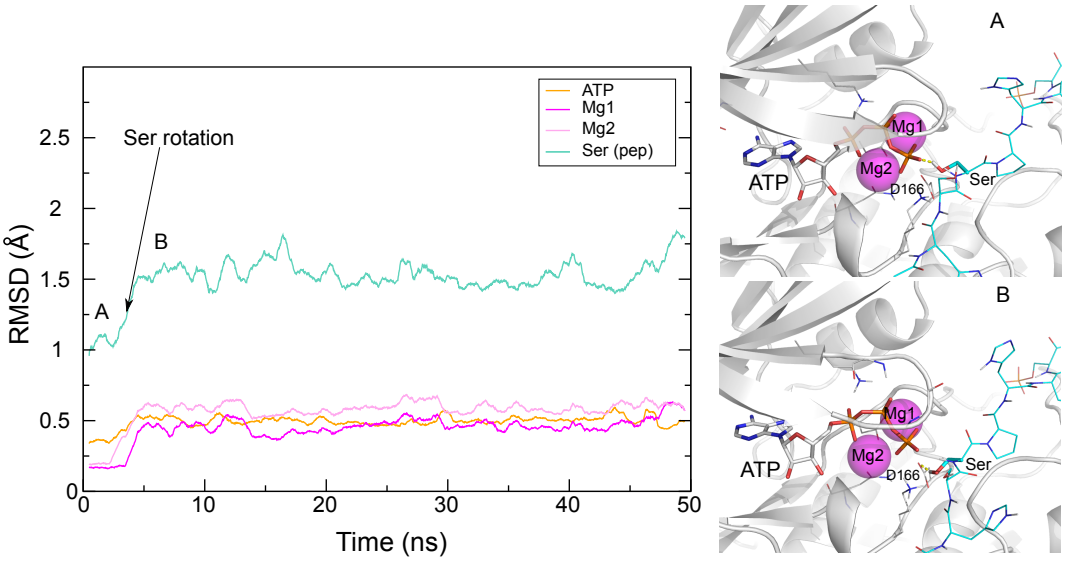


Figure 3.2. Time evolution of the Root-Mean-Square Deviation (RMSD) of ATP, Mg1, Mg2 and the serine that undergoes the phosphorylation in the simulated Michaelis complex. Two representative snapshots from A and B are represented on the right panel showing the position of every element.

After 3 ns it was observed that the Ser OH proton rotated pointing from the γ-phosphate towards the base (Figure 3.2) and it remained in this hydrogen-bonding arrangement for the rest of the simulation. Therefore we proceeded with the reaction in which the serine oxygen attacks the γ-phosphate while it is deprotonated by the aspartic acid residue acting as the base, namely Asp166 in PKA and Asp181 in GSK3β (Figure 3.3-a). Departing from this configuration we started the QM/MM simulation for the phosphoryl transfer over the course of 5 ps of MD simulation. The finding that deprotonation occurs immediately after phosphoryl transfer is most likely related to the fact that approach of the Ser hydroxyl would be very difficult if the proton was removed

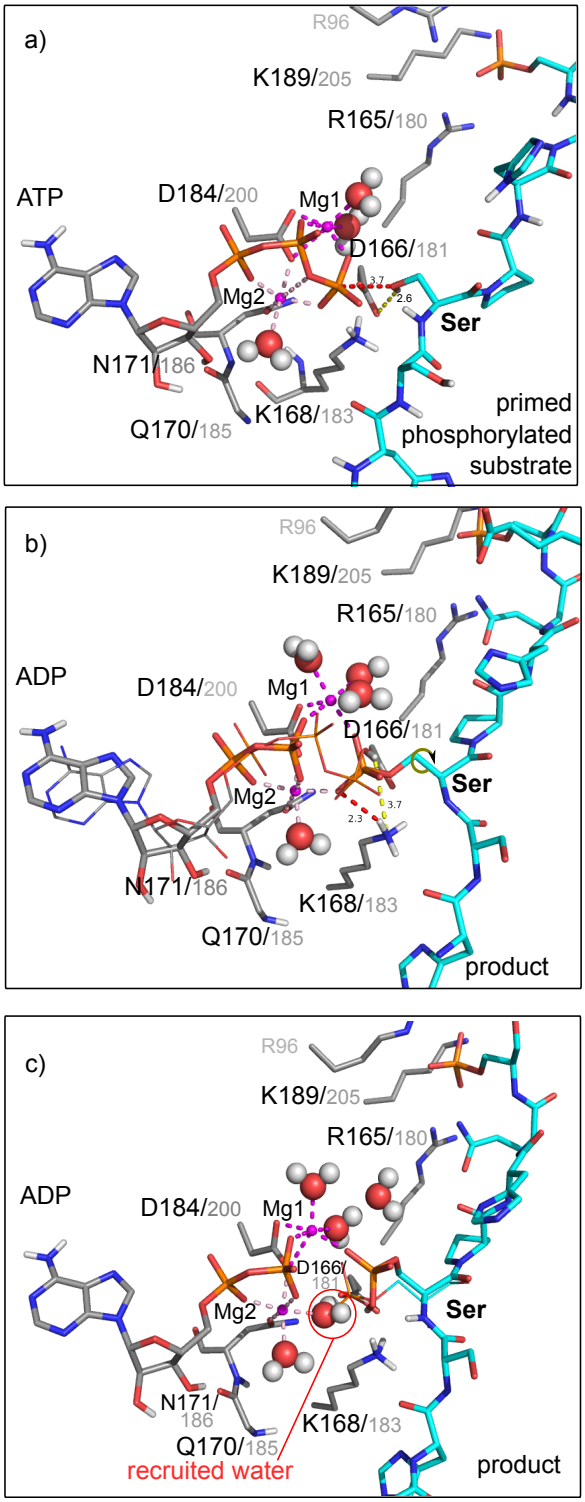
prior to the attack due to electrostatic repulsion by the phosphate. This result is also consistent with reports from Valiev et al. [57] who also described a late deprotonation event using different levels of theoretical chemistry methods.

Subsequently, we simulated, using classical MD, the resulting system containing ADP, phosphoserine in the peptide substrate and the protonated aspartic residue in the enzyme. The complex was found to be very stable during 200 ns of simulation and no significant structural rearrangements were found (Figure 3.4).

Excellent agreement was found between our simulated complex and the crystal structure describing the spontaneous hydrolysis of ATP in PKA (PDB code 1RDQ). Nonetheless, the phosphoserine conformation and the Mg^{2+} coordination spheres appeared to be quite different from those found by Taylor et al. [58] in their crystals after the full transfer of the phosphoryl group. In our simulated system, the two Mg^{2+} ions appeared to be coordinated by oxygens from both the phosphoserine and the ADP molecule, in addition to the water molecules that complete the coordination sphere (as found in the 1RDQ crystal structure). This charge *sandwich* of negative-positive-negative groups appears to be dynamically very stable and we could not see any spontaneous evolution from this point. We then reasoned that it was probably necessary first for a proton to be transferred so as to regenerate the conjugate base of the aspartic carboxylate (Figure 3.3-a). As a consequence of transferring this proton by means of a reaction coordinate using QM/MM, we did indeed observe the resolution of the formerly very stable intermediate and the rearrangement of the phosphoserine residue, the two Mg^{2+} ions and the ADP molecule so as to resemble very closely the final structures described by Taylor et al [2 - 4].

- Proton transfer reaction

We departed from the previous reaction intermediates in which the phosphoserine residue of the product is stabilized by both Mg^{2+} ions and the basic aspartic acid is protonated (Figure 3.3-b). Examination of the final structures of the reaction determined by X-ray crystallography strongly suggested that the final acceptor of this proton from the aspartic acid residue should be the O2 atom of the phophoserine residue on the basis of the change observed in this region of the complexes when 1RDQ (spontaneous ATP hydrolysis) and 4HPT (total transfer of the phosphoryl group to the substrate) structures were compared.



This transfer would then weaken the otherwise strong interactions between the Mg^{2+} ions and the ADP and phosphoryl group and allow the complex to evolve. Two different pathways are possible for this reaction to take place (Figure 3.3-b):

a) the proton is directly transferred from Asp166/181 to the phosphoryl group, or b) the highly conserved neighbouring lysine, i.e. Lys168/183, first donates one of its protons to the phosphate and then recovers its original protonation state by abstracting the hydrogen from the aspartic acid residue.

Whatever the pathway for this transfer, the resulting system consists of a product with monoprotonated phosphoserine, ADP^{3-} and two Mg^{2+} ions whose coordination spheres are completed with water molecules. Classical MD simulations of this ensemble show that the complex evolves from an initial configuration in which the Mg^{2+} ion is coordinated with the phosphoserine (Figure 3.3-b) to another in which a bulk water

Figure 3.3. Snapshots of crucial reaction steps in GSK3β: a) proton transfer from the Ser to be phosphorylated to Asp166 (yellow dotted line) and Ser-O attack of the γ-phosphate in ATP (red dotted line). b) Catalytic Asp166 is protonated and phosphoryl group (sticks) is transferred from ATP to the Ser (ATP from the previous step as lines). c) Deprotonation of Asp166 by outgoing phosphorylated Ser (sticks) overlaid on its previous conformation (lines). Waters are shown with oxygen and hydrogens as small spheres.

molecule solvates this moiety and favours its separation from the metal coordination sphere (Figure 3.3-c). The subsequent rearrangement of the phosphoserine residue in this limited space drives the complex to the situation that can be observed as the final state of the phosphoryl transfer reaction in the 4HPT and 4IAK crystal structures. The time evolution of the RMSD shown in Figure 3.4 reveals that this product is very stable and no significant deviations from the starting configuration are observed for ADP, Mg^{2+} ions and the phosphoserine.

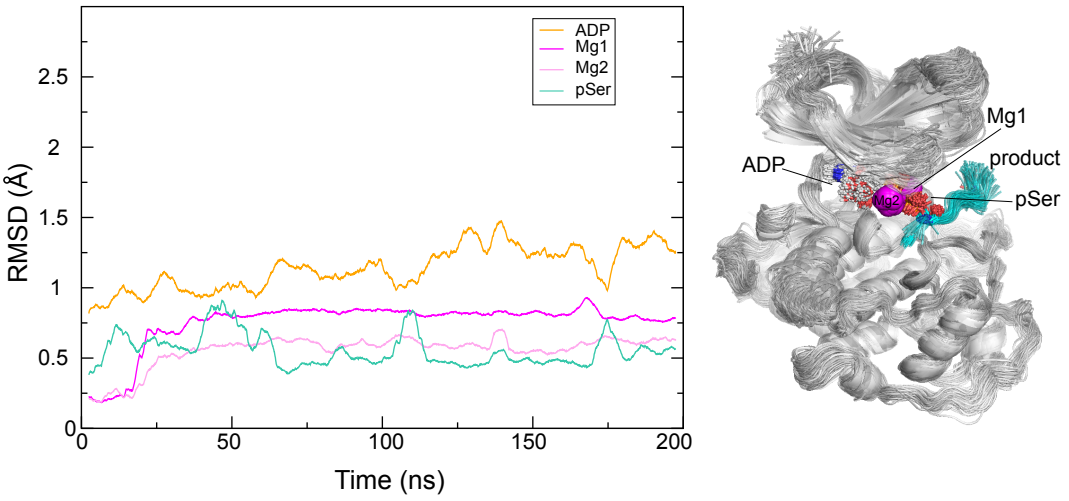


Figure 3.4. Evolution of the RMSD of ADP, Mg1, Mg2 and phosphoserine after the phosphorylation reaction. On the right, an overlay of MD snapshots separated by 1 ns shows the stability of the product.

Due to the very good agreement between the simulation results and available high-resolution crystal structures, we observed that the active sites from the reactant and the product are geometrically very similar, suggesting that the elements of the complex before and after the reaction are almost identical for both systems, GSK3β and PKA. However, there are two major changes in this phospho transfer reaction: the former is the orientation of the Ser undergoing through phosphorylation, which started pointing towards the aspartic acid and ended pointing away the Asp181 and looking to the solvent. The other change is the ions coordination. While Mg1 maintains its interactions with $PO3^-$ group and phosphoserine, Mg2 loses coordination to the phosphoserine gaining extra water in the product structure.

3.1.1.2 GSK3β catalytic cycle and lithium inhibition

The GSK3β catalytic site, as already stated, is almost identical to that of PKA with only two replacements in the amino acids lining the substrate binding pocket, namely Phe187→Ser203 and Thr51→Asn64. However, this resemblance does not necessarily mean that the steps of the catalytic cycle are strictly similar to those reported for PKA, particularly if we consider that the activation mechanisms are different for both enzymes: PKA requires the phosphorylation of Thr197 while GSK3β needs to be phosphorylated on Tyr216 and prefers substrates that are phospho-primed at key positions [6-7].

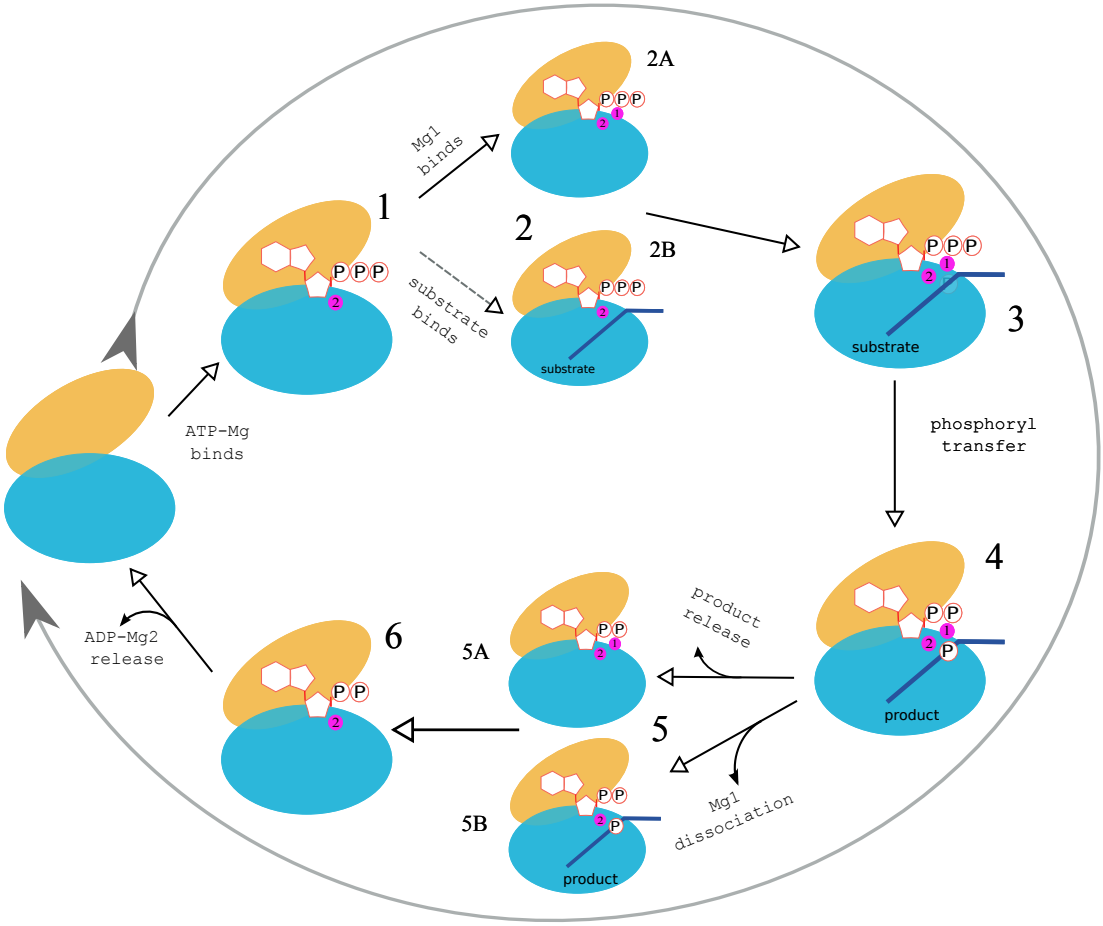


Figure 3.5. Stepwise decomposition of a complete catalytic cycle for PKA and GSK3β. Step 1: ATP-kinase complex; step 2: Binding of the second Mg^{2+} ion and the substrate make up the Michaelis complex; step 3: Phosphoryl transfer reaction; step 4: Proton transfer event; step 5: Release of both product and one Mg^{2+} ion; step 6: Release of ADP-Mg^{2+} and beginning of new cycle.

To assess whether these important differences had an impact on the generality of the mechanism reported above and on lithium inhibition in particular, we simulated steps 2, 3, 4 and 5 (as named in Figure 3.5) with both Mg^{2+} ions and with one Mg^{2+} and one Li^{+} ion to analyse their respective stabilities of the respective complexes.

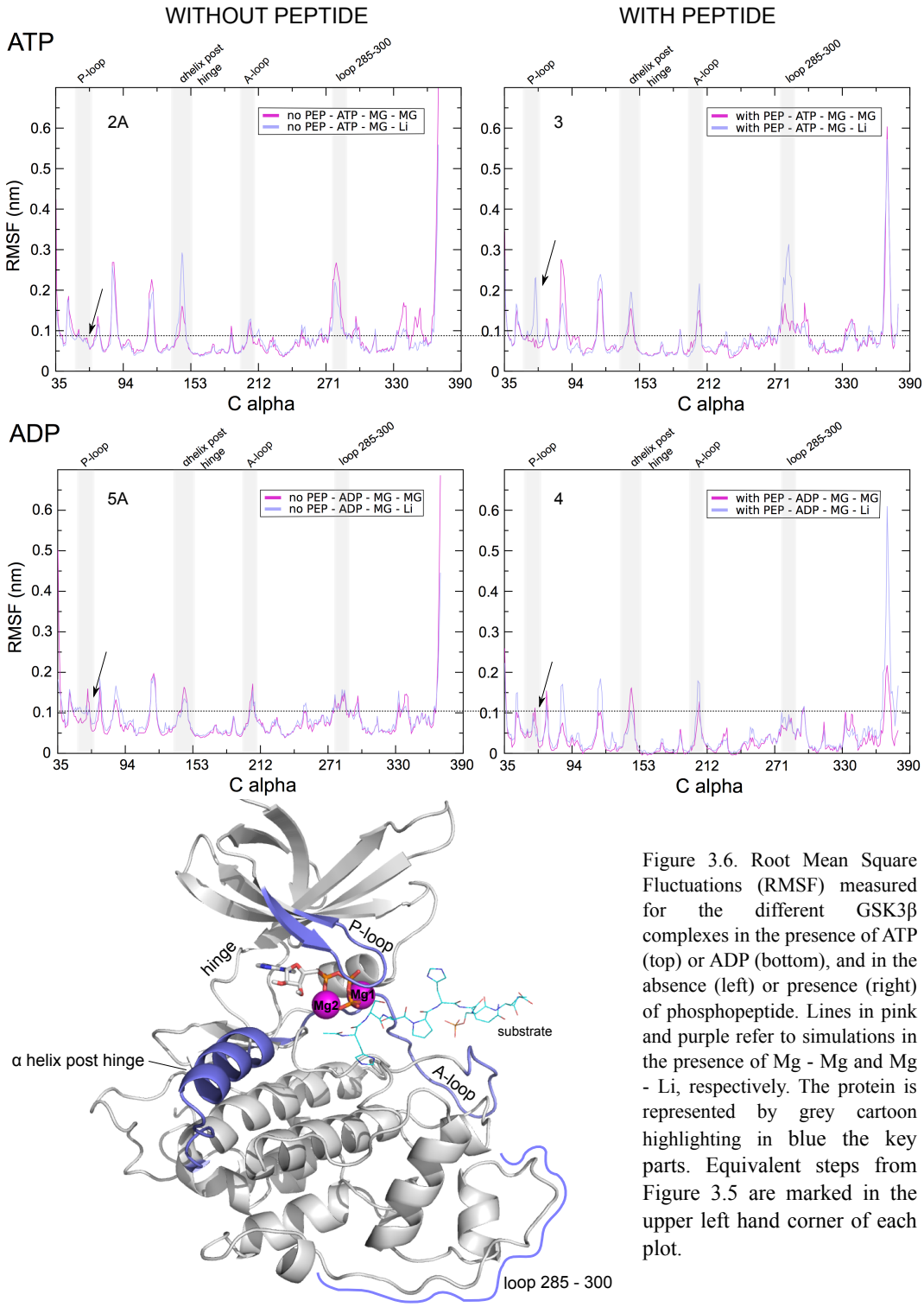


Figure 3.6. Root Mean Square Fluctuations (RMSF) measured for the different GSK3β complexes in the presence of ATP (top) or ADP (bottom), and in the absence (left) or presence (right) of phosphopeptide. Lines in pink and purple refer to simulations in the presence of Mg - Mg and Mg - Li, respectively. The protein is represented by grey cartoon highlighting in blue the key parts. Equivalent steps from Figure 3.5 are marked in the upper left hand corner of each plot.

As lithium has been shown to be a competitive inhibitor of GSK3β with respect to Mg^{2+} [174], we studied which of the two divalent cations is the most likely candidate to be replaced by Li^{+}. The oxygens from β- and γ –phosphoryl groups of ATP, Asp200 and two water molecules coordinate Mg1, while Mg2 at the high-affinity site has a trigonal bipyramidal geometry in its coordination with two oxygens of the α- and β-phosphoryl groups of ATP, one carboxylate oxygen from Asp200, the carboxamide oxygen from the side chain of Asn186 and a water molecule. It was shown by Bastidas et al. [58] that for PKA Mg1 is partially, if not completely, expelled following phosphoryl transfer (steps 5B and 6 in Figure 3.5), which implies that it may not bind as strongly and possibly may not influence the active site as much as Mg2. Additionally, we found that Mg2 recruits a water molecule immediately following phosphoryl transfer to maintain an octahedral coordination state. This recruited water molecule is also present in the crystallographic study published by Bastidas et al. in their PKA structures showing partial and complete phosphoryl transfer. The presence of this water is most likely due to the strong binding character of Mg2, which may help to catalyse the phosphoryl transfer reaction and/or stabilize the transition state. Taking all this information into account we conclude that Li^{+} most likely competes for binding to the same site as Mg1.

After the unrestrained MD simulations of steps 2A, 3, 4 and 5A, we observed that the time evolution of the Root Mean Square Fluctuations (RMSF) profiles of the eight systems (+ATP with (3) and without (2A) peptide and with and without Li+ in Mg1 position; +ADP with (5A) and without (4) phosphorylated peptide and with and without Li+ in Mg1 position) was very similar in all cases, although differences were apparent in three regions: the P-loop or glycine-rich loop which plays an essential role in stabilizing the phosphate; the A-loop which contains the phosphorylation site and whose conformation, among others, dictates whether or not the substrate can bind; and loop 285-300 underneath the α-helix G. These three regions showed higher flexibility for steps 2A and 3, which can be taken as an indication of increased stability upon phosphorylation (4 and 5A) and suggesting the possibility of an allosteric effect or connection between the active site and the C-terminal part of the kinase which are located far away one from each other. In addition, when Li^{+} replaces Mg1, the RMSF profile of state 3 gets significantly disturbed within the P-loop region, 3 (GSK3β with ATP and peptide).

Since the stability of this Gly-rich loop is essential for the phosphorylation reaction to take place, we propose that this could be the conformational state to which Li^+ binds. Inhibition of enzyme activity would then ensue because the orderly arrangement that normally leads to the near-attack conformation is thus prevented.

3.1.1.3 Discussion

The results of our simulations confirm the general picture that we obtained for PKA. First, in the Michaelis complex the serine residue of the peptide substrate is close to the γ-phosphate of ATP while its proton points towards the Asp181 carboxylate. The ammonium group of Lys183 is hydrogen-bonded to both ATP and the aspartic acid acting as the base that stabilizes the complex together with the interactions of the P-loop (Ser66 and Gly65). The fact that Ser66 also interacts with the substrate results in a complex that is tightly restrained. This limited motion appears to be key for the reaction to proceed. Phosphoryl transfer appears to occur through a mechanism in which the phosphoryl group approaches the serine hydroxyl whose proton is fixed by a hydrogen bond with the aspartate. In this case, proton transfer from the serine to the carboxylate also appears to be a late event. Thereafter, ADP, the intermediate product and the base keep their relative positions in an arrangement very similar to that reported above for the PKA-substrate complex and also to that observed in PDB entry 1RDQ showing PKA and the products of the spontaneous hydrolysis of ATP (ADP + P_i). The GSK3β-phosphopeptide intermediate is also stabilized by strong charge-charge interactions. Only upon proton transfer from Asp166 to the phosphate does the complex evolve to an

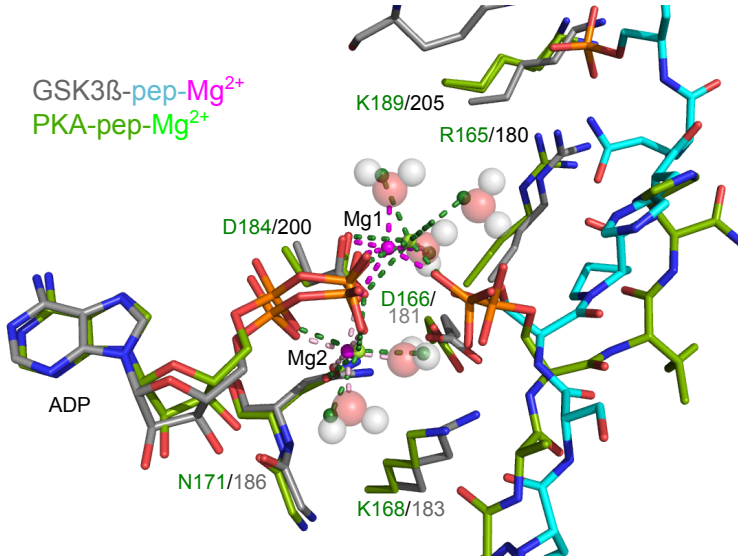


Figure 3.7. Superposition of the active sites of the product complexes: GSK3β-Mg_2-ADP-phosphopeptide after QM/MM simulations. GSK3β protein and peptide carbon are coloured grey and cyan respectively, Mg^{2+} ions are magenta and waters are represented by spheres for oxygen and hydrogen atoms. The X-ray structure of PKAc-Mg^{2+}-ADP- SP20 (PDB ID 4IAK) is shown with light green carbon atoms, lemon Mg^{2+} ions, small dark green spheres represents the crystallographic waters).

arrangement similar to that observed for PKA which facilitates product release (Figure 3.7).

Therefore, despite the general differences in substrates and activation mechanisms between PKA and GSK3β, the catalytic cycle of these two kinases appears to follow the same general trend. Importantly, phosphoryl transfer by itself is not enough to explain the structural changes in the active site that have been observed in the crystallographic structures. It is also crucially important to consider the subsequent proton transfer that allows the charge-charge interactions between Mg^{2+} ions and phosphoserine to weaken.

Regarding lithium inhibition, it has repeatedly been shown that Li^{+} is a GSK3β competitive inhibitor with respect to Mg^{2+}, but not with respect to ATP or the substrate. In our simulations we observed that not only did Li^{+} destabilize key regions (P-loop) that are crucial for the phosphoryl transfer reaction, but one of the most likely states for this destabilization to occur in presence of Li^{+} is before the phospho transfer reaction takes place, when the GSK3β-ATP-Mg-peptide complex is formed.

3.1.2 SUBSTRATE SPECIFITY OF THE SHIP2 PHOSPHATASE DOMAIN AND ALLOSTERIC EFFECTS OF THE C2 DOMAIN.

Recently the crystal structure of a SHIP2 fragment containing the phosphatase (5-Ptase) and C2 domains was solved in our group (manuscript in preparation). The two domains interact by means of an interface consisting of a central largely hydrophobic patch and a peripheral polar region (Figure 3.8). The substrate and metal binding sites are located on the top part of the 5-Ptase domain. Biochemical analysis demonstrated that the C2 domain enhances activity towards the phosphatidylinositol (3,4,5)-trisphosphate ($PI(3,4,5)P_3$ or PIP_3) lipid substrate by more than 10 fold, whereas only approximately 1.5 fold for the soluble Inositol 1,3,4,5-tetrakisphosphate (IP_4) substrate. The C2 domain is therefore required for full activity, but also plays an important role in substrate specificity. From the crystal structure alone, the mechanism of how the C2 domain affects SHIP2 catalysis is not clear since it is positioned on a side of the 5-Ptase domain (Figure 3.8) that is opposite to the catalytic centre. In addition, the catalytic mechanism is still not fully understood for the 5-Ptase family. To understand how the C2 domain regulates the catalytic activity it is essential to shed light on the exact catalytic mechanism and to model the Michaelis complex.

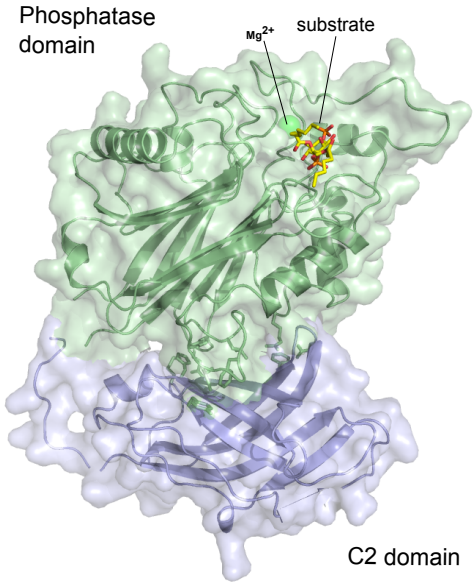


Figure 3.8. Crystal structure of phosphatase (green) and C2 (blue) domains as solved in our group. The substrate (stick representation), the Mg^{2+} ion (green sphere) and the loop linking phosphatase and C2 domains have been modelled. The interacting residues at the interface are shown as sticks.

3.1.2.1 Dephosphorylation reaction

Until 2001, the 3D structure of 5-phosphatases remained somewhat obscure. However, new knowledge arising from X-ray crystallographic work on the catalytic domain of several inositol polyphosphate 5-phosphatases [175], [176], together with biochemical data [177], [178], has been instrumental to realize the high degree of functional and structural homology of these phosphatases with the large and diverse family of Mg^{2+}-dependent nucleases, including the Apurinic/Apyrimidinic (AP) endonucleases. AP endonucleases and the 5-phosphatases not only catalyse chemically analogous reactions (the cleavage of a sugar-phosphate bond) that require divalent cations (e.g., Mg^{2+}) but they also share all the key residues involved in catalysis, as clearly evidenced when we performed a multiple sequence analysis [179] using more than 100 sequences from both families of enzymes (selected sequences are shown in Figure 3.10). Moreover, those conserved residues within the catalytic domain adopt identical arrangements in the 3D structures (Figure 3.9).

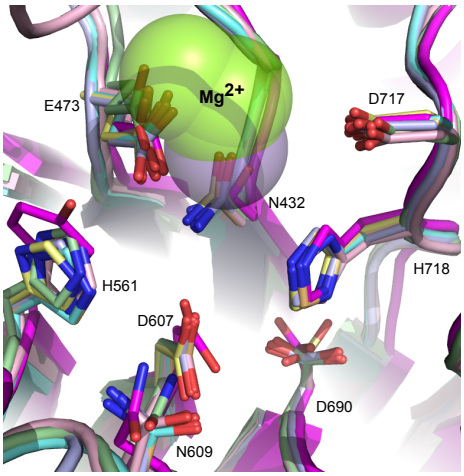


Figure 3.9. Superposition of the catalytic site of the 1DE9.pdb AP endonuclease (shown in magenta) and different 5-phosphatase family members (SHIP2, shown in green, *S. pombe* synaptojanin IPP5C in pink, INPP5B in complexes with different products [not shown here for clarity] in cyan, orange and violet, and OCRL in yellow). Mg^{2+} and Mn^{2+} cations are represented as green and purple spheres, respectively. Residue numbers are shown for SHIP2.

Using recent crystallographic data available in our group, we produced a model of SHIP2 in complex with the substrates PIP_3 and IP_4 that are consistent with the AP endonucleases' mechanism (Figure 3.11). In addition, we created models using the X-ray crystal structure of another SHIP2 phosphatase (PDB entry 3NR8), which differs from our recent crystal structure mainly in one loop conformation located at the upper part and is involved in substrate recognition (Figure 3.10-b).

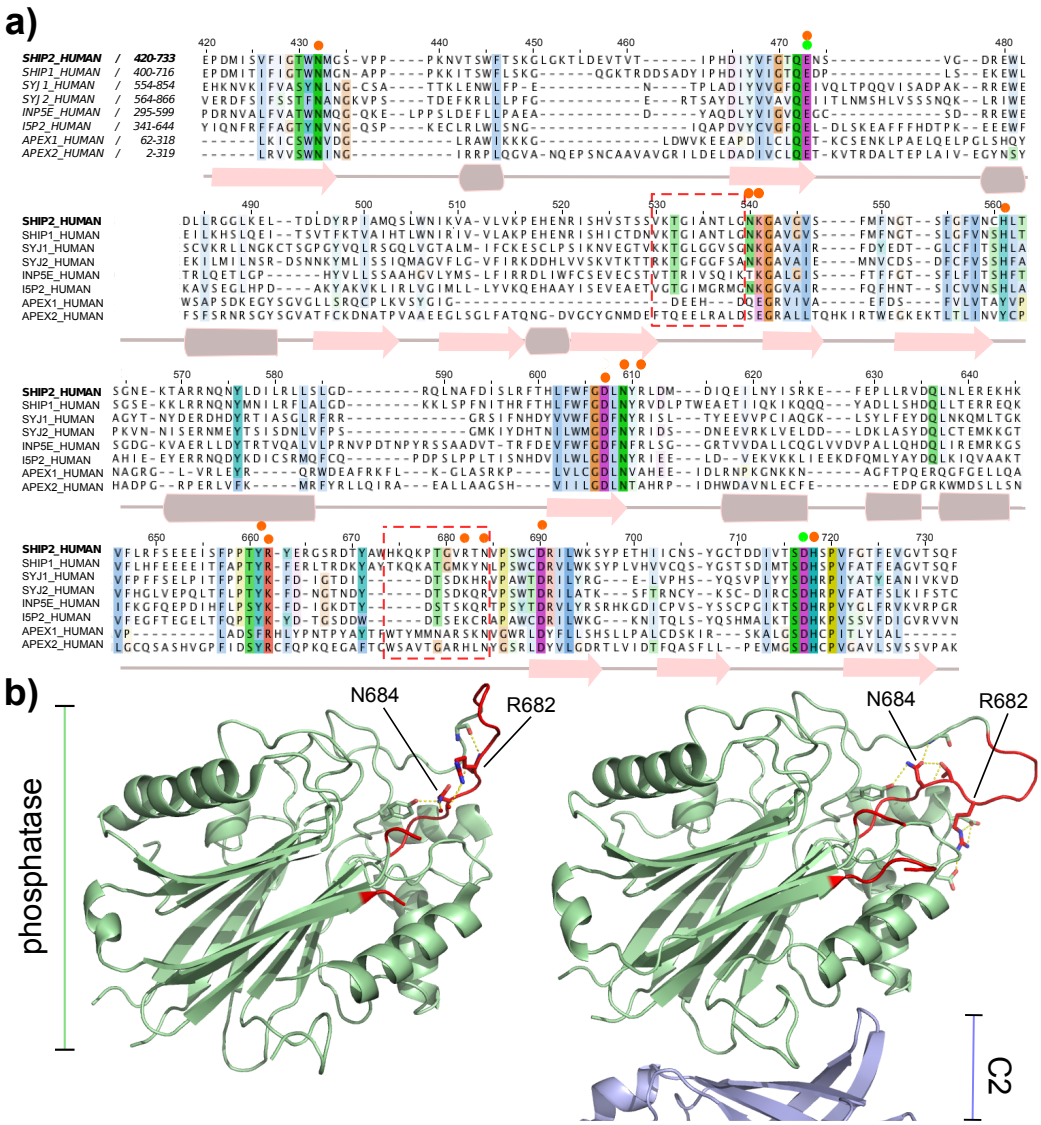


Figure 3.10. Structural and sequence comparison of the 5-phosphatase catalytic module. a) Sequence alignment of 5-phosphatases of SHIP2, SHIP1, SPsynaptojanin1 and 2 (SYJ1 and SYJ2), INP5E, I5P2, APendonuclease1 and AP2endonuclease2. The sequence colouring scheme is based on ClustalW by conservation and the residues involved in substrate recognition are indicated with an orange circle while green circles indicate residues coordinating the metal. b) Structural analysis of SHIP2 with two variable regions highlighted by dashed red circles and Arg682 and Asn684 shown in sticks. From left to right: phosphatase domain alone (PDB code: 3NR8), phosphatase in complex with biphenyl pentakisphosphate (B5P) showing interactions between B5P with Arg682 and Asn684 in red dashes (PDB code: 4A9C) and phosphatase + C2 domains showing the interactions of Arg682 and Asn684 with other residues as yellow dashed lines. The latter structure corresponds to a recent X-ray crystal structure solved in our lab that is not published yet.

We observed that several residues in the active site are able to adopt different conformations, especially residues in the loop that varies its conformation in our starting X-ray structures. For example, the characteristic Arg682 from the SHIP2 family, which is specific to SHIP 5-Ptases, explaining the requirement for 3-phosphorylated substrates for SHIP, is not well conserved among other phosphatases and AP endonucleases, and is able to adopt two different conformations (Figure 3.10-b).

Consistent with our findings, on the one hand X-ray crystal structures of the Ptase domain show that if the loop is in an "in" conformation, pointing into the active site (PDB codes: 3NR8 and 4A9C in Figure 3.10-b), the Arg682 is positioned in such a way that it can interact with the substrate. On the other hand, 6 of the 8 Ptase+C2 structures present in the asymmetric unit of the crystal structure solved in our group show an alternative conformation where the loop is pointing outside the substrate binding site and the Arg682 is interacting with Asp613 and/or Asp615. Potentially, this "out" conformation is important to allow access to the catalytic site to facilitate substrate entry and binding. Interestingly, this out conformation has only been only observed in crystal structure where the C2 domain is present, whereas in all 4 available structures of the isolated Ptase domains, the loop is either in an in-conformation or is disordered.

Another residue that can adopt different conformations is Asn684 (an Arg in other 5-Ptases, see Figure 3.10-b), which has in other phosphatases been observed to play an important role in phosphoinositide recognition [175]. Asn684 can either point towards the substrate site when the loop is in the "in" position, or it can point upwards interacting with Tyr671, Thr683 and the backbone of the His674, in which case it does not promote substrate binding (right structure on Figure 3.10-b).

In order to clarify the catalytic reaction, we modelled the Michaelis complex based on the proposed catalytic mechanism of AP endonucleases (Figure 3.11). We were able to position all the required elements for catalysis as follows: the target phosphate was oriented and polarized via its interaction with the imidazole of His718 (which was paired with Asp690) and Asn609. In addition, we propose that His561, which has been shown to be required for activity [64], forms a hydrogen bond to the O2P atom of the

scissile phosphate. Nucleophilic attack is initiated by Asp607, which aligns and activates the attacking water molecule. Asp607 is oriented by hydrogen bonds to the side chain of Asn432 and the backbone amide of Asn609 [64]. This mechanism entails the presence of an attacking (catalytic) water molecule bound to Asp607 in the Michaelis complex. This water, included in our simulations and only present in some X-ray structure (PDB codes: 3MTC, 1I9Y, 4CML), is crucial for placing all the key components to achieve the reaction.

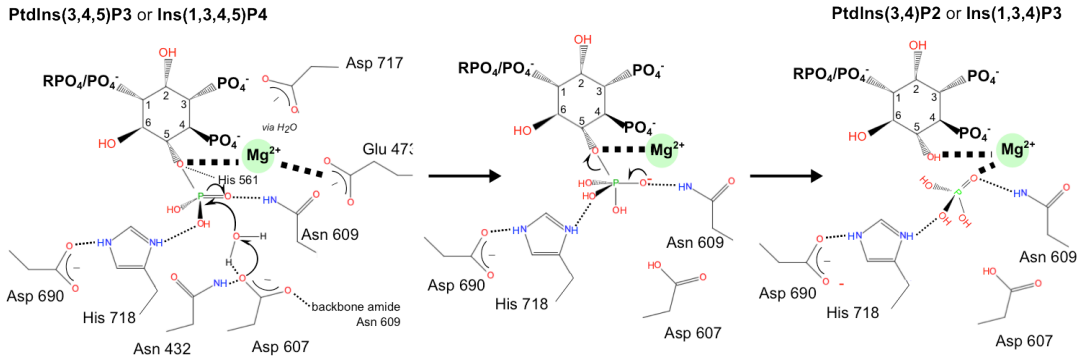


Figure 3.11. SHIP2 catalytic reaction pathway proposed based on AP nucleases [64].

Taken together, all of these observations suggest a common catalytic mechanism [64] that we will further study by means of MD simulations.

Initially, we performed unbiased MD simulations of the following eight systems with both loop conformations, "in" and "out", and the 5-Ptase with and without the C2 domain with either the IP_4 headgroup or with the PIP_3 lipid, containing 8-carbon acyl chains. After 300ns of simulations within the in-conformation (the in-attack conformation), all the key residues remain stable. However, there is a significant difference in the stability of the substrates in presence or absence of the C2 domain, being the latter one more stable with lower values of RMSD compared to the headgroup of Ptase alone, which is around 2 Å lower (Figure 3.12).

In addition to the apo-SHIP MD simulations, we run MD simulations either on the phosphatase alone or the phosphatase + C2 structures, also using both loop conformations and in the absence of any bound substrate, to study the behaviour of the system.

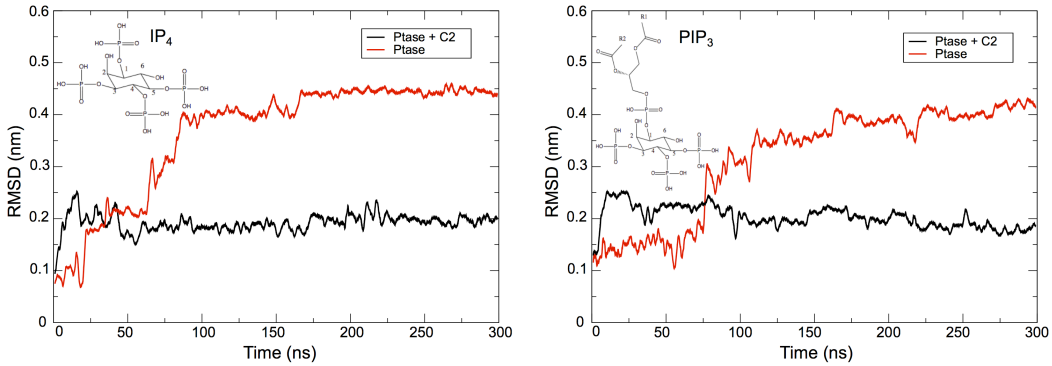


Figure 3.12. The RMSD of substrate's head group is shown through the simulation compared to the starting structure for the Ptase (red plots) and Ptase+C2 (black plots) in presence of IP_4 (left) or PIP_3 (right).

3.1.2.2 How does the C2 domain regulate SHIP2 activity?

The enzyme kinetics data performed in our group showed that the C2 domain enhanced significantly SHIP2's overall activity (wild-type (wt) K_{cat} shown in Table 3.1). In addition, the results suggest that the C2 domain is involved in substrate selectivity. Whereas for the Ptase IP_4 is the preferred substrate, the Ptase+C2 protein performs approximately 4 times faster with the PIP_3 lipid.

	Phosphatase		**Phosphatase + C2**	
	I(1,3,4,5)P_4	**PI(3,4,5)P_3diC8**	**I(1,3,4,5)P_4**	**PI(3,4,5)P_3diC8**
	K_{cat} (s^{-1})		K_{cat} (s^{-1})	
wt	1.32 ± 0.02	0.69 ± 0.02	2.02 ± 0.04	7.83 ± 0.26
F593D+L597D	1.20 ± 0.05	4.44 ± 0.17	1.39 ± 0.06	4.79 ± 0.49
R649A	1.64 ± 0.05	1.60 ± 0.09	1.79 ± 0.05	6.37 ± 0.42
D613A+D615A	NA	NA	1.45 ± 0.05	7.10 ± 0.57
R665A	0.53 ± 0.03	0.23 ± 0.02	2.07 ± 0.09	7.02 ± 0.37
R682A	0.13 ± 0.08	ND	0.24 ± 0.07	1.74 ± 0.23
N684A	ND	ND	ND	ND
NA: Not available; ND: Not detectable				

Table 3.1. Experimentaly determined K_{cat} comparing wt Ptase, wt Ptase + C2 (in grey), and the different mutants in for both substrates, IP_4 and PIP_3. (Data provided by Dr. J. Le Coq).

Analysis of long unbiased MD simulations of 4.35 μs at 300 K of both, the Ptase alone and the Ptase plus the C2 domain with no substrate bound (Figure 3.14-b) showed different patterns of flexibility. In particular, we observed that loop 3 exhibited increased flexibility in the absence of the C2 domain, since loop 3 is located at the domain interface between Ptase and C2 domains (Figure 3.14-a). Further, we observed higher flexibility of the loops 2 and 4 in the presence of the C2 domain, although both loops are located very distant from the C2 domain. In spite of the loops being the most disordered elements in protein structures, and flexibility is one of their intrinsic characteristics, using the structure-based approach, ACA, based on elastic network models [180], [181], we were able to allosterically connect loop regions that showed higher flexibility in the simulations and were located within distant locations in the protein (Figure 3.14-d). When the C2 was present, the motions of loops 1, 2 and 4 appeared to be coupled, whereas that of loop 3 was not coupled as it lost part of the flexibility. Interestingly, in the absence of the C2 domain, the coupling of loop 1 was strongly reduced.

Surprisingly, an element in the Ptase domain formed by three helices (shaded grey in Figure 3.14-a/b) showed a significantly reduced RMSF values in the presence of the C2 domain. Interestingly, this helical section reaches from the domain interface almost to the active site, hence providing a possible allosteric connection between the C2 domain and the substrate binding site. When the C2 is present the connection is stabilized and the catalytic reaction is favoured. Additionally, we performed PCA on the MD simulations. For the first eight PC of each simulation we observed that in the absence of the C2 domain there is a large spring-like motion of the helical section mentioned above (coloured grey in Figure 3.14-c). This motion was not observed when the C2 was present suggesting that stabilizing this helical part could allosterically communicate with the active site in the Ptase domain.

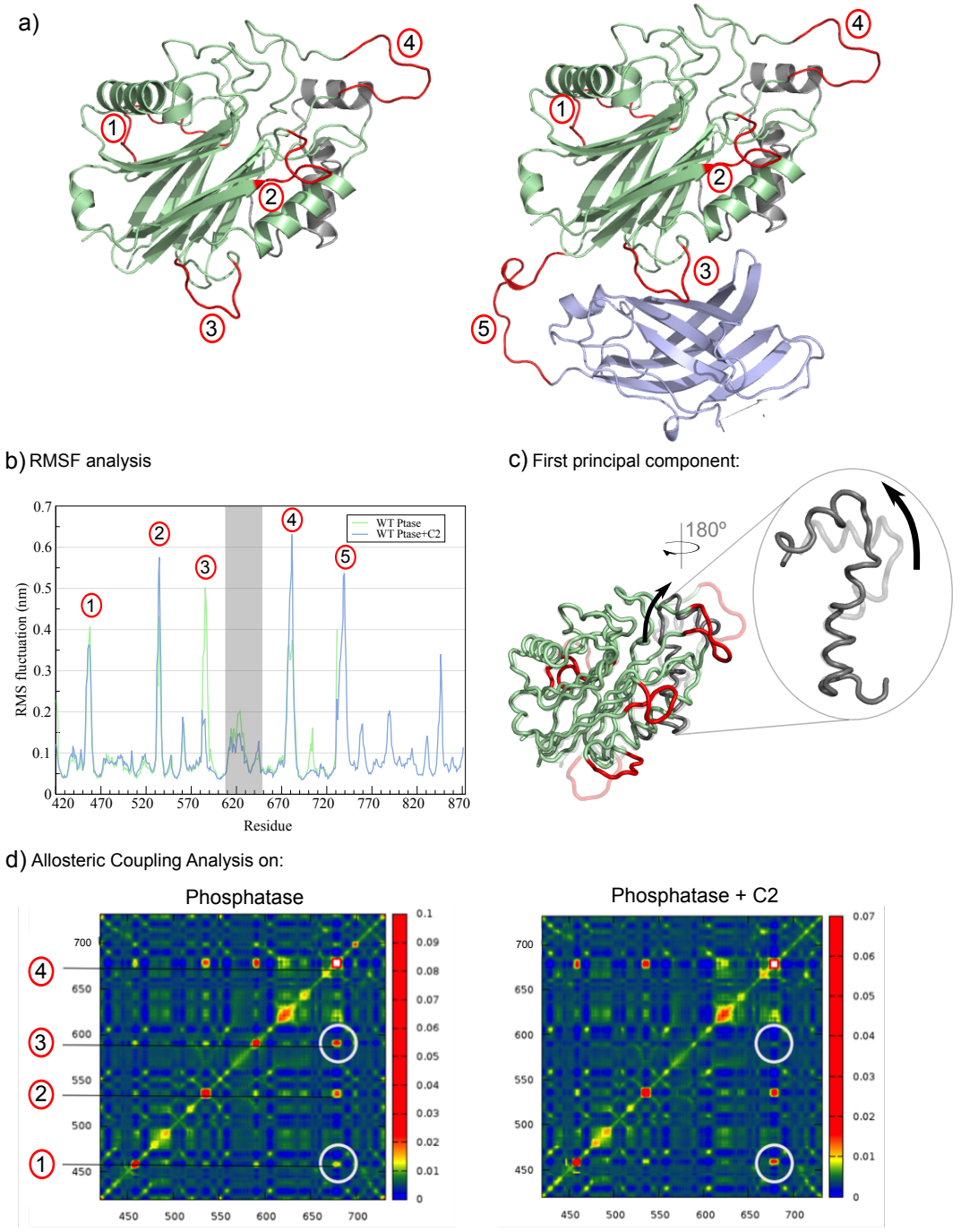


Figure 3.14. Comparison of the dynamics of the Ptase with Ptase+C2. a) Cartoon representation of both SHIP2 fragments with phosphatase coloured in green and C2 in blue. Different loops are numbered and coloured red. b) RMSF analysis of both MD simulations. Most loop regions show high RMSF values. The region corresponding to the helical section is shaded grey. c) Projection of the first eigenvector of the phosphatase PCA: the two extreme projections along the MD trajectory on the average structure are shown. d) Allosteric coupling analysis on the phosphatase (left) and phosphatase+C2 (right). For clarity only the analysis of the phosphatase region (residue 410 to 730) is shown. The largest differences are enclosed in white circles.

Inversely related to the helical effect, there is an already mentioned increase in the fluctuations of loops 2 and 4 (Figure 14-b). This effect mostly disapears for the MD simulations starting from the loop 4 "in" and when the substrates are bound (Figure 3.15).

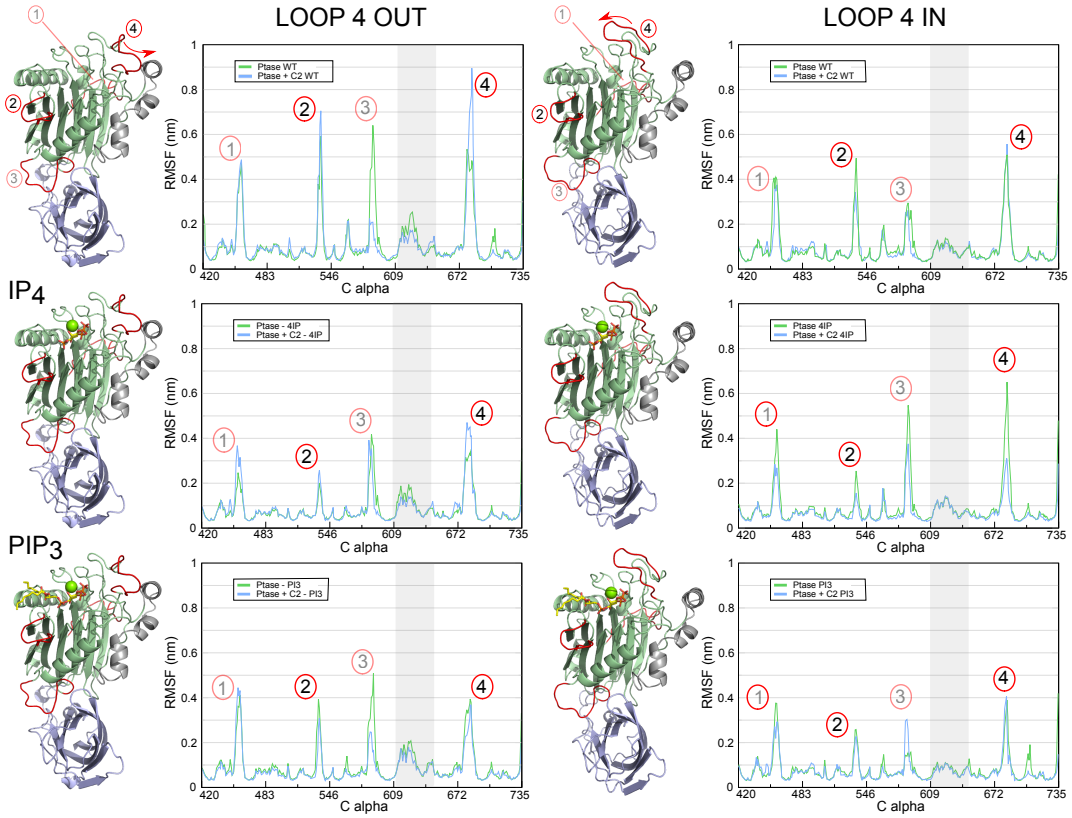


Figure 3.15. RMSF plots of all MD simulations: in absence of substrate (top panels), in presence of IP_4 (middle) or PIP_3 (bottom), starting with Loop 4 in an "out" position (left side) or with loop 4 in an "in" position (right side). Ptase measurements are coloured in green and Ptase + C2 in blue. Cartoon representations with loops highlighted in red colour are included on the left side of each plot for clarification.

Therefore the stabilization of this helical part when the C2 is present is very similar to the effected by the substrate in the absence of the C2 domain. This could suggest that the reduction in fluctuation of the helical part in presence of the C2 domain may favour substrate binding.

Since our analysis supports the idea that the C2 domain stabilized the phosphatase allosterically, reducing the fluctuations of a helical section in the Ptase domain and

increasing the fluctuations of loops 2 and 4, we analysed the main interactions between the helical part, loops 2 and 4. By performing cluster analysis on the MD trajectories we observed how the conformational space was differently explored not only depending on which conformation the MD started, but also, and more significantly, depending on the presence/absence of C2 domain. As we can appreciate in Figure 3.16, in the Ptase simulation with an initial conformation in which loop 4 is pointing outside the binding site, loop 4 and especially the interactions between Arg682 and Asp613 or Asp615 were more stable (Figure 3.16-a in red plot). However, when we started our simulations in the same conformation but with the C2 domain, the interaction between the arginine and the aspartic acids is broken in two clusters (Figure 3.16-a – clusters 2 and 3) allowing loop 4 to switch from the "out" to the "in" conformation.

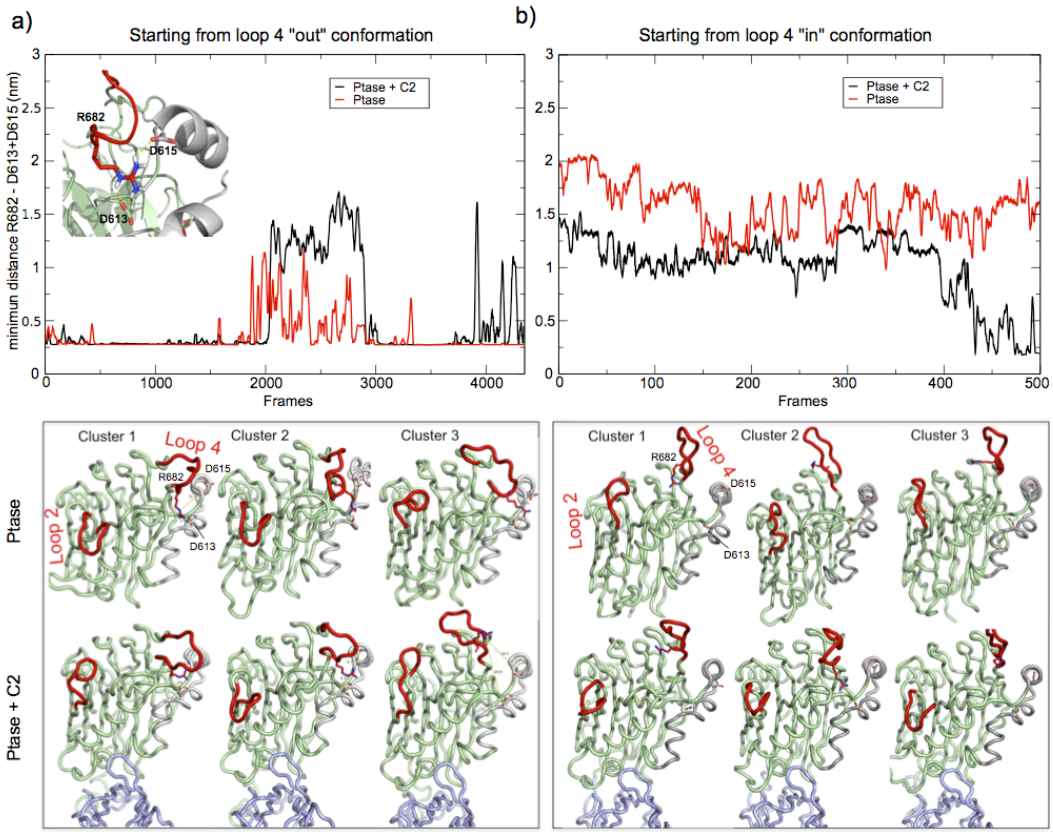


Figure 3.16. Analysis of simulations starting with loop 4 in an "out" (a) or "in" (b) conformation. The minimun distance between the amino groups of Arg682 and the hydroxyl groups of Asp613 and Asp615 are plotted in the upper graphs. Representative structures of the 3 most frequent states from the corresponding cluster analysis is shown at the bottom.

A similar effect was observed for the simulations starting with the loop conformation in an "in" position; while the Ptase alone did not explore any other conformation rather than the one pointing towards the binding site or similar, the Ptase + C2 domain could reach different conformations, including the complete opening of the loop, forming the salt bridges between Arg682 and Asp613/Asp615 at the end of the simulation (black plot and cluster 2 in Figure 3.16-b). Switching between loop 4 conformations is likely important to allow substrate entry and binding (requires "out" conformation) and correct positioning of substrate for catalysis (requires "in" conformation), hence providing a feasible mechanism how the C2 domain allosterically increases catalytic turnover.

3.1.2.3 Site-directed mutagenesis

Several residues either involved in the Ptase-C2 interface, the allosteric communication to loop 4, or directly in substrate binding were subjected to site-directed mutagenesis: F593D+L597D, D615A+D613A, R649A, R665A, R682A, and N684A.

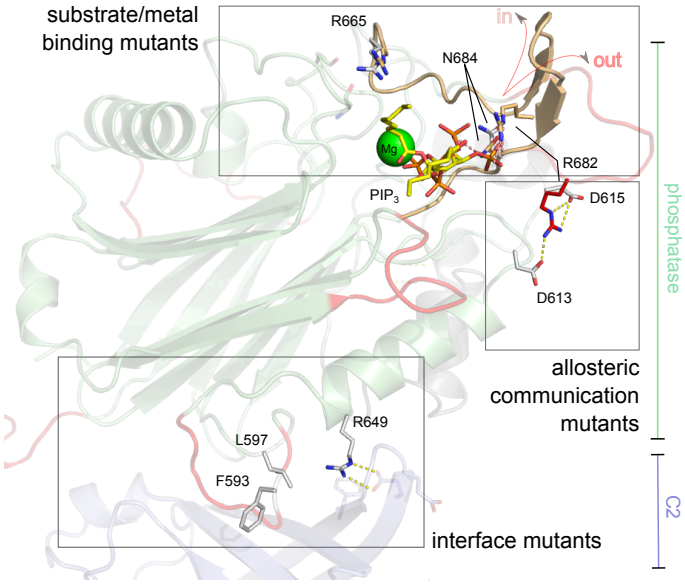


Figure 3.17. Structures of SHIP2 with the two different loop 4 conformations at the top of the phosphatase domain shown in red ("out" conformation) and orange ("in" conformation). Other important loops are coloured in red while the helical part is in grey and the rest of the phosphatase in green. The amino acids that were mutated and the substrate (PIP_3) are shown as sticks.

- ***Interface mutants*:** the hydrophobic mutant F593D+L597D is located in loop 3 of the Ptase domain. Phe593 and Leu597 are at the centre of a hydrophobic core joining the Ptase with the C2 domain. In agreement with our predictions, the Ptase + C2 mutant decreased its activity for both substrates (IP_4 and PIP_3) reaching similar activity levels that Ptase alone (Table 3.1). While for the Ptase mutant in presence of PIP_3, the mutant

increased its activity compared to the wild-type (wt), although we did not observe any change in presence of IP_4. We performed simulations of this mutant and observed a differential behaviour of the mutations on the Ptase versus Ptase + C2 protein (Figure 3.18-a). The mutations reduce the RMSF of loops 2 and 4 for Ptase+C2, but have have no effect on the Ptase alone, indicating that the mutations remove the effect of the C2 domain in enhancing loop 2 and 4 fluctuations, that are likely important for high activity. On the other hand, the mutations stabilize the helical region in the Ptase, which is not affected in Ptase+C2. Although this effect on the Ptase is more difficult to reationalise it could be related to the observed increase in catalytic activity observed for the Ptase mutant.

The mutant R649A at the polar interface causes a slight decrease in activity of Ptase + C2, while for the Ptase mutant there is a considerable increase, suggesting that neutralisation of Arg649 might contribute to the increased activity of Ptase + C2.

- ***Allosteric communication mutant***: To confirm the relationship between the C2 domain and the substrate binding site, we generated D613A/D615A mutant located at the already mentioned helical zone (Figure 3.17). These two residues stablishes polar interactions with Arg682 only when the loop 4 adopts the out-conformation and favouring this state. This mutant could only be generated for Ptase + C2, but was not stable for Ptase only, confirming that the C2 domain is stabilising this helical section. The mutations in Ptase+C2 only affected IP_4 kinetics, but not PIP_3, which is consistent with the view that loop 4 "out" is important for headgroup entry, wherease additional lipid interactions in PIP_3 appear sufficent to place the substrate.

- ***Substrate/metal binding mutants:*** Most of the residues involved in substrate and metal binding are conserved among 5-Ptases (Figure 3.10). However, some critical residues are diverged in the SHIP phosphatases. At the position of Gly434 in SHIP Ptases, there is a Asn or Gln in all other 5-Ptases that is critical for coordinating the Mg ion. According to SHIP2 crystal structures and our simulations, Arg665 could substitute the coordination, however, it is too far for a direct interaction. Our simulations suggest that Arg665 is coordinating the magnesium through a network of water molecules (Figure 3.18-b). However, in our MD simulations, we have observed that this

coordination is maintained for the Ptase domain alone throughout the simulation, whereas in presence of the C2 domain the maintenance of these interactions is not as stable, reaching alternative Arg665 conformations throughout the simulation (Figure 3.18-b). Indeed, the kinetics experiments confirmed the importance of Arg665 in the Ptase alone as its mutant to Ala exhibits significant lower activity, whereas for Ptase + C2 mutant the activity was similar to wt (Table 3.1).

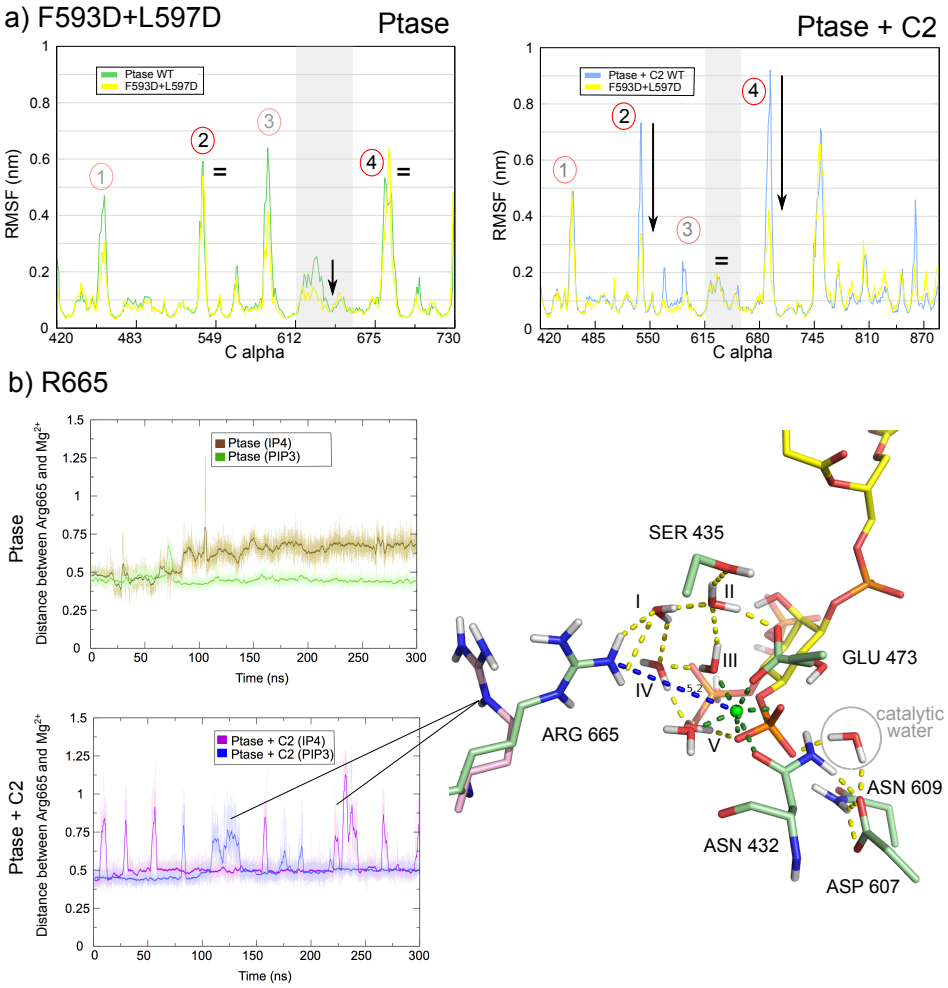


Figure 3.18. a) Comparison of RMSF for wt and F593D+L597D mutant SHIP2. b) The plots on the left hand side show distances between the Arg665 and the Mg^{2+} for the Ptase alone (top part) in presence of IP_4 (brown) or PIP_3 (green), and for Ptase + C2 (bottom part) with IP_4 (pink) and PIP_3 (purple). On the right hand side, a frame extracted from Ptase with PIP_3 shown in yellow sticks is shown. The Mg^{2+} and its coordination are shown in green (sphere and dashed lines respectively). The waters network is marked with roman numbers and its hydrogen bonds shown as dashed yellow lines.

Lastly, we tested the mutants Arg682 and Asn684. As explained earlier the former is within loop 4 and can either be in an "out" conformation, interacting with D613/D615, or can point into the substrate binding site (Figure 17) and we proposed that Arg682 recognizes the 3-phosphate in PIP_3, making it indispensible for the catalytic reaction to take place (Figure 3.19). When we tested the activity of the mutant experimentally, it was strongly reduced for the Ptase and Ptase + C2 and with both substrates (IP_4 and PIP_3, Table 3.1), confirming our predictions. Moreover, Asn684, which as explained earlier we also predict to be indispensable for substrate recognition at the 4-phosphate, showed almost no activity in activity experiments (Table 3.1).

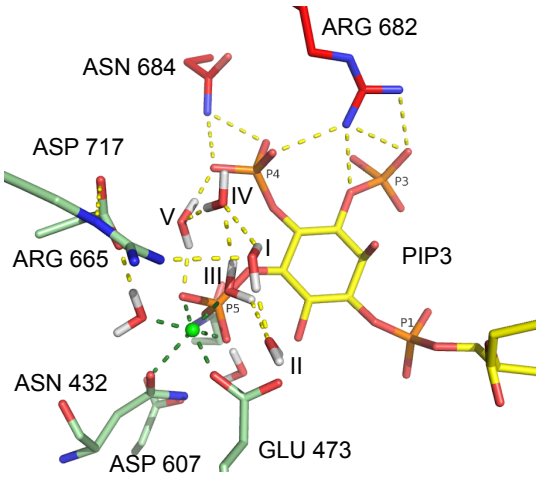


Figure 3.19. The head group environment of PIP_3 (in yellow sticks). Ptase residues and residues belonging to the loop 4 are shown in green and red sticks respectively. The Mg^{2}+ is shown as green sphere, and its coordination as dashed green lines. Water network is shown as roman numbers and dashed lines.

3.1.2.3 Discussion

Many studies have demonstrated the importance of 5-Ptases in a range of human diseases, including diabetes, cancer and Lowe's syndrome. However, many details on molecular mechanisms are still elusive. Integrating crystallographic, biochemical, and computational data from 5-Ptases + nucleases [175], [182], [64], allowed us to generate a model for PIP_3 and IP_4 binding Ptase domain in complex with the C2 domain. These models enable us to interrogate detailed catalytic mechanisms and regulation insights regarding the C2 role.

By applying several computational methodologies and experimental validation, we

propose a new regulatory mechanism involving the C2 domain, proximal to the catalytic 5-Ptase in SHIP2. We showed how this domain can stabilize a helical part in the Ptase domain through direct interactions and that this affects the dynamics of an important loop next to the substrate binding site (loop 4). We show that this loop, which is specific to SHIP Ptases (Figure 3.10), can adopt two different conformations, both important for efficient catalysis, and that the switch between the 2 conformations is favoured in presence of the C2 domain. The "out" conformation allows sufficient space for the substrate to access the binding site, and the "in" conformation is required for substrate binding via the 3-phosphate. We believe that the latter explains why SHIP, in contrast to other 5-Ptases, exhibits a strong preference for the 3-phosphorylated PIP_3 over the $PI(4,5)P_2$ substrate (reference). We further show that the loop movement regulates Asn684 positioning for 4-phosphate binding of the substrate. We experimentally validated the importance of these mechanisms by determining the catalytic activity of relevant SHIP mutants (Table 3.1).

Further studies will be required to determine how additional regions in SHIP2 affect the mechanisms we describe. In summary our data provides important new insights into SHIP2 regulation and catalytic mechanisms that could be utilized to design novel compounds targeting SHIP via orthosteric or allosteric binding sites.

3.2 Activation mechanisms of kinases

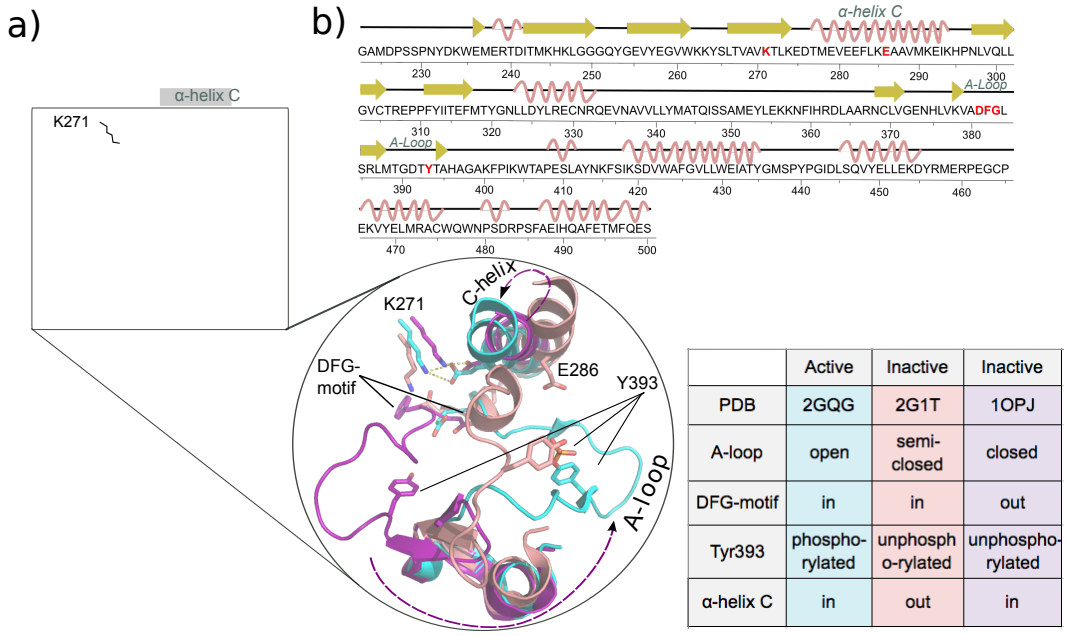
3.2.1 FREE ENERGY LANDSCAPE OF ACTIVATION LOOP CONFORMATIONAL CHANGES IN ABL: FROM THE UMBRELLA SAMPLING TO THE METADYNAMICS VIEW

The structure of Abl kinase domain has been solved by X-ray crystallography in the active and phosphorylated conformation as well as in a number of inactive conformations [37], [183]. Activation of Abl depends on phosphorylation of Tyr393 in the activation loop (A-loop), both in oncogenic forms and in the cellular form of Abl [42]. We reconstructed the corresponding free energy landscape to characterize the dynamics of the A-loop from the closed "inactive" conformation to the open "active" conformation and vice versa, to identify hidden intermediates of the reaction, and to understand how phosphorylation affects the opening/closing transition. To this end, we first applied US (see 1.2.2.6) along a reaction path described by an interpolation between the crystal structures of the open and closed forms. Although this methodology was successful for the phosphorylated protein it was not sufficient for the unphosphorylated form. For this reason, we applied a second approach using metaD together with PTmetaD and WTE (see 1.2.2.6).

The purpose of using both methodologies is that each one allows studying, within their computational limits, different transitions and, in the end, phosphorylated and non-phosphorylated energetic profiles can be characterized. This characterization leads on the appearance of many intermediate states that can be used as new targets for protein-ligand docking in drug discovery.

3.2.1.1 Structural insights

The activity of the Abl kinase domain is regulated by means of a series of conformational changes such as A-loop opening or α-helix C rotation, and through the major phosphorylation site Tyr393 in the A-loop (Figure 3.20-a) [41]. To understand these events, we applied MD simulations (see 1.2.2.5) to obtain atomic and energetic details characterizing the conformational space of the different transition states of the catalytic domain of Abl kinase.



	Active	Inactive	Inactive
PDB	2GQG	2G1T	1OPJ
A-loop	open	semi-closed	closed
DFG-motif	in	in	out
Tyr393	phospho-rylated	unphospho-rylated	unphospho-rylated
α-helix C	in	out	in

Figure 3.20. a) Schematic of the kinase domain and cartoon representation of the α-helix C and A-loop in inactive (purple), inactive Src-like (light pink), and active (cyan) conformations [25]. Residues involved in salt bridge (Lys271 and Glu286) and phosphorylation (Ty393), are highlighted. Important motifs such as the A-loop, the DFG motif and the α-helix C are labelled and summarized in the table. b) Abl amino acid sequence and assigned secondary structure.

To reach the inactive state a number of highly conserved motifs must adopt the autoinhibited conformation. An example of an X-ray structure of this inactive state is 1OPJ.PDB (purple cartoon in Figure 3.20-a), in which Tyr393 is unphosphorylated and

the A-loop is blocking the catalytic site thus preventing substrate binding. In addition, the DFG motif is in an "out" conformation and the phenylalanine (Phe382) occupies the ATP site. Finally the α-helix C is in an intermediate state with Glu286 pointing inside the ATP site. Apart from this inactive conformation, a Src-like inactive conformation was crystallized (2G1T.PDB) [183], providing an additional state where the A-loop is semi-closed and partially folded, the DFG motif in an "in" state but unable to accommodate ATP since the A-loop is occupying the cleft and the α-helix C is rotated outward (light pink cartoon in Figure 3.20).

For the catalytic reaction to take place, the position of the α-helix C has to change so that a salt-bridge between Glu286 and Lys271 can occur. The DFG motif must flip allowing Asp381 to participate in the coordination of the catalytic Mg^{2+}. The A-loop, which is the most flexible part of the catalytic domain, has to be extended to promote substrate binding and tyrosine phosphorylation. This phosphorylation has several consequences in the stabilization of the active conformation not only for Abl but for several other kinases [184], [185]. Even though all of these characteristics are well understood, thanks to a variety of crystal structures, the transition from the inactive to the active state, the conformational changes involved, and the order of events are still unclear.

Using US and metaD calculations, we described in detail the free energy landscape pertaining to conformational changes of the A-loop and characterized the effect of tyrosine phosphorylation on the dynamics of Abl.

3.2.1.2 Strategy to define the reaction coordinate

In principle, a sufficiently long unbiased MD trajectory should sample the complete conformational space of a system and allow reconstructing the underlying free energy landscape. However, our reaction of interest consists of a complex conformational transition, where the different closed and open states are likely to be separated by relatively high free energy barriers that lead to slow and/or infrequent transitions.

Because of this caveat a complete sampling by unbiased MD simulations is unfeasible with ordinary computer resources.

To establish a relevant reaction coordinate that describes the conformational change between the inactive (closed) and active (open) A-loop states [186], we had to consider different issues. On one hand, the selection of the relevant conformations for the description of the mechanism of opening of the A-loop out of the large number of different known conformations, which makes the scenario more complex. We first collected all available structures of the Abl Kinase from the PDB and analysed them by means of structural and multiple sequence alignments [179]. On the basis of the results from this analysis, we extracted 15 structures containing Abl in several distinct conformations. Out of these, we selected the 7 most representative structures, namely PDB entries 1OPL, 1OPK, 1OPJ, 2GQG, 2G1T, 2G2F, and 2F4J, which were individually simulated using standard MD for 100 ns. All the structures remained conformational stable within the expected thermal fluctuations. Thereafter, we used PCA (Figure 3.21) to characterize the conformational space covered by the unbiased MD simulation trajectories. A clear separation between the three main conformations was observed upon projection onto the plane described by the first two PCA eigenvectors: the open conformations were located on the left-hand side of the plot covering the same conformational space (Ev1<1) whereas the only two closed-inactive conformations (PDB: 2G1T and 1OPJ) were located on the right-hand side (Ev1>1), being the 2G1T closer along the first PC to the open-active pull of structures. We finally selected two structures as a representative of the closed and open states, both in DFG-in conformations, namely 2G1T and 2GQG respectively. As a reaction coordinate, we chose the first eigenvector (Ev1*) of a covariance matrix of these two structures. For the closed to open transition, Tyr393 was unphosphorylated while for the reverse reaction coordinate it was phosphorylated.

Figure 3.21 shows the projection of the crystal structures themselves (red diamonds) and projections of the trajectories from the corresponding unbiased MD simulations of 100 ns each. The reaction coordinate chosen for the US is drawn as a dashed red line (Ev1*). The figure shows that both active and inactive conformations were stable and

very distant from each other, which supports our view that exploration of any transition between them using unbiased MD is very likely to be unsuccessful.

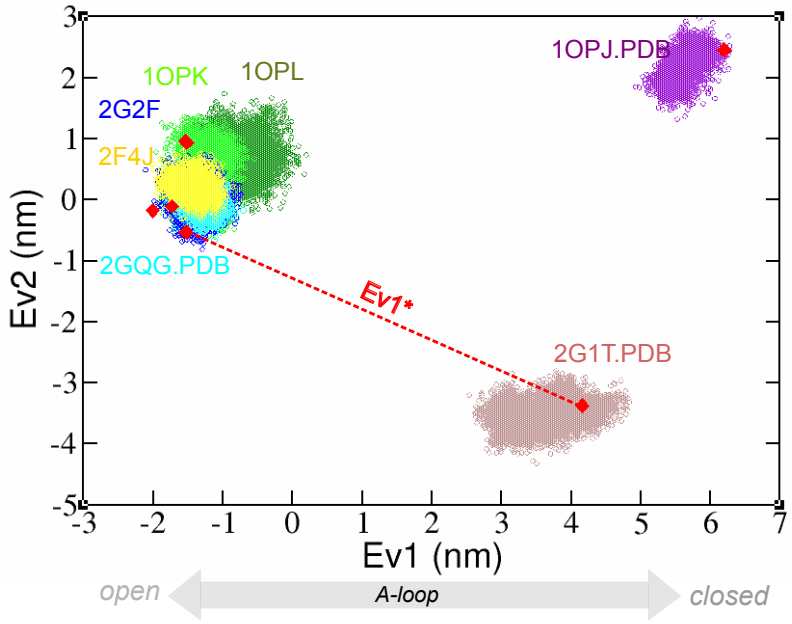


Figure 3.21. Projection onto the plane defined by PCA eigenvectors Ev1 and Ev2 of the crystal structures and snapshots taken from the simulated unbiased MD trajectories. Note that 1OPJ and 2G1T correspond to DFG-out and DFG-in conformations, respectively.

3.2.1.3 Free energy landscape

After the two initial structures were energy minimized and equilibrated for 280 ns at 300 K, we calculated the PMF [16] along the relevant reaction coordinate, Ev1*. A total of 20 ns for each reaction coordinate were analysed for ten different windows equidistant along our Ev1*, with k = 10 kJ/mol·nm^2. The energy profile of the phosphorylated kinase starting from the open-active state and the un-phosphorylated kinase starting from the closed-inactive state are shown in Figure 3.22-a. For the phosphorylated kinase (in blue), we only observed one stable minimum (labelled A' in Figure 3.22-a) at Ev1* = -3.385, in an active conformation with an extended A-loop and Glu286 in α-helix C pointing towards the catalytic site (Figure 3.22-b). For the

unphosphorylated kinase we also reached only one stable minimum (labelled A in Figure 3.22-a) corresponding to the closed-inactive state. However, several local minima appeared (B, C) when two additional windows with a higher *k* were added (windows centred at Ev1*= 0.0 and 0.5 with *k* = 100 kJ/mol·nm^2, histograms are shown in Figure 3.22-c). The B and C states were located at Ev1* = 0.690 and 1.264, respectively, with a PMF = ~ 55 kJ/mol, corresponding to the starting opening of the most hydrophilic part of the A-loop (from Thr389 to Ile403) (B) and the beginning of the helix unfolding from the most hydrophobic part of the loop, from Leu384 to Asp391. As other studies have shown for different kinases [29], the hydrophilic part is the most flexible part of the loop and consequently initiates the opening process. For the opening of the second half of the loop two key changes must take place: (i) unfolding of the helix formed by residues Leu384, Ser385, Arg386 and Leu387, and (ii) a rotation movement of α-helix C (coloured in cyan in Figure 3.22-b) from an outward position with Glu286 pointing outside the ATP site to a position where Glu286 (shown as sticks in Figure 3.22-b) points into the catalytic pocket.

However, with the limitations of our simple and single reaction coordinate, we were not able to see rotation of the α-helix C. In order to facilitate this rotation we included three more windows starting from the opposite site of the Ev1* (Ev1*= 1.5, 2.5 and 3.5, shown as dashed lines in Figure 3.22-c). Although the histograms of the windows overlap quite well and the PMF values were lower (dashes line Figure 3.22-a), the transition did not occur.

The sampling of the unphosphorylated form detected one stable minimum (A) and several transition states (B, and C) whereas in the sampling of the phosphorylated form only one state (A') was observed. Analysis of the different patterns in flexibility revealed the reasons why the phosphorylated system did not explore any other conformation.

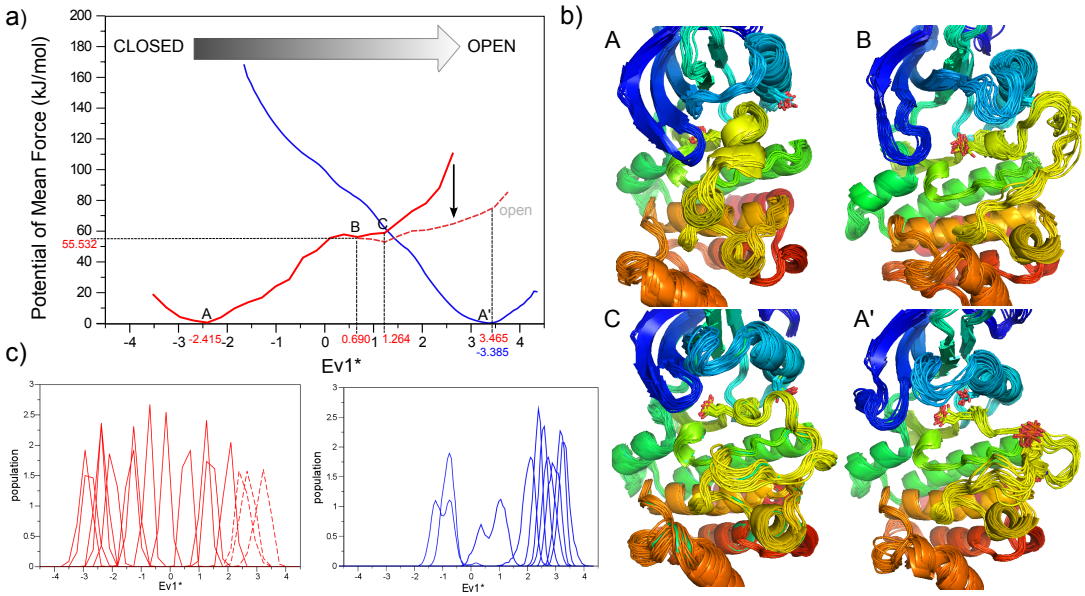


Figure 3.22. a) PMF profiles for the open-to-closed phosphorylated Abl (blue) and closed-to-open for unphosphorylated Abl (red) transitions. The dotted line corresponds to the three extra windows that were necessary to complete the reaction coordinate. b) Structures from the clusters corresponding to states A, B, C and A' in the PMF diagram. c) Histograms of the configurations showing that sufficient overlap between adjacent windows was achieved in both US simulations (same colour code as in a)).

3.2.1.4 Stabilization of the phosphorylated active state

The RMSF were larger in the unphosphorylated kinase than in the phosphorylated one (red and blue, respectively, in Figure 3.23-a), especially along the N-lobe and even more remarkably at the A-loop. In particular, we noted that the flexibility of Tyr393 was drastically reduced upon phosphorylation.

As shown in Figure 3.23-b, when Tyr393 was phosphorylated in an extended conformation of the A-loop, three positively charged residues, namely Arg362, Arg386 and His396, formed a highly stable cluster around the phosphotyrosine side chain. This increased the stability of the interface between the two lobes and consequently of the entire catalytic domain, thus resulting in lower fluctuations of the N-lobe residues 230 to 375. These interactions and stabilization of the phosphorylated Abl kinase provide a rationale for the finding that this state is the most favourable upon phosphorylation.

A recent study [186] has shown a similar pattern for the highly homologous tyrosine kinase Src. Whose active conformation is stabilized upon phosphorylation while the unphosphorylated form displays higher flexibility and also favours some intermediate states. To assess whether the results obtained for Abl depended on the choice of Ev1* we switched to metaD to study the same reaction coordinate.

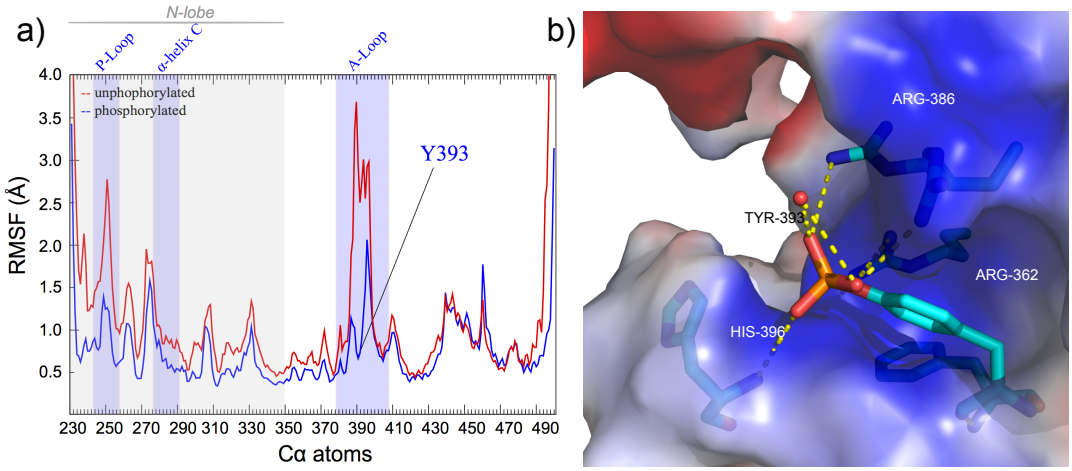


Figure 3.23. (a) Root Mean Square Fluctuations (RMSF) of Abl Cα atoms computed on both trajectories, phosphorylated in blue and unphosphorylated in red. N-lobe marked with grey background and important motifs highlighted in blue. (b) Close-up of the local environment around Tyr393. Neighbouring residues are shown as sticks and an enveloping surface is coloured according to the calculated molecular electrostatic potential (MEP), with blue and red colours representing positive and negative values, respectively.

3.2.1.5 PTmetaD using additional variables

To test whether the opening - closing of the unphosphorylated A-loop was better described by a multidimensional approach, we performed extensive WTE-PTmetaD (described in section 1.2.2.6) along a total of 1500 ns. The energy bias was obtained after 10 ns of WTE of the following seven temperatures (K): 295, 310, 325, 341 358, 376, and 395. For the T=300 the energy bias was null, so any reweighting step was needed. Two collective variables (CVs) were defined for describing the reaction process: CV1 consisting of an s-path [187], the low free energy path obtained along 9 frames extracted from a MD simulation steered from the closed and semi-closed X-ray structures (PDB 1OPJ and 2G1T respectively) to the open X-ray form (PDB 2GQG).

The contact maps [188] for this CV were chosen with cutoff values of 5 Å and 35˚, extracted after minimization and equilibration for at least 5 ns at 310 K. CV2 corresponds to an s-path on the in-out conformations of the DFG motif. By adding this new combination of dimensions we included the rotation of both α-helix C and Glu286 and also the DFG flip (in/out conversion). These features were not included in the US because both structures were in the DFG-in conformation. While in the inactive and closed A-loop the DFG adopts two different positions: “in” (2G1T.PDB) and “out” (1OPJ.PDB), the extended and active state only adopts a DFG-in conformation. To enhance the exploration of the different conformations, in addition to the structures used in US (2GQG and 2G1T), we included a second closed inactive structure, 1OPJ, with the DFG in an “out” conformation.

The free energy surface (FES) of the white replica is shown in Figure 3.24-a as a function of CV1 and CV2, and in Figure 3.24-c as a function of CV1 alone. The simulation explored several different conformations at the current state of the calculation, and the results were comparable to those obtained in the US. However, according to what we anticipated in the US results, we also observed a high stabilization of different intermediate states where the highly flexible A-loop tends to adopt a number of different semi-open conformations rather than the fully closed or open one. As we could observe in the FES (Figure 3.24-a/c) the absolute minima (G and H) along these two CVs corresponded to the semi-open A-loop, with part of the loop folded in a small helix, similar to that observed in the crystal structure with PDB code 2G1T, also known as the Src-like inactive structure. In addition, alternative intermediate states corresponding to the fully open (D, E) and closed conformations (I, J, K) were reached, although higher energy values were needed: 30 kJ/mol and 20 kJ/mol respectively (Figure 3.24-c).

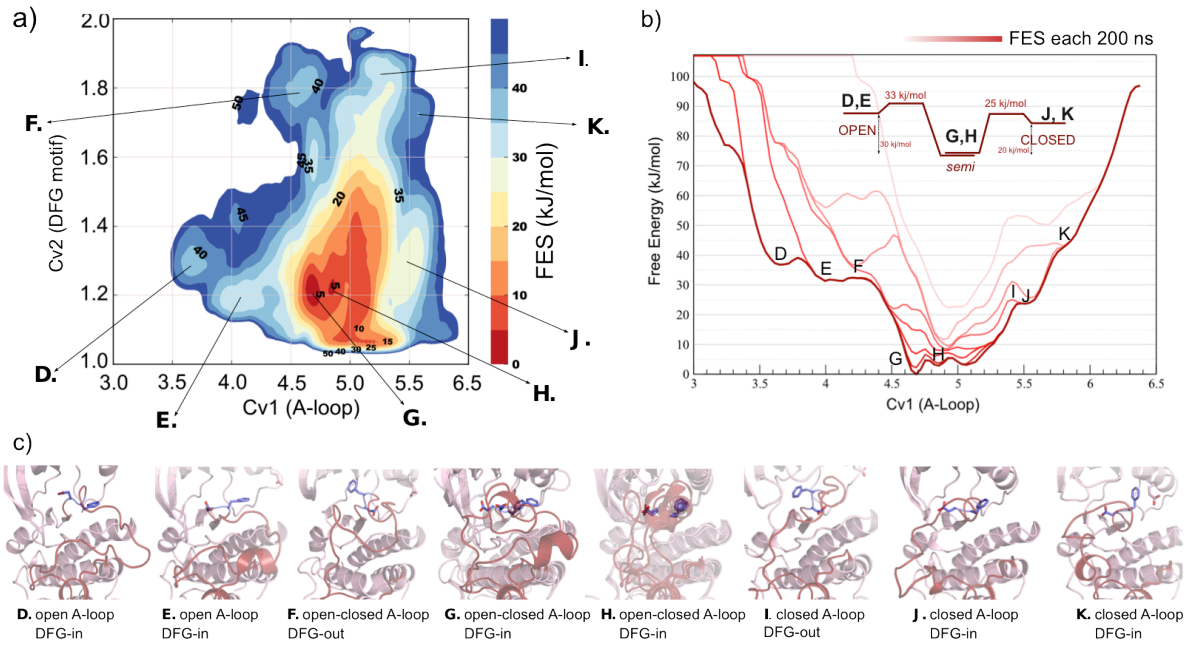


Figure 3.24. a) Free energy surface of unphosphorylated Abl as a function of CV1 (A-loop) and CV2 (DFG motif). The contour lines are drawn every 5 kJ/mol. States corresponding to different minima and plateaus are marked with letters. b) 2D energy profile corresponding to CV1 (A-loop transition). c) Representative structures of states D-K are shown as cartoons.

These results confirm the hypothesis put forward for other kinases [29] that unless Tyr 393 is phosphorylated, the fully open state will not be favoured, and a semi open state will be preferable. Apart from the variety of A-loop conformations that we could observe, we also sampled different DFG conformation. For the DFG-out, we sampled structures along the simulation either with a closed A-loop conformation (I) or with a semi-open A-loop (F), but never in a fully open state. In contrast, for the states corresponding to a DFG-in conformation (D, E, G, H, J, and K) the A-loop showed intermediate open-closed states (G, H), a completely open state (D), or a completely closed form (J, K). This suggests that, in agreement with previous results for other kinases [34], the DFG-in conformation is favoured over DFG-out.

3.2.1.6 Discussion

We have used US and PTmetaD simulations to characterize the conformational stability of the active state of the catalytic domain of Abl kinase when Tyr393 in the A-loop is either unphosphorylated or phosphorylated. The computed free-energy profile shows that, while occurrences of an active-open conformation are allowed when the A-loop is unphosphorylated, the kinase domain does not exclusively remain in a catalytically competent state. In agreement with the recent study for Src kinase [29], when the A-loop shows higher flexibility, the half open - or the so-called Src-like state - is stabilized. Considerable flexibility and fluctuations can occur in the overall neighbourhood of this active-like conformation once the A-loop has been opened but remains unphosphorylated. The A-loop can adopt a wide variety of extended conformations and the A-loop can either accommodate the substrate or trans-phosphorylate Tyr393, although the fully-open state is not completely stabilized unless the tyrosine is phosphorylated. Therefore, the free-energy surface suggests that the kinase domain of Abl could temporarily maintain a semi-open extended conformation for the A-loop regardless of the α-helix C rotation. This could lead to the hypothesis that even rarely visiting these states would be sufficient to allow Tyr393 to become phosphorylated, thereby stabilizing the kinase domain into its active state.

The steep free-energy observed in Figure 3.22-a could explain why phosphorylated Abl kinase only crystallises in the active conformation, as the phosphorylated A-loop is unlikely to return to a conformation with the inactive outward α-helix C. This stabilized phosphorylated conformation is favourable for substrate binding.

These observations are in agreement with previous experimental work showing that the kinase domain of Abl enhances its activity when the A-loop is phosphorylated [189].

3.2.2 FOCAL ADHESION KINASE (FAK) ACTIVATION AND CONFORMATIONAL REARRANGEMENTS

In the paper *"PI(4,5)P2 mediates integrin signalling by triggering activation of Focal Adhesion Kinase via clustering and conformational rearrangements, Goñi, GM.; Epifano, C.; Boskovic, J.; Camacho-Artacho, M.; Zhou, J.; Bronowska, AK.; Martín, MT.; Eck, MJ.; Kremer, L.; Gräter, F.; Gervasio, FL.; Perez-Moreno, M.; Lietha, D. PNAS. 2014" (Annex 1)*, we studied the activation mechanism of FAK using biochemical, structural and cell biology approaches. We showed that FAK in the basal state adopts a closed conformation, which is maintained through two interfaces involving two regions of the FERM domain, namely the F2 and F1 lobes. These lobes are responsible for C-lobe and N-lobe kinase binding, respectively, as observed in the crystal structure [53].

We observed that the fully closed conformation requires ATP bound to the FAK kinase. On the other hand, $PI(4,5)P_2$ binding to the FERM domain induces a relaxed and partially open conformation and this event initiates the activation sequence of FAK. Since $PI(4,5)P_2$ appears to revert the "closing" effect of ATP and therefore the two events appear to be related, our computational goal and contribution to this work was to better understand the effect of ATP. For this purpose we employed MD simulations, NMA and ACA.

3.2.2.1 ATP affects kinase flexibility

In Goñi et al. [21] we observed that ATP and $PI(4,5)P_2$ have opposite conformational effects. While our collaborators in Heidelberg studied the $PI(4,5)P_2$ effect, we investigated the effect of ATP/Mg^{2+} using computational methodologies such as MD simulations and NMA.

The unbiased MD simulations with the FAK kinase domain both in the presence and in the absence of ATP showed different patterns of flexibility especially regarding α helices C and G.

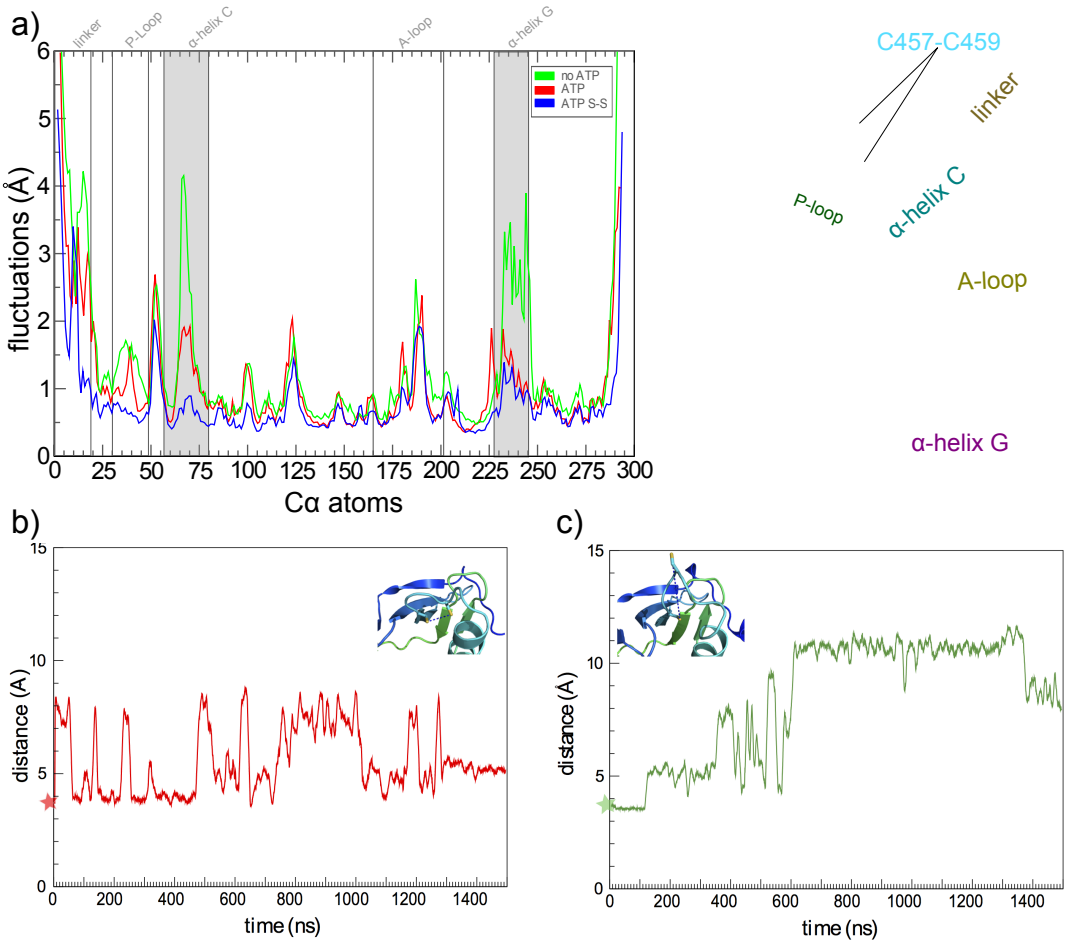


Figure 3.25. a) On the left, the RMSF plot of the three simulations: without ATP (green), with ATP (red) and with ATP and the disulphide bridge between Cys456 and Cys459 formed (blue). On the right, the FAK kinase is shown in cartoon representation; highlighting in different colours are motives relevant for the RMSF. b) Time evolution of the S(Cys456)-S(Cys459) distance in the presence of ATP. c) Same plot as in b) but in the absence of ATP.

We prepared the system using the kinase domain of the catalytic active structure including part of the linker (from residue Thr394 to Ser410) in complex with AMP-PNP (where the imidodiphosphoric acid nitrogen bridging the beta and gamma phosphates was replaced with an oxygen atom), the magnesium ion and the water molecules

involved in metal coordination (extracted from the PDB entry 2J0L). The system was solvated, minimized, and equilibrated by running 1000 steps of steepest descent and 2000 steps of conjugate gradient. In a second stage each minimized system was heated from 0 K to 300 K using an NVT ensemble for 100 ps. Following this, 100 ns of equilibration were carried out using the NPT ensemble at 1 bar. Subsequently, we performed standard MD simulations for 1.5 μs in the presence and in the absence of a bound ATP molecule. The simulations were analysed for differences in backbone flexibility by plotting RMSF values for each residue. We found that the presence of ATP in the active site had the largest stabilizing effect on the α-helices C and G of the FAK kinase domain, reducing the RMSF values for both helices (highlighted in grey in Figure 3.25-a).

These two sites exactly map the autoinhibitory interaction sites with the FERM domain (see Figure 3.26). In agreement with experimental data, we therefore propose that binding of ATP to the FAK kinase induces a tightly closed FERM-kinase conformation by rigidifying the interaction interfaces. Furthermore, we notice that in the presence of ATP not only was the αC helix rigidified but also the loop entering it. This loop contains two cysteines, Cys456 and Cys459, which in the presence of ATP are for most part of the simulation time in a favourable position for forming a disulfide bridge (~4-5 Å, Figure 3.25-b in red), whereas in the absence of bound ATP they equilibrate into an alternative conformation that is not compatible with disulfide formation (>10 Å apart, Figure 3.25-c in green). Although in the 2J0L crystal structure (shown in cartoon in Figure 3.25) the disulfide bond was not formed, likely due to the presence of high concentrations of reducing agent in the crystallisation condition, there are other crystal structures of the FAK kinase where it is formed: PDB codes: 3BZ3, 2IJM, 2ETM, 1MP8, 2JKK, 2JKM, 2JKQ, 4K9Y, and 4EBW.

In order to reveal whether the disulfide bond would affect the protein flexibility, we simulated the kinase in the presence of ATP and with the disulfide bond formed (Figure3.25-a blue line). We observed that in presence of the bond the majority of the

domain is stabilized, most notably the α-helix C, indicating that formation of the disulfide bond could further stabilise the closed autoinhibited conformation.

Next, we analysed possible conformational changes using two different computational approaches: NMA (see 1.2.6) and ACA (explained in section 3.1.2.2).

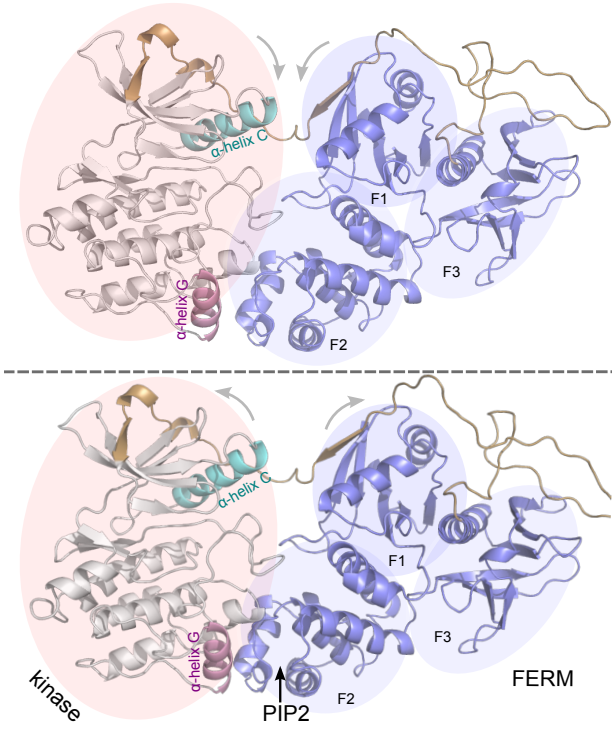


Figure 3.26. First normal mode from NMA. Highlighted in red is the kinase and in blue the FERM domain. The direction of the motion is represented by the two arrows on top of each snapshot showing the "closed" conformation (top) and the "partial-open" conformation (bottom).

NMA calculations were performed using the FAK-FERM crystal structure (PDB code: 2J0J). Because the X-ray structure lacks part of the linker and the A-loop, we modelled those parts using the ModLoop server [190]. The resulting model was minimized and the first 25 normal modes of the protein were obtained using the NOMAD-ref server [191]. The first 6 modes correspond to the three translational and three rotational degrees of freedom for the whole molecule. The seventh eigenvector was the lowest frequency mode, and it is most likely to be functionally relevant [192]. This motion

corresponds to a partial opening of the top part of the protein (Figure 3.26), which is in agreement with the partial opening in the absence of ATP or the presence of PI(4,5)P_2 observed experimentally in Goñi et al. [21].

Moreover, by performing an ACA of the FAK kinase domain we found a strong allosteric coupling between helices C and G despite their distal locations (Figure 3.27). Additionally, the helix E located below the ATP site and helix EF at the end terminal part of the A-loop, and the linker located at the top of the N-lobe also exhibited strong coupling.

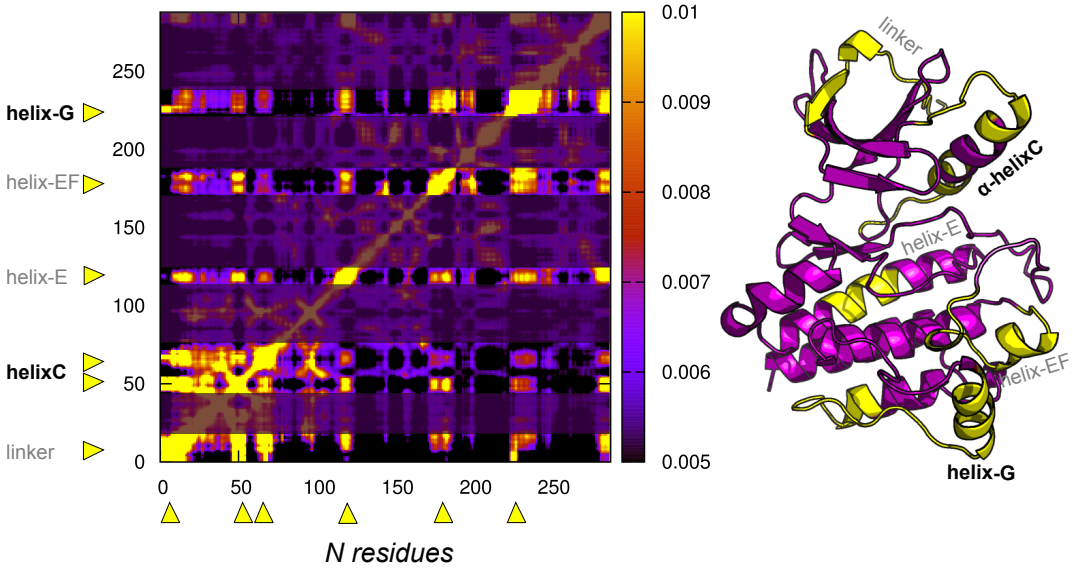


Figure 3.27. Allosteric coupling matrix showing in yellow the regions of the protein that are highly coupled. To emphasise the relevant parts we have shaded out the rest of the plot. Coupling far from the diagonal occurs allosterically and is therefore more relevant.

3.2.2.2 Discussion

Despite the fact that the FAK autoinhibition mechanism and the players involved in regulation have largely been described, the activation mechanism was not well understood. In this study we addressed this issue and provided new useful insights that were included in the joint experimental and computational study published in *Goñi, GM. et al (Annex 1)*.

The previously reported crystal structure of the FERM and kinase region of FAK revealed how the FERM domain docks onto the kinase domain to induce autoinhibition [53]. Additionally, $PI(4,5)P_2$ was identified as an important activator of FAK [193]. The multidisciplinary studies reported in Goni et al [21] has led together with previous studies to the description of a multistep activation sequence for FAK (Figure from Annex 1 included): Firstly, we showed that $PI(4,5)P_2$ production in focal adhesions by PIP5KIγ is required for FAK activation. Further, we demonstrated that $PI(4,5)P_2$ interacts specifically with a basic region in the FERM domain of FAK. Electron microscopy studies revealed that, surprisingly, binding to $PI(4,5)P_2$ induces the formation of large FAK clusters. Using a purified conformational FRET sensor in combination with MD studies we demonstrated that in these clusters FAK undergoes domain opening, which allows efficient autophosphorylation and recruitment of the Src kinase. Src, in turn, phosphorylates the FAK kinase, resulting in full FERM-release and catalytic activation.

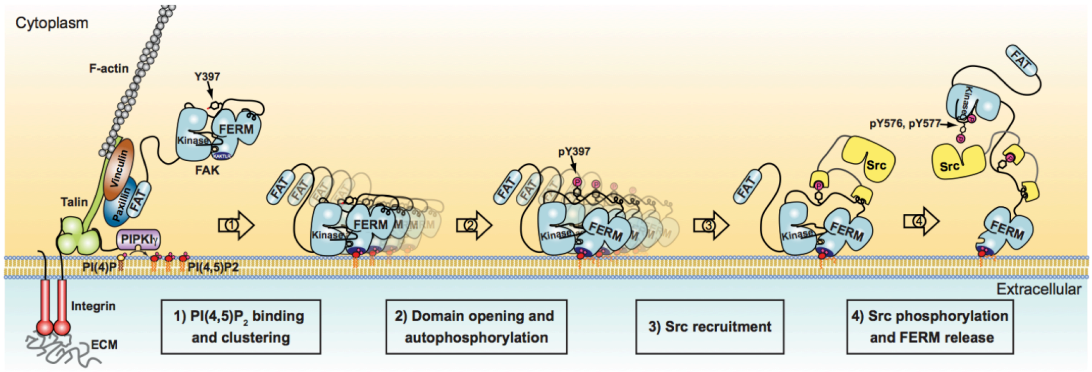


Figure from Annex 1. Multistep activation sequence initiated by $PI(4,5)P_2$ generated in integrin adhesion structures. The model is based on work carried out in *Goñi, GM et al. (Annex 1).*

The contribution described in this thesis has been to explain how ATP binding to the kinase stabilizes two key regions, helices C and G, which directly interact with the regulatory FERM domain in the autoinhibited form. This mechanism suggests new strategies for allosteric drug discovery involving these regions.

3.3 Inhibiting Focal Adhesion Kinase

We pursued three different computational strategies: The first strategy is related to the design of pyrimidine derivatives that bind to the active site, competing with ATP; the so-called *type I* inhibitors. The second one is to design compounds, known as *type II* inhibitors, that bind also to the catalytic site but when the protein is in an alternative inactive conformation (i.e. DFG-out). Finally, the third strategy aims to discover allosteric compounds (binding outside the ATP site) to inhibit FAK signalling. We pursue this third goal using two different approaches: (i) designing optimised analogues of current hits experimentally identified by fragment screening using ^{19}F-NMR [194] and/or Surface Plasmon Resonance (SPR) [195], and (ii) performing a combined process of MD simulations and large-scale VS to identify novel scaffolds interacting with (transient) allosteric pockets.

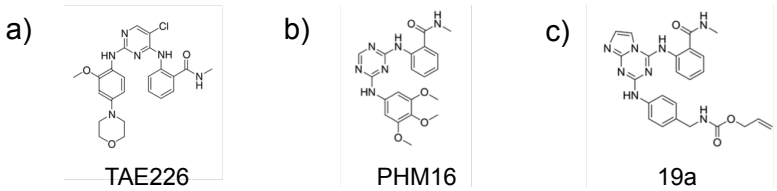
3.3.1 STRATEGY 1: TYPE I INHIBITORS

In 2008 it was shown that TAE226, a compound developed by Novartis, not only inhibited FAK (Figure 3.28-a) very efficiently (IC_{50} = 5.5 nM) [53] but also bound Insulin Receptor (IR) (IC_{50}=2.6 nM) and Insulin-like growth factor I receptor (IGFIR) (IC_{50}=160 nM) with high affinity. Due to resulting toxicity, further development of TAE226 was stopped. In collaboration with *Laboratorie de Chimie et Biochimie (Université Paris Descartes),* we explored the possibility of replacing the pyrimidine ring in TAE226 with a triazine. In addition to potentially reducing off-target activity, triazine compounds have also the benefit of being less protected by patents. Following this research line, compound PHM16 (Figure 3.28-b) was developed and shown to have very poor affinity for IR (IC_{50} >> 1 μM) [196], while retaining considerable affinity for FAK (IC_{50} ≈ 1 μM). In addition to FAK, PHM16 also targets FGFR-2 (IC_{50} = 0.37 μM). This combined inhibition of FAK and FGFR-2 suggested an improvement of the anti-tumour efficacy by improved anti-angiogenic effects [196]. However, although the selectivity against IR was improved, there was still no selectivity against IGFIR (IC_{50} ≈ 1 μM). Further, the affinity of PHM16 for FAK was significantly reduced compared to TAE226. As shown by a crystal structure solved in our group, the binding mode of PHM16 is almost identical to that of TAE226 (Figure 3.28). However, the reduced

affinity is likely due to the missing chlorine atom in PHM16 and because the hydrogen bond made by the triazine nitrogen is weaker than that made by the pyrimidine in TAE226 (Figure 3.28-b). In an attempt to enhance the binding affinities and selectivity towards FAK, an imidazotriazine series of compounds was developed by fusing an imidazole ring with the triazine ring. As the X-ray structures for TAE226 and PHM16 were already solved (Figure 3.28-a/b) and imidazotriazine compounds were expected to interact in similar modes, we analysed possible interactions and binding modes of imidazotriazine compounds using docking tools.

Inhibition assays indicated that the most potent compound of the imidazotriazines series contained the imidazo[1,2-*a*][1,3,5]triazine scaffold (compound **19a**), which showed very good inhibition for FAK (IC_{50} = 0.05 µM) and no effect on either IR or IGFIR (IC_{50} >> 1 µM) [42]. Its regioisomer **19b** showed lower affinity for FAK (0.88 µM). To understand the differences in affinity of such similar compounds we proceeded to analyse the binding modes of both ligands using Autodock 4.2 [197].

(Results publised in *Annex 2*: *Dao, P.; Smith, N.; Tomkiewicz-Raulet, C.; Yen-Pon, E.;* ***Camacho-Artacho, M****.; Lietha, D.; Herbeuval, JP.; Coumoul, X.; Garbay, C.; Chen, H. 2014. Design, Synthesis and evaluation of novel imidazo[1,2-a][1,3,5]triazines and their derivatives as Focal adhesion kinase inhibitors with antitumor activity. J. Med. Chem. 2014.*)



IC_{50} (µM/L)	TAE226	PHM16	19a
FAK	0.0055	~ 1	0.05
IGFIR	0.16	~ 1	>>1
IR	0.0026	>>1	>>1

Figure 3.28. FAK type I inhibitors: a) in yellow sticks compound TAE226 (PDB code: 2JKK), b) PHM16 in cyan (PDB code: 4C7T), c) **19a** in green sticks. Table. IC_{50} values for all the three compounds: TAE226, PHM16 and **19a** towards FAK, IGFIR and IR.

The FAK target structure was extracted from the X-ray structures of the kinase in complex with TAE226 and PHM16 (PDB codes: 2JKK [53] and 4C7T [196], respectively). In both structures the DFG motif adopts a similar helical conformation (Figure 3.28-a/b) and the common scaffold of TAE226 and PHM16 bound identically. However, the scenario changes for **19a** and **19b** (Figure 3.29). After docking both compounds it was very clear that the two ligands may bind using different poses, which likely explains their differences in binding affinities.

For **19a** (Figure 3.29 in green sticks), the triazine ring, the two aniline rings (A and B) and the carboxamide moiety were placed in very similar positions as for PHM16 [196] bound to FAK. The two hydrogen bonds to the kinase hinge region made by PHM16 with its [1,3,5]triazine ring and the aniline moiety (ring B) were conserved in **19a**. However, additional interactions were made by the imidazo[1,2-*a*] ring in **19a**, most notably with Met499 (Figure 3.29-A). The CO of the carboxamide group of **19a** was located near the DFG motif (Asp564-Phe565-Gly566) of the activation loop of the kinase domain and formed a hydrogen bond with the backbone nitrogen of Asp564 of the DFG motif. This hydrogen bond, together with hydrophobic contacts between ring A and Leu567, stabilized the short helical conformation of residues 565-568 and the typical DFG conformation also seen in the PHM16- and TAE226-bound structures, with the side chain of Asp564 pointing up towards the kinase N-lobe. The CO of the alloc group in **19a** formed an additional interaction with Glu506, which together with the interactions made by the imidazo[1,2-*a*] ring may contribute to the increased binding affinity for FAK, compared to PHM16.

On the contrary, **19b** cannot bind in the same mode as **19a**, because the imidazole ring would be too close to Glu500 in the hinge region, leading to steric clashes with this residue (Figure 3.29 in orange, surface shown).

As a result, the imidazo and triazine rings are pushed further out of the pocket, preventing hydrogen bonds between **19b** and the kinase hinge; the hydrogen bonds that are possibly formed are much more solvent exposed and therefore provide a lower binding energy. This likely explains the lower affinity of **19b** as compared to **19a**. As shown in Figure 3.29-E and F, **19a** but not **19b** fitted well into the pocket.

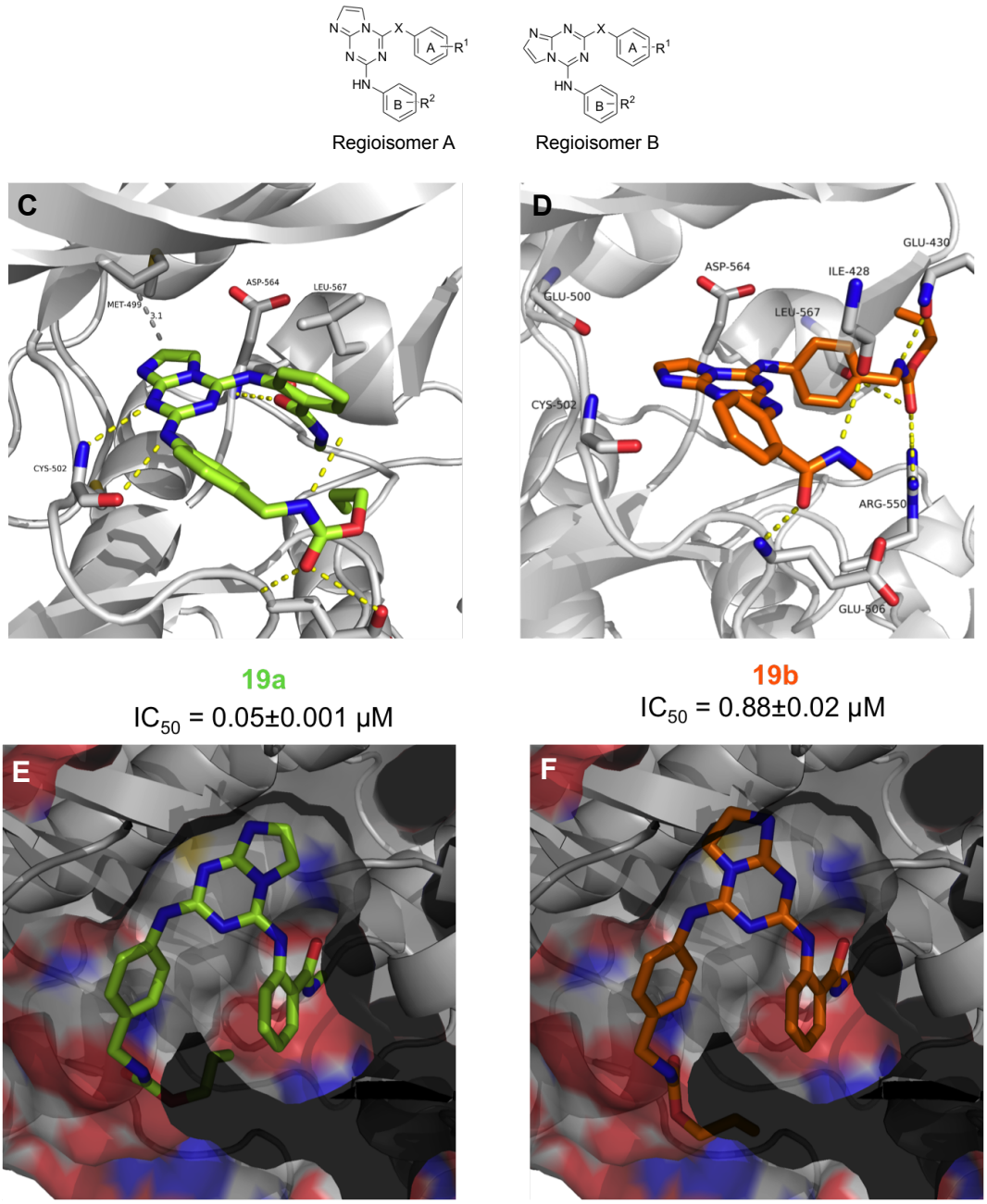


Figure 3.29. Docking of **19a-b** into the ATP binding pocket of FAK. C-F) The lowest energy binding modes obtained by molecular docking are shown for compound **19a** in light green (C-E) and **19b** in orange (D-F). Hydrogen bond interactions between receptor and ligands are shown as yellow dashes lines. A slice through the FAK kinase above the ligand-binding site with the protein surface shown semi-transparent indicates a good fit for **19a** into the ATP binding pocket. The **19b** compound is positioned in a hypothetical position that corresponds to the **19a** binding mode. Clashes of the imidazo ring in **19b** with Glu500 in the kinase hinge region prevent this binding mode for **19b**.

3.3.2 STRATEGY 2: TYPE II INHIBITORS

3.3.2.1 Rational Drug Design

As a template we used inhibitors developed for a very close homologue of FAK, the proline-rich tyrosine kinase PYK2 [19]. It was shown recently [198] that inhibition of FAK could be compensated by Pyk2, illustrating the importance of inhibiting both targets. These two FAK proteins share very high sequence identity, especially in the kinase domain, with 60% of sequence identity [19] and a 45% within the whole molecule. At the outset of this project there were only crystal structures of FAK in complex with classical type I inhibitors, and with the DFG motif in either an "in" position or a helical position as previously mentioned [199]. For Pyk2, however, Han S. et al [200] found Pyk2 to adopt a DFG-out conformation when bound to the low-affinity compounds BIRB796 or PF-4618433 (Figure 3.30). These new diarylurea inhibitors were the first compounds found to induce or stabilize a DFG-out conformation for this tyrosine kinase family.

We designed similar DFG-out compounds for FAK by: (i) modelling a DFG-out conformation for FAK (segment Asp594 to Arg569) based on the segment Asp567 to Arg572 from Pyk2 (PDB codes 2JKK and 3FZT respectively) [201]. (ii) The fourteen missing loop residues (from Tyr570 to Leu584) were built with ModLoop [190] and then minimized, so no clashes or structure artefacts were present. We then fitted two Pyk2 inhibitors PF-431396 (type I inhibitor, pdb:3FZR) and PF-4618433 (type II, pdb: 3FZT) according to their crystal structures into the modelled DFG-out pocket of FAK (Figure 3.32-a). Affinity maps were then calculated using cGRILL (see 1.3.3.1), (Figure 3.32-b) confirming a continuity in the binding pocket starting in one extreme of the pocket, where the PF-

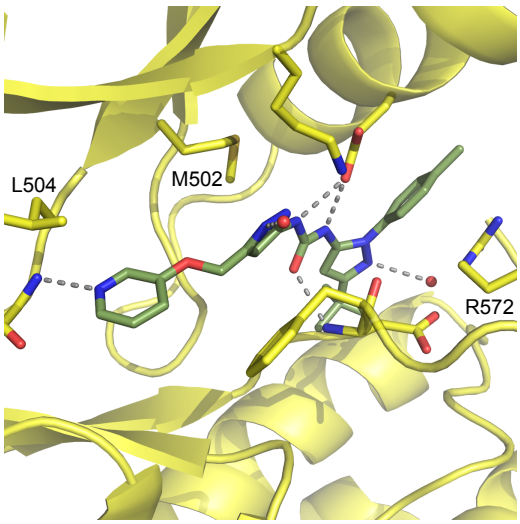


Figure 3.30. Pyk2 structure in yellow in complex with PF-4618433 in green (PDB: 3FZT).

4618433 is located (on the right hand side in Figure 3.32-c) and finishing at the opposite side at the position of PF-431396 (left side in Figure 3.32-c). Moreover, the good fitting between the functional groups of the two molecules and the affinity maps supported our design strategy of merging both molecules into one, thus creating a hybrid molecule (Figure 3.31). This hybrid consisted of the pyrazole ring and the diarylurea group from PF-4618433, so as to form the corresponding hydrogen bonds with the catalytic Lys and Glu, and the *p*-tolyl group, which can make a stacking interaction with Arg572 and van der Waals interactions with the bottom part of the α-helix C. On the other side, the pyrimidine ring and the oxindole ring from PF-431396 were conserved to establish the corresponding interactions with the hinge in the adenine-binding pocket. The final design was a relatively large compound that almost covered the whole interface between the C-lobe and the N-lobe (Figure 3.32-d/e).

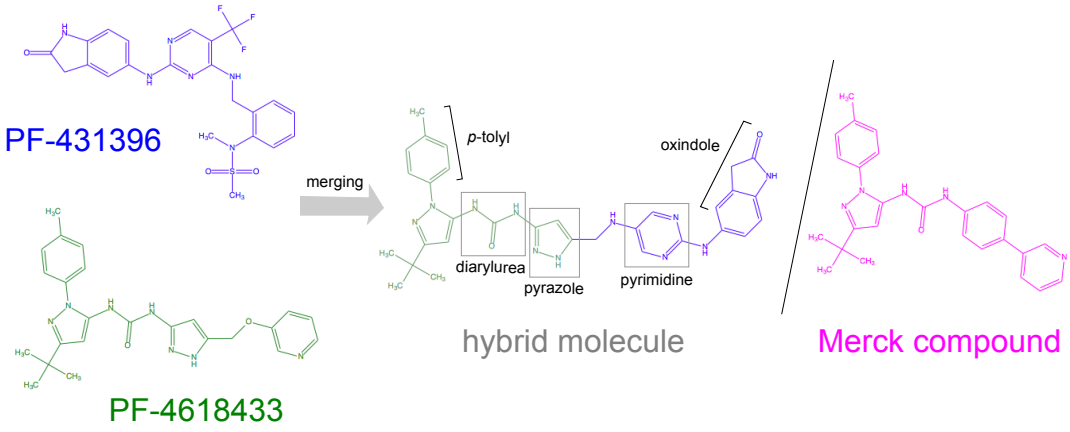


Figure 3.31. 2D structures representing the strategy for the hybrid generation. At the top left hand side, the type I inhibitor PF-431396 in blue, and below, in green, the type II inhibitor PF-4618433. The hybrid molecule with the important chemistries highlighted, and the Merck compound shown in magenta.

Unfortunately, while the hybrid compound was being synthesized by our collaborators, a very similar compound targeting FAK in a DFG-out conformation and its crystal structure bound to FAK was published by Merck [202]. The Merck compound, which was discovered by means of an SPR screening with an in-house library, exhibits submicromolar potency (IC_{50}=266 nM) for FAK and similar potency for PYK2 (IC_{50} = 414 nM). In Figure 3.32-f the crystal structure obtained by Gradler et al. is shown in magenta superimposed onto our predicted model in grey.

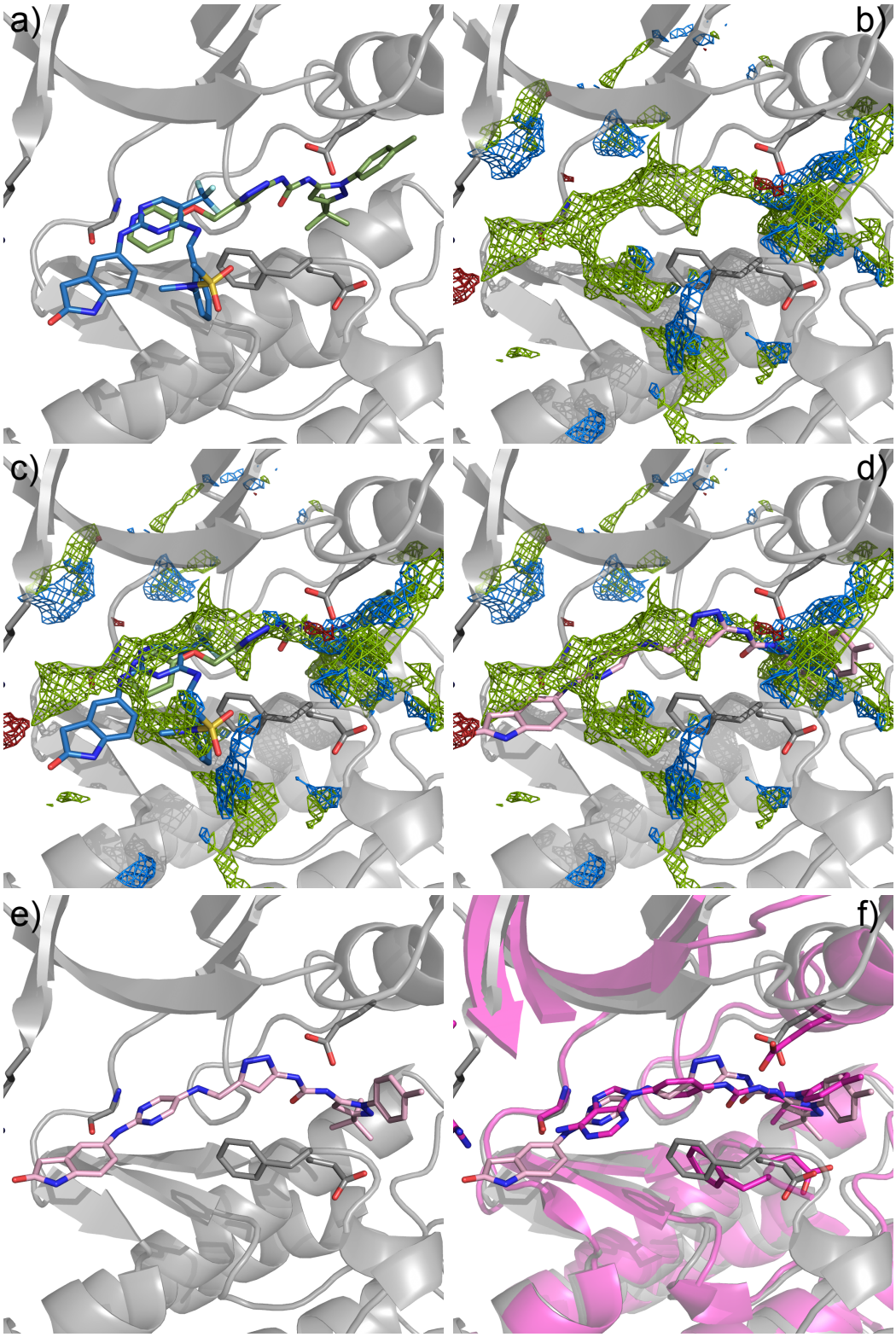


Figure 3.32. a) Superposition of the PYK2 type I (in blue) and type II inhibitor (in light green) on the FAK homology model (grey cartoon). The DFG residues D564 and F565 (in stick representation) are modelled to adopt a DFG out conformation. b) cGRILL affinity maps: green (hydrophobic map), red (H-bond acceptor) and blue (H-bond donor). c) Superposition of (a) and (b). d-e) Type II hybrid compound in pink matching the cGRILL affinity maps (d). f) Superposition of the recently published Merck compound in magenta and our type II hybrid compound in light pink.

3.3.2.2 Virtual screening using the DFG-out conformation

In an alternative strategy, we performed VS of a library extracted from the ZINC database (http://zinc.docking.org) containing ~5,000,000 commercially available molecules. We used the program Glide [153] and the FAK DFG-out model as the rigid target centring the grid box at the DFG motif. The 500 best scored molecules were then selected for visual inspection and consideration of polar and hydrophobic interactions, solvent effects, "forbidden regions" (when a scaffold is occupying a particular direction that leads to steric clashes with the putative boundaries of the receptor), etc.

Out of these 500 molecules, we selected the three compounds shown in Figure 3.33 to be purchased and subjected to crystallography directly. Although we usually performed SPR analysis to check the binding affinity, in this case it was not possible because of the sticky properties of the compounds that could damage the SPR chip.

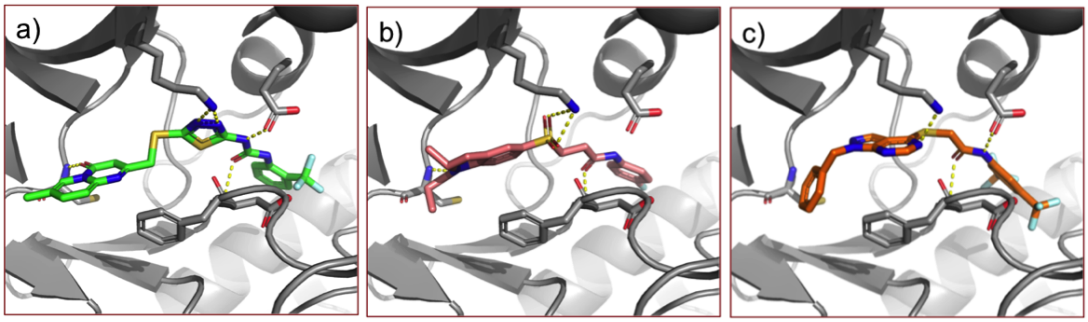


Figure 3.33. The three compounds selected after VS. The FAK model is shown in grey while the putative ligands are shown as sticks with C atoms in different colours. Polar interactions are depicted as yellow dashed lines.

3.3.3 STRATEGY 3: TYPE III - ALLOSTERIC INHIBITORS

For this strategy we focused mainly on a fragment-based approach, although in some particular cases we used a library containing larger drug-like compounds from our *in-house* compound library.

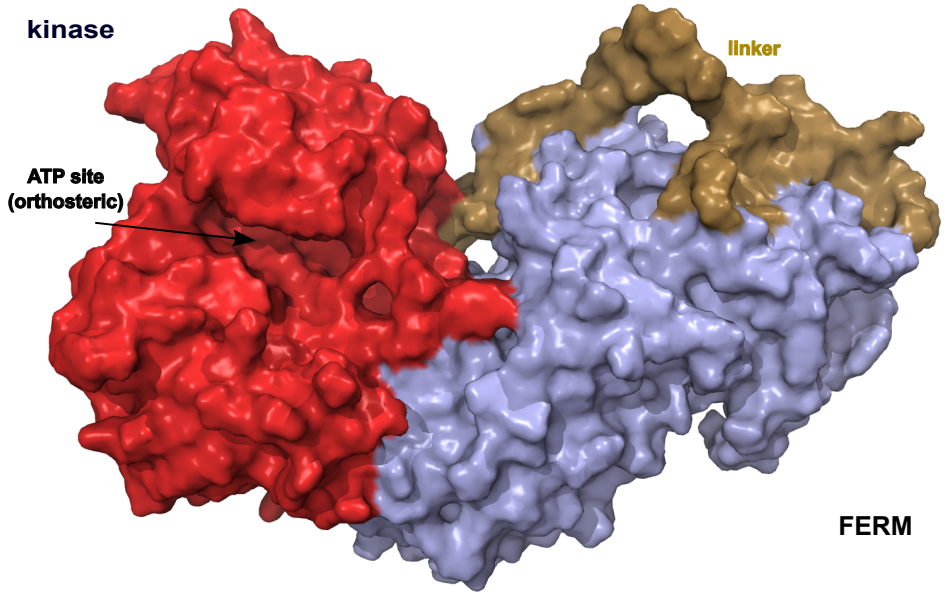


Figure 3.34. Van der Waals surface of the FAK-FERM complex in the autoinhibited form (2J0J.PDB). The kinase, FERM and linker domains are coloured in red, blue and brown, respectively.

3.3.3.1 Optimization of hits from experimental screening

We used NMR-based FBLD to identify fragments targeting the kinase or FERM domains of FAK [202]. By using fragment-based approaches we were able to sample higher chemical diversity with fewer compounds and also to target smaller allosteric pockets while retaining high ligand efficiency. As an example, in Figure 3.34, we can visualize how the surface is full of small cavities and crevices around the surface, some of which can be considered targetable.

To this end, we performed fragment screening by ^{19}F-NMR [203] to detect low-affinity fragments in the micromolar range using an in-house library of fluorinated compounds. Several hit fragments (17 out of 371) were taken for further analysis including validation by SPR [195] and structural characterisation by X-ray crystallography. The

most promising hit "**hit 17**" exhibited a dissociation constant (K_d) of 513 µM. The crystal structure solved in complex with the FERM domain revealed that the compound bound at the F2 lobe of the regulatory domain at the interface with the kinase domain in the autoinhibited FAK conformation [53]. However, due to the solvent exposed nature of the pocket and because of the ligand flexibility, it was difficult to determine the exact pose of the ligand on the basis of the electron density maps. We therefore docked the compound using Autodock 4.2 in order to propose possible binding poses. The two best-scoring poses (green and magenta in Figure 3.35) fitted well into the electron density maps.

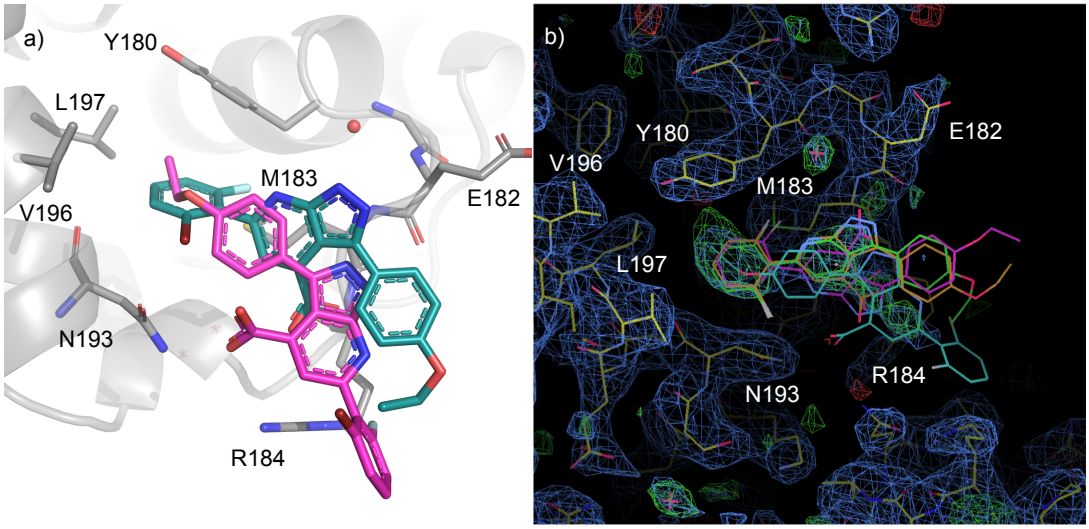


Figure 3.35. a) The two top scoring docking poses for **hit 17** from autodock are shown in green and magenta. b) Several different **hit 17** orientations fitting the electron density maps are shown. 2Fo-Fc maps are shown in blue, difference Fo-Fc maps in green (for positive values) and red (negative values). Both maps are countered at 1σ (image courtesy of Dr. D. Balzano).

We further characterized the binding site using cGRILL to obtain the affinity maps and the putative interaction hotspots (see 1.3.3.1). As we can see in Figure 3.36-a, a hydrophobic region deep within the binding site (highlighted with a white sphere), which is not fully occupied by either pose of **hit 17**, can accommodate a hydrophobic moiety; this possibility will be discussed below. The two red spheres marked with A and B in Figure 3.36-a indicate that this position is favourable for the carboxylic group of the ligand, in agreement with the magenta pose, whereas the lipophilic and hydroxyl affinity map (coloured green and purple respectively in Figure 3.36-b) close to Glu182

would support the green pose better. Indeed, it is possible that the presence of both orientations could result in the poor electron density observed.

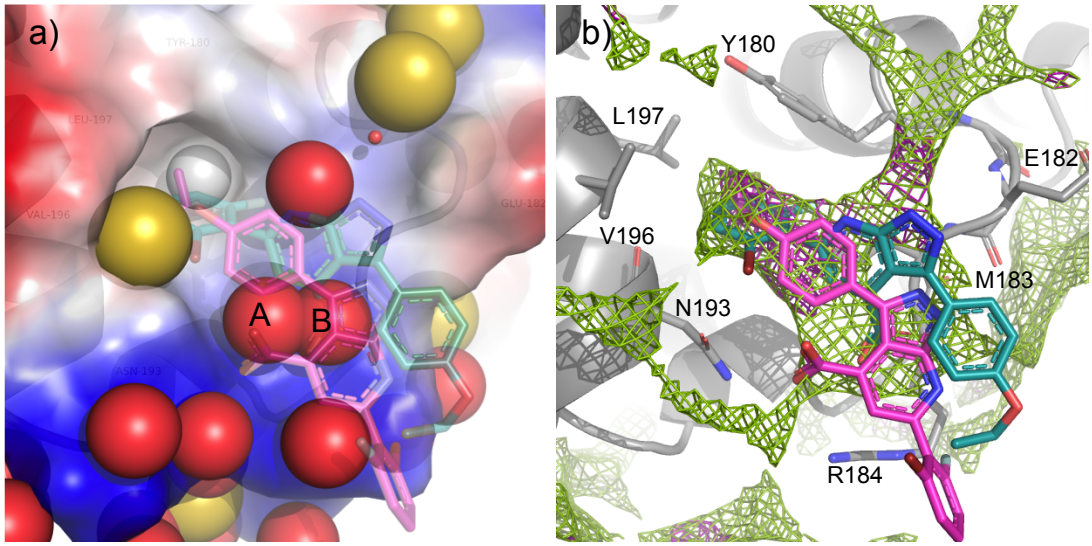


Figure 3.36. cGRILL analysis of the pocket in the FERM domain harbouring **hit 17**. a) Docked possibilities of **hit 17** are shown in green and magenta sticks. The electrostatic potential is colour-coded on the molecular surface and the hotspots are represented as spheres of different colours (positively charged hydrogen-bond donating nitrogens in blue spheres, neutral hydrogen-bond accepting oxygens in red spheres, mixed hydrogen-bond accepting and donating hydroxyls in yellow and hydrophobic areas as white spheres). The water molecule is represented as a small sphere in red. b) cGRILL affinity maps shown as meshes in different colours (lipophilic in green, hydrogen bond donor in blue, hydrogen bond acceptor in red, and hydroxyl and hydrophobic-like in purple and cyan respectively). The side chains of the residues making up the pocket are shown as sticks.

Having identified two likely poses of **hit 17**, we next proceeded to utilize the structural information to design analogues for both poses with the aims to discern the favoured binding pose and to increase the binding affinity. To do so, we treated each pose separately:

(i) Pose 1 (turquoise in Figure 3.37-a) has the 1-bromo-5-fluoro-benzene ring inside the hydrophobic cavity, forming van der Waals interactions with Met183, Val193 and Leu197, and π-π interactions with Tyr180. Our strategy was to divide the molecule into three different parts to increase the interactions with the pocket: the bromo-fluoro-benzene ring, the pyrazolpyrimidine and the carboxylic group. In case of the benzene group we created analogues by replacing the Br with either an amine or iodine to

enhance the interaction with the Asn193, or adding a chlorine atom in –para position to interact deeper into the pocket (Figure 3.37-c/d). For the pyrazolpyrimidine modifications, we designed analogues by adding hydrophobic (cyclopentane) or hydrophilic groups (*tert*-butyl-ether and hydroxyl) to displace the crystallographic water or to replace it (Figure 3.37-e-g). Finally, the carboxylic group in the pyrimidine ring was replaced with a carboxamide to form interactions with Asn193 and Arg184 (Figure 3.37-h).

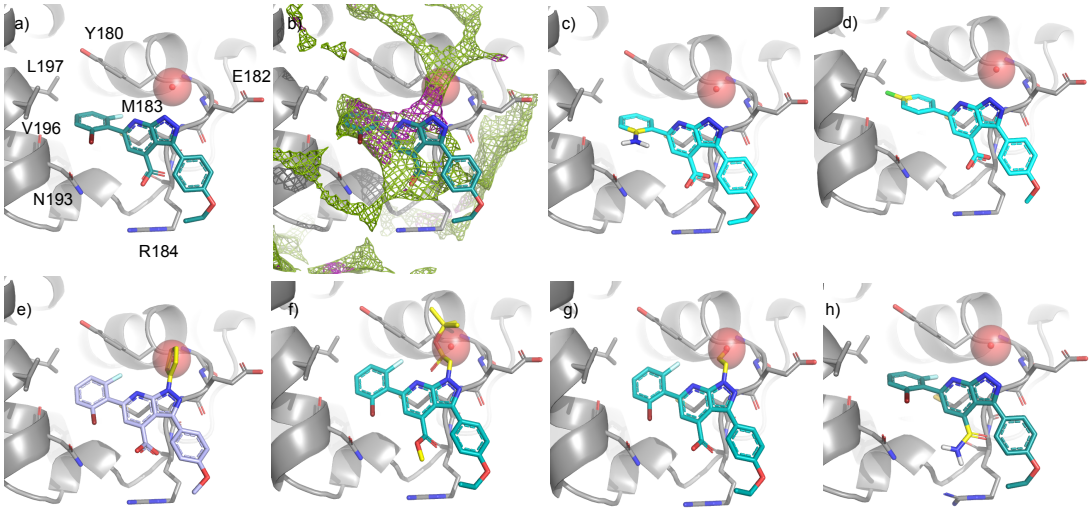


Figure 3.37. a) Starting pose 1 from docking results in turquoise with interacting residues labelled and the crystallographic water shown as a red sphere. b) cGRILL maps (lipophilic in green and hydrophobic in purple) superimposed on pose 1. c-h) Analogues of Pose 1 in blue. The protein is shown in grey while the different analogues are coloured in different blue tones with modifications highlighted in yellow.

(ii) In pose 2 (magenta in Figure 3.38-a) the phenylethoxy group is located inside the pocket and the carboxylic group forms interactions with Asn193. For modifications of this orientation we focused on the ethoxy moiety, which we replaced by an ether-isopropyl, isopropyl or trifluoromethyl group to fill better the hydrophobic cavity (Figure 3.38-c, d and e respectively). Further, we again tried to displace the water molecule by adding a methoxyethyl (Figure 3.38-f) or a propine group (Figure 3.38-g). Finally, we tried to enhance the interactions with Asn193 by replacing the carboxylic group with a carboxamide (Figure 3.38-h).

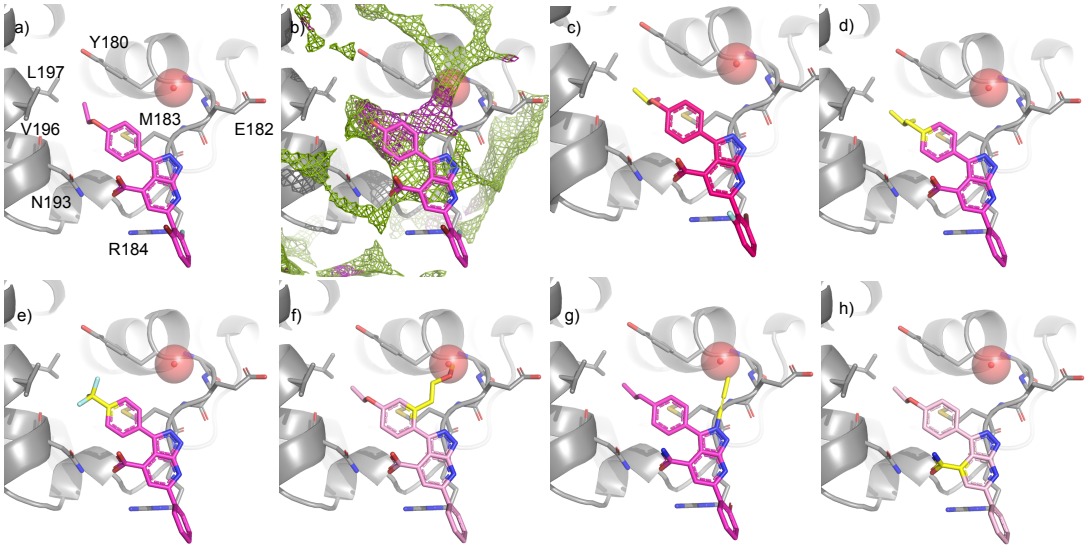


Figure 3.38. Analogues of Pose 2 (c - h). The protein is shown in grey while the different analogues are in pink tones with modifications shown in yellow. cGRILL maps in b and the water molecule in sphere.

These analogues were synthesized at the Experimental Therapeutics Program in the CNIO, screened by SPR to check their binding affinity (Table 3.2) and co-crystallized to verify our predictions. One of the SPR positive hits (K_d = 104 µM), analogue "**A19**" in Figure 3.37-d, exhibits much clearer electron density compared to the original hit 17 compound which allows to place **A19** unambiguously. The chlorine modification on the phenyl ring appears to select pose 1, in which the *p*-chlorophenyl moiety is lodged inside the hydrophobic pocket on the FERM F2 lobe (Figure 3.39) and the affinity is improved 5 fold compared to hit 17 (Table 3.2).

Compound	K_d (µM)	Ligand density	Pose	Image
FH17	513 ± 250	partial	not conclusive	Figure 3.37/38-a/b
A34	ND	none	not determined	Figure 3.37-c
A19	104 ± 17	yes	determined	Figure 3.37-d
A40	563 ± 230	none	not determined	Figure 3.37-e
A41	234 ± 44	none	not determined	Figure 3.37-f
A47	NA	none	not determined	Figure 3.37-g
A46	NA	none	not determined	Figure 3.38-c
A38	ND	none	not determined	Figure 3.38-e
ND: not detectable; NA: not analysed				

Table 3.2. K_d from SPR analysis of Hit 17 and its analogues.

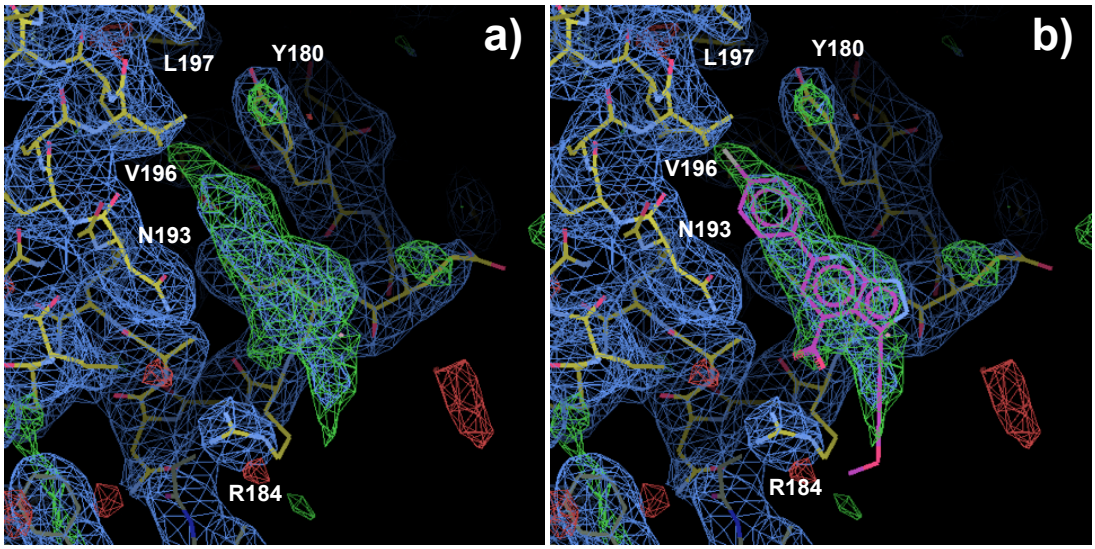


Figure 3.39. Electron density maps and fitted **A19** in its complex with the FERM domain in B). The 2Fo-Fc density at 1σ showed in blue; difference Fo-Fc maps in green (for positive values) and red (negative values). Both maps are countered at 1σ. Image courtesy of Ph.D. student M. Acebrón.

3.3.3.2 Identification of Transient Allosteric Pockets

In the following strategy, we established a collaboration with the group of D. E. Shaw (DE Shaw Research, New York, USA), who has one of the most powerful computers specifically built to perform MD simulations. This allowed us to run long unbiased MD simulations of the full FERM+kinase region with the aim of identifying alternative conformations that reveal potential pockets that are formed temporarily and can be targeted for discovery of type III ligands.

- MD simulations

We run 100 μs of unbiased MD simulations to extensively explore the conformational space. The simulations were run on ANTON [94] and the system was prepared for this specific computer. The FERM+kinase domain region consists of 645 amino acids (starting from Arg 35 and ending with Ile 679), the missing parts in the crystal structure, such as the linker and the A-loop, were modelled (see 3.2.2.1) as we did for the NMA. The protein was solvated in a cubic box of TIP3P water molecules [204] and then energy minimized. The CHARMM force field was used [205], [206] and output was

saved every 1 ns. On each frame of the simulation the program SiteMap (Schrödinger [142]) was used to identify targetable pockets and ranked according to their DScore (see 1.3.3.1). The top 50 pockets were visually inspected and six pockets were selected to perform VS (Figure 3.40-b) based on DScore, number of occurrences, divergence from the crystal structure and an available strategy to develop identified ligands into allosteric inhibitors (Table 3.3) [21].

The strategies that were considered for allosteric inhibition are:

I. Linking FERM- and kinase domains: This approach consists of stabilizing the autoinhibited closed conformation of FAK with compounds that interact with both FERM and kinase domains. Any pocket located close to the domain interface can be considered for this approach. (pockets 1, 31, 42).
II. Competing with FAK activators: In particular we considered pockets close to the PIP_2 binding site [21] activation (pocket 1).
III. Unclustering: preventing the clustering needed for FAK activation ([21], *Annex I, Figure 8, step 1*) (pockets 3, 14).
IV. Back-pocket exploration: located close to DFG-motif, between α-helix C and the A-loop, or in some cases reaching below the DFG (pocket 16).

Pocket ID	Max Dscore	Frame in Anton trajectory	Average Dscore	Number of Occurences	Classification	Strategy
1	1,366	85475	0,796	1894	FERM	I. Linking FERM - kinase domains and II. Compiting with activators (e.g. PIP2)
3	1,328	19741	0,875	12114	FERM	III. Unclustering
14	1,241	47904	0,922	44978	FERM	III. Unclustering
16	1,235	19227	0,945	20731	KINASE	IV. Targeting the back pocket
31	1,191	291	0,729	4691	KINASE	I. Linking FERM - kinase domains
42	1,158	2261	0,918	7164	INTERFACE	I. Linking FERM - kinase domains

Table 3.3. Pockets selected out of the 50. From left to right the following parameters are shown: maximun druggability score (DScore), first frame where the pocket appears (Frame in Anton), average DScore from all occurrences, number of occurrences of each pocket along the simulation, classification indicating where the pocket is located, either kinase domain, FERM domain or at the interface and the strategy to follow.

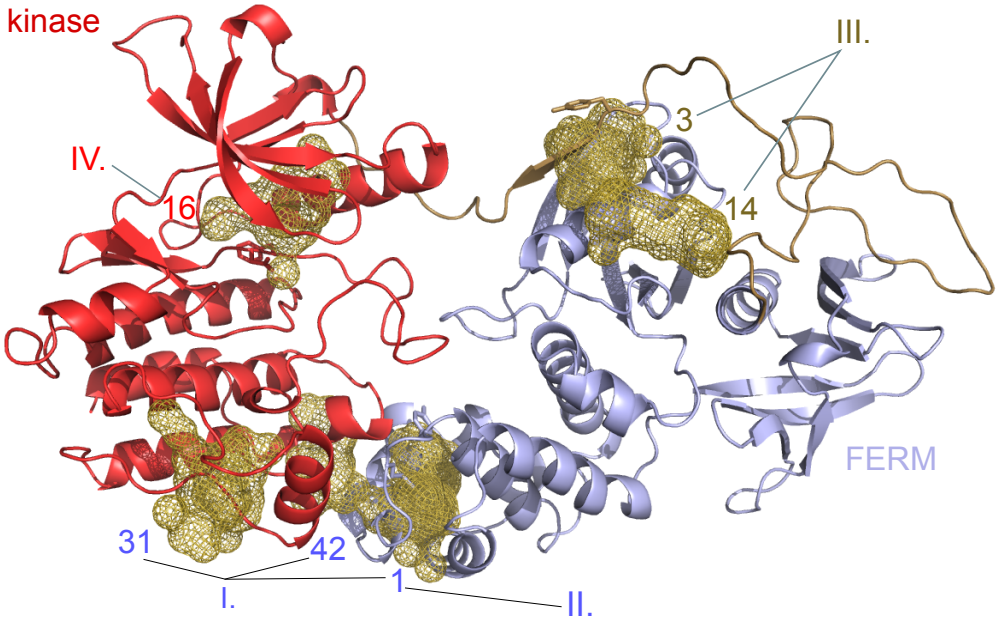


Figure 3.40. Structure of autoinhibited FAK, with the kinase domain in red and the FERM domain in blue, showing the 6 pockets as gold meshes. Roman numerals refer to the strategy followed for each pocket.

3.3.3.3 Virtual Screening on Allosteric Pockets

After pocket selection, VS was performed on each pocket using the Glide program [153] with a virtual library containing ~530.000 fragments selected from the ZINC database [207] . The top 500 fragments for each pocket were visually inspected and approximately 7 fragments were selected for each of them.

To assess the binding stability of the interacting fragments, standard MD simulations of their respective complexes with FAK (31) were performed. The compounds were previously parameter assigned using CHARMM General Force Field (CGenFF) and CGenFF Program [208], [209]. In cases where the molecules could not be assigned directly with the toolkit, parameters were manually assigned. To confirm whether those molecules remained stably bound in the pocket or they are undocked the RMSD was computed (*Annex 3*). A total of 18 compounds remained stably bound for ~200 ns and eight of them were purchased in the first round.

3.3.3.4 Experimental Validation

For validation, SPR experiments were carried out for all the compounds and for compounds with K_d values <150 µM, further competition assays were achieved (Table 3.4).

Compound	K_d (µM)	Competition assay	Co-crystallized	Pocket targeted Nº
VSC1	28 ± 40	yes	✓	31
VSC2	370 ± 46	no	✗	42
VSC3	145 ± 53	yes	✓	31
VSC4	62 ± 15	yes	✓	3
VSC5	400 ± 6	no	✗	3
VSC6	463 ± 100	no	✗	42
VSC7	402 ± 6	yes	✓	42
VSC8	117 ± 48	yes	✓	14

Table 3.4. Summary table showing (from left to right): the K_d from SPR analysis, if the compound where subjected to a competition assay with the ATP competitor Crizotinib, if it was subjected to co-crystallization and the intended pocket to be targeted.

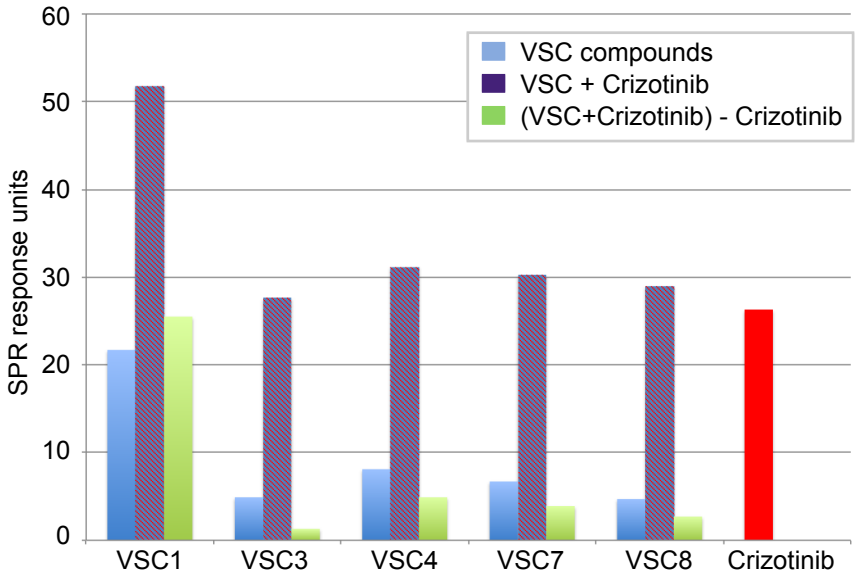


Figure 3.41. Competition assay of **VSC1**, **VSC3**, **VSC4**, **VSC7** and **VSC8** with the ATP site inhibitor Crizotinib (K_d=0,2 µM). Shown are the SPR responses of: in blue 50 µM of the **VSC compounds**, in purple 50 µM of **VSC** plus 0,5 µM of Crizotinib, and in green the response of VSC+Crizotinib with the Crizotinib response (red) subtracted. (Data courtesy of Ph.D. student M. Acebrón)

To probe whether the compounds bound in the orthosteric site or, as predicted, outside the ATP site, we prepared a competition SPR assay were binding of the compounds was measure, first alone at 50 µM and then in presence of 0,5 µM Crizotinib (see 1.3.3.5). VSC binding is considered non-ATP completive if the VSC and Crizotinib responses are additive, or ATP competitive if the VSC compound does not significantly increase the Crizotinib response. As shown in Figure 3.41 **VSC1** exhibits a response that is additive to the Crizotinib response, suggesting that this compound binds none-ATP competitive (Figure 3.41). For the other compounds the response is too small compared to Crizotinib, rendering the experiment inconclusive.

Further co-crystallization has been performed with each compound and the diffraction data obtained is currently being processed.

3.3.3.4 Discussion

We performed FBLD, using either experimental (^{19}F-NMR) or computational screening (VS) with small-molecule fragments (<300 Da) to increase the probability of finding new hits that target allosteric binding sites. Fragments with low molecular weight and low molecular complexity can sample chemical space more efficiently than larger molecules. Although, HTS have clearly identified many value compounds, they had so far limited success for discovery of allosteric regulators, in particular for kinases. In order to target small pockets outside the active site, our approach focuses on identifying small fragments with weak potency as starting points for further structure guided optimization by merging, linking, or growing fragments into drug-like molecules [162], [166].

Herein, we demonstrate that starting from a small library of fluorinated fragment compounds, we were able to detect and verify several allosteric ligands for FAK. Although we are still at an initial stage, we were able to optimize one ligand (A19), resulting in increased affinity and conclusive structural information, which will facilitate further structure guided optimization. Similar strategies are increasingly being pursued in the big pharma with promising results [210].

Computational tools are contributing hugely to many aspects of drug design and are of particular importance in FBLD, either for initial screening or during fragment growth.

In this thesis we present how a variety of computational tools, such as VS, pocket characterization methodologies (e.g. cGRILL), docking and modeling can help in various stages during the drug discovery and development process in order to understand and/or improve a given drug molecule [57].

After all, tomorrow is another day!
Gone with the wind (1942)

Chapter 4

4. Conclusions

4.1 GSK3β phospho-transfer reaction mechanism

- The GSK3β phospho-transfer reaction appears to occur through a mechanism in which the phosphoryl group approaches the serine hydroxyl whose proton is stabilised by a hydrogen bond with the aspartate 181 side chain.
- The proton transfer from Asp181 to the phosphoryl group is essential in the phospho-transfer because it weakens the strong charge-charge interactions between the Mg^{2+} and the phosphoserine.
- Li^{+} is a GSK3β competitive inhibitor with respect to one of the bound Mg^{2+} ions, Mg1, but not with respect to ATP or the substrate. Moreover, one of the most likely states for this near-attack conformation destabilization to occur in presence of Li^{+} is before the phospho-transfer reaction takes place, i.e: in presence of GSK3β, ATP, Mg^{2+}, and the peptide preventing the phosphoryl transfer reaction.

4.2 Substrate specifity of the SHIP2 phosphatase domain and allosteric effects of the C2 domain

- The C2 domain of SHIP2 is an integral part of the catalytic unit and allosterically affects SHIP2 catalysis. This mechanism involves mainly two regions: (i) three helices in the Ptase domain communicating directly with the substrate binding site, and (ii), a contiguous *loop* (*loop 4*) acting as a gate keeper for substrate entry to the binding site.
- Arg682 and Asn684 in loop 4 recognize the P3 and P4 phosphate respectively from the inositol ring of the substrate. Both are required for the reaction to take place.

- In absence of C2 domain, Arg665 plays an essence role in the reaction, participating in the octahedral Mg^{2+} coordination by stabilizing a network of water molecules.
- The presence of the C2 domain promotes an open *loop 4* conformation, allowing interactions between Arg682 and Asp613 and/or Asp615. This open conformation facilitates access to the substrate cavity, therefore promoting substrate binding.

4.3 Activation mechanisms of kinases

- The A-loop can adopt a wide range of extended conformations. Among them, the fully open state is stabilized when the activation loop is phosphorylated (on Tyr393 for Abl).
- When the A-loop is unphosphorylated, the half open - or so-called Src-like - state of Abl is the most stable conformation and the lowest energy state, while the completely closed state, which requires a DFG flip, represents a higher energy "local minimum", which requires the presence of a type II inhibitor (e.g. imatinib) for easy access.
- In FAK, the ATP binding not only stabilizes the binding site, but also two key regions, helices C and G, which directly interact with the regulatory FERM domain in the autoinhibited conformation of FAK.

4.4 Inhibition of FAK

- Different from compound **19b**, compound **19a** is a high affinity, selective FAK inhibitor. Although both compounds are highly similar (identical functional groups but different positions), their binding modes are not. The interactions of compound **19a** are more numerous and less solvent exposed compared to **19b**.
- By applying computational tools we were able to design analogues of initial fragment hits with higher affinity for allosteric sites on FAK, providing improved structural data for further compound optimisation.

- Fragment screening is a suitable strategy for initial screening to target allosteric pockets. This has allowed us to obtain structural information and we have succeeded in identifing several initial hits.
- Combining MD and VS allows for the discovery of compounds that target temporary allosteric pockets that are not visible in crystal structures. Via this method we identify several FAK binders and confirm for one compound a binding mode outside the ATP binding pocket.
- We demonstrate that combining fragment screening with structure based computational strategies can offer an alternative and/or complementary drug discovery pipeline compared to high-cost HTS campaigns.

Those are my principles, and if you don't like them... well, I have others.
Groucho Marx

Chapter 5

5. Resumen

5.1 Antecedentes y objetivos

Las proteína quinasas son unas enzimas esenciales para prácticamente todos los procesos celulares ya que son las encargadas de catalizar las reacciones de fosforilación. La mayoría de ellas son consideradas potenciales dianas farmacológicas debido a su implicación en numerosos procesos tumorales, como pueden ser la proliferación celular, la invasividad, la desregulación de la apoptosis, la formación de metástasis y la neo-angiogénesis. En los últimos años, avances en muchas y variadas disciplinas han hecho posible la identificación de numerosas macromoléculas diana, el conocimiento de sus secuencias, e incluso la determinación con resolución atómica de sus estructuras tridimensionales y la de sus complejos con inhibidores. Este progreso ha contribuido enormemente a la comprensión de la plasticidad estructural y regulación de numerosas dianas biológicas, entre las que se encuentran un buen número de proteínas-quinasas.

Se están utilizando también una variedad de métodos teóricos, con un importantísimo componente computacional, para predecir la estructura de proteínas y su comportamiento dinámico, aspecto clave a la hora de comprender las funciones de péptidos, proteínas y ácidos nucleicos, así como las de sus complejos entre sí o con moléculas orgánicas.

En este contexto, la presente tesis doctoral acomete el estudio de tres temas relacionados: (i) las reacciones de fosforilación llevadas a cabo por una serie de quinasas y la de desfosforilación catalizada por una fosfatasa presente en la membrana celular, (ii) el proceso de activación de dos de estas quinasas, y (iii) la aplicación de distintos métodos computacionales a la búsqueda de potenciales inhibidores de algunas proteína-quinasas seleccionadas.

El primer tema consiste en un estudio a nivel atómico de los mecanismos de fosforilación y defosforilación. La reacción de fosforilación es una de las modificaciones postraduccionales más utilizada en los procesos celulares de transducción de señales. La familia de proteínas que lleva a cabo esta modificación son las quinasas, las cuales catalizan la transferencia del fosfato γ del ATP hasta el grupo hidroxilo de un aminoácido en concreto, generalmente serina (Ser), treonina (Thr) o tirosina (Tyr). En consecuencia, estas enzimas funcionan como interruptores imprescindibles regulando diferentes funciones biológicas.

Un ejemplo de quinasa para la que aún no se conoce con exactitud su mecanismo de activación es la conocida como glucógeno sintasa kinasa 3β (GSK3β). Esta quinasa juega varios papeles principales en numerosas vías de señalización, siendo su desregulación (principalmente hiperactivación) crítica a la hora de desarrollar patologías como la diabetes *mellitus*, la enfermedad de Alzeimer, otros trastornos neurológicos y cáncer. La actividad de esta serina/treonina quinasa está controlada por la fosforilación. Para poder entender el proceso de activación y el mecanismo de transferencia del fosfato con mayor exactitud utilizamos como modelo otra quinasa que está muy bien estudiada a nivel bioquímico y estructural, la proteína kinasa A (PKA). Estudiamos en primer lugar los pasos de la reacción de fosforilación catalizados por esta enzima sobre un péptido sustrato típico con objeto de validar todos nuestros protocolos teóricos ("control positivo") y poder después entender cuáles son las diferencias en GSK3β que hacen que esta enzima presente determinadas características especiales, como son la inhibición específica por iones litio o la necesidad de que la mayoría de péptidos sustrato deban estar previamente fosforilados para ser reconocidos.

A continuación estudiamos la "SH2 domain-containing Inositol 5'-Phosphatase 2" (SHIP2), codificada en el gen *INPPL1* (*IN*ositol *P*olyphosphate *P*hosphatase-*L*ike *1*), como modelo de fosfatasa que lleva acabo una reacción de defosforilación. Los miembros de la familia a la que pertenece esta enzima son fosfoesterasas dependientes de Mg^{2+} que están encargadas de la regulación de los niveles tanto de fosfoinosítidos como de inositoles, que son moléculas de señalización que regulan la proliferación y la supervivencia celular, la reorganización del citoesqueleto y el tráfico de vesículas, reclutando proteínas efectoras de las membranas celulares. Para llevar a cabo el estudio de esta fosfatasa, modelamos los substratos $PI(4,5)P_3$ e IP_4 y simulamos la dinámica de

la proteína en presencia y ausencia de los mismos, analizando así la influencia de ambos inositoles sobre la conformación de la enzima. Todas las fosfatasas están compuestas por un dominio llamado "5-fosfatasa" que contiene alrededor 300 aminoácidos. Además de este, SHIP posee un dominio adicional y diferente al resto de los dominios de otras fosfatasas llamado C2. Este dominio C2 fue cristalizado en complejo con la 5-fosfatasa en nuestro laboratorio y comparado con el dominio fosfatasa en términos cinéticos. Para poder arrojar un poco de luz sobre este sistema y la influencia del dominio adyacente C2, al igual que sobre la preferencia por los distintos sustratos, analizamos los diferentes complejos y todas las posibles combinaciones mediante dinámica molecular y cálculos de energía de unión ligando-receptor. Además, propusimos algunas mutaciones puntuales que nos deberían ayudar a explicar el mecanismo de activación de esta proteína.

Como segundo tema, estudiamos los cambios conformacionales en la activación de dos quinasas diferentes: Abl y Focal Adhesion Kinase (FAK). Para la primera llevamos a cabo un estudio energético aplicando métodos de cálculo de energía libre y caracterizamos la influencia del estado de fosforilación sobre la flexibilidad de esta enzima. Para la segunda, además de simular la proteína en presencia y ausencia de ATP para tratar de comprender cuáles son las zonas afectadas por la unión del sustrato mediante dinámica molecular, llevamos a cabo un análisis de modos normales para entender mejor el efecto alostérico del ATP, previamente identificado mediante métodos experimentales (FRET).

Por último, procedimos a la elaboración de un protocolo novedoso para la búsqueda de inhibidores de quinasas y, más concretamente inhibidores de FAK, mediante la conjunción de una serie de técnicas tanto computacionales como experimentales. Empleamos desde larguísimas simulaciones de dinámica molecular para intentar visitar conformaciones alternativas de la proteína de interés (y así poder explorar diferentes bolsillos en los que podrían tener cabida distintos inhibidores) hasta técnicas de cribado virtual utilizando quimiotecas de más de medio millón de fragmentos moleculares. Sometimos los candidatos seleccionados a análisis posteriores (e.g. resonancia de plasmón superficial) e intentamos validar los positivos mediante co-cristalización con la proteína diana.

5.2 Metodología

La metodología fundamentalmente empleada en esta tesis doctoral consiste, en su mayor parte, en cálculos teóricos realizados *in silico* y visualización tridimensional de las estructuras macromoleculares y sus ligandos mediante gráficos por ordenador. El abanico de técnicas utilizadas cubre desde el modelado y simulación de estas estructuras mediante métodos de mecánica molecular y mecánica cuántica hasta la dinámica molecular, los análisis de modos normales, el cálculo de energías libres y la predicción de unión ligando-diana. Se puede considerar que todas estas metodologías han supuesto una auténtica revolución en el campo de la biología estructural y su aplicación a la farmacología.

El modelado molecular ha estado presente desde los años 50, cuando Linus Pauling generó el primer modelo satisfactorio de una macromolécula, un péptido con conformación de hélice α, basándose en datos de difracción de rayos X y utilizando los principios generales de la estructura molecular. Posteriormente, en 1958, Kendrew y sus colaboradores construyeron el primer modelo de una proteína, la mioglobina, al que rápidamente siguieron, de forma progresiva pero incesante, los miles de modelos de estructuras basadas en datos bien procedentes de rayos X o radiación de sincrotrón, microscopía electrónica o espectroscopía de RMN que hoy en día tenemos a nuestra disposición en el repositorio público conocido como *Protein Data Bank* (PDB).

Además de esta explosión de conocimiento estructural, el hecho de poder visualizar las moléculas en 3D también ha supuesto un avance extraordinario. Desde los modelos de "bolas y palitos" de Kendrew o los modelos de "alambres" de Rubin y Richardson hasta los sofisticados programas de que disponemos hoy en día. Pero no sólo hemos visto progreso en los modelos, que al fin y al cabo son imágenes estáticas de las moléculas. También las simulaciones de estos sistemas que comenzaron en los años 70, con los parámetros de los hidrocarburos de Norman L. Allinger, han ido evolucionando haciendo que hoy en día seamos capaces incluso de simular atómicamente el plegamiento de una proteína .

Dentro de los distintos métodos computacionales, dependiendo del grado de precisión y la magnitud de nuestros sistemas tendremos que elegir uno u otro. La química computacional, por ejemplo, nos permite estudiar los sistemas moleculares como un conjunto de núcleos y electrones cuando hablamos de mecánica cuántica. Por su parte, la mecánica clásica, comúnmente conocida como mecánica molecular, se basa en las leyes de la mecánica Newtoniana, que necesita asumir varias aproximaciones para poder simplificar los sistemas sometidos a estudio. De esta manera, los átomos quedan representados como masas esféricas y los enlaces entre los átomos como fuerzas armónicas o muelles. También nos encontramos con métodos híbridos que comprenden los dos anteriores, la mecánica cuántica y la mecánica molecular, tratando cada parte del sistema de una manera según nos interese y obteniendo los beneficios de ambas.

En 1964 se llevó a cabo por Rahman la primera simulación de argón líquido, y 10 años más tarde se simuló mediante dinámica molecular el agua líquida. Después llegarían las simulaciones de las primeras proteínas, aunque con tiempos nada comparables a los que somos capaces de simular hoy en día. Además la cantidad de parámetros que pueden extraerse de las simulaciones ha abierto innumerables posibilidades a la hora de comparar sistemas, energías, estabilidades termodinámicas, reacciones químicas e incluso la predicción de unión entre distintas proteínas o de proteínas y ligandos, que también estudiamos en esta tesis doctoral.

La interacción entre moléculas constituye el lenguaje básico de los sistemas biológicos, estando la mayoría de las enfermedades ligadas a un mal funcionamiento de estas interacciones. Un mejor conocimiento de cómo ocurren no sólo nos ayuda a entender el proceso sino también a poder buscar alternativas que reemplacen o interfieran con esta función alterada para poder evitarla o revertirla. Siguiendo esta línea de investigación, nos centramos en la búsqueda de nuevos ligandos capaces de bloquear uno de los sistemas de señalización más involucrados en la generación o mantenimiento de los tumores, como es el llevado a cabo por ciertas proteína-quinasas.

Para llevar a cabo el estudio o la predicción de la unión del ligando con su diana tendremos que tener en cuenta varios factores dependiendo de si disponemos de información estructural tanto de la diana como del ligando: necesitaremos caracterizar el bolsillo, seleccionar las quimiotecas más adecuadas para cada diana, generar las

diferentes conformaciones de los ligandos, realizar un cribado con todas estás moléculas y evaluarlas mediante una función de tanteo para poder establecer una adecuada priorización de candidatos. Para ello, hoy en día disponemos de varias herramientas computacionales tremendamente sofisticadas tanto para el análisis de las propiedades físico-químicas del bolsillo, estableciendo cuáles son los residuos que pueden participar en la unión al ligando (e.g. cGRILL, GRID, PP_SITE, LigBuilder), como para realizar el acoplamiento automatizado de una única molécula (*docking*) o colecciones de las mismas (cribado virtual) (e.g. AutoDock, Glide, DOCK, CRDOCK).

5.3 Conclusiones

5.3.1 MECANISMO DE LA REACCIÓN DE FOSFORILACIÓN EN GSK3β

- La transferencia del fosfato en GSK3β ocurre mediante un mecanismo en el que el grupo fosfato se aproxima al grupo hidroxilo de la serina, el cual tiene el protón fijado al aspártico 181 mediante un enlace de hidrógeno.
- La transferencia del protón es imprescindible en la reacción porque debilita las interacciones existentes carga-carga entre el Mg^{2+} y la fosfoserina.
- La inhibición por litio no sólo ocurre compitiendo con uno de los Mg^{2+}, Mg1, sino que además es plausible que esta inhibición ocurra previa a la fosforilación, es decir, en presencia de GSK3β, ATP, Mg^{2+}, y el péptido impidiendo así la transferencia del fosfato.

5.3.2 EFECTO ALOSTÉRICO DEL DOMINIO C2 Y ESPECIFICIDAD DE SUSTRATO POR SHIP2

- El dominio C2 es parte del dominio catalítico de SHIP2, al cual estabiliza mediante un mecanismo alostérico. Este mecanismo implica principalmente dos regiones: la primera formada por tres hélices, que comunica el C2 con la parte superior de la fosfatasa donde se encuentra el sustrato, y el *loop* contiguo (*loop* 4) que actúa de tapa. Además, el *loop* 4 contiene los residuos Arg682 y Asn684.

- La Arg682 reconoce el fosfato P3 del anillo de inositol y la Asn684 reconoce el fosfato P4, siendo ambos necesarios para que ocurra la reacción de fosforilación.
- La Arg665 es crucial para que la reacción tenga lugar en ausencia del C2, participando en la coordinación octaédrica del Mg^{2+} a través de una red de moléculas de agua.
- La presencia del dominio C2 favorece una conformación abierta del *loop* 4, en la que la Arg682 interacciona con los Asp613 y Asp615, facilitando de este modo la apertura de la cavidad donde se une el sustrato y favoreciendo su unión.

5.3.3 MECANISMO DE ACTIVACIÓN DE LAS QUINASAS

- El A-loop de Abl es capaz de adoptar múltiples conformaciones. De entre todas ellas, la conformación totalmente abierta es la más estable cuando la enzima activada se encuentra fosforilada en la Tyr393.
- Cuando Abl no está fosforilada, la más estable es la conformación semi-abierta mientras que la conformación totalmente cerrada sólo es estable en presencia de un inhibidor tipo II (e.g. imatinib) que le permita adoptar esa conformación.
- En FAK, el ATP no solo estabiliza el sitio de unión sino especialmente las hélices-α C y G. Estás dos hélices, además, están en contacto con el dominio regulatorio FERM en la forma autoinhibida.

5.3.4 INHIBICIÓN DE FAK

- A diferencia del compuesto **19b**, el compuesto **19a** es un inhibidor selectivo de FAK de alta afinidad. Pese a que ambos compuestos son muy similares (mismos grupos funcionales pero diferente posición), su modo de unión no lo es. Las interacciones que se producen en el **19a** son más numerosas y están menos expuestas al disolvente que las del compuesto **19b**.
- Mediante el uso de herramientas computacionales somos capaces de diseñar análogos para incrementar su afinidad y la probabilidad de que puedan ser cristalizados más fácilmente para dilucidar su modo de unión receptor.

- El cribado de fragmentos es una estrategia acertada para un primer cribado dirigido a bolsillos alostéricos. Esto nos ha permitido obtener información estructural con la que poder identificar varios fragmentos iniciales.
- Buscamos bolsillos alóstericos de FAK combinando DM y cribados virtuales de fragmentotecas. Estos bolsillos son posteriormente validados mediante técnicas experimentales (e.g. SPR y co-cristalización).
- Demostramos que estrategias computacionales basadas en la estructura del receptor ofrecen una alternativa o complemento útil a los costosos experimentos de HTS dirigidos a la identificación de candidatos a fármacos.

En la noche que me envuelve,
Negra como un pozo insondable,
Doy gracias al Dios que fuere
Por mi alma inconquistable.

En las garras de las circunstancias
No he gemido ni llorado.
Ante las puñaladas del azar
Si bien he sangrado, jamás me he postrado.

Más allá de este lugar de ira y llantos
Acecha la oscuridad con su horror,
No obstante la amenaza de los años
Me alla y me allará sin temor.

Ya no importa cuán recto haya sido el camino,
Ni cuántos castigos lleve a la espalda.
Soy el amo de mi destino:
Soy el capitán de mi alma.

Invictus (2009)

Chapter 6

6. Bibliography

[1] E. H. Fisher and E. G. Krebs, “Phosphorylase activity of skeletal muscle extracts.,” *J. Biol. Chem.*, vol. 216, pp. 113–120, 1955.

[2] E. H. Fisher and E. G. Krebs, “Conversion of phosphorylase b to phosphorylase a in muscle extracts,” *J. Biol. Chem.*, vol. 216, pp. 121–132, 1955.

[3] W. D.A, P. J.P, and E. Krebs, “An adenosine 3’,5'-monophosphate-dependant protein kinase from rabbit skeletal muscle.,” *jounal Biol. Chem.*, vol. 243, no. 13, pp. 3763–3766, 1968.

[4] S. Taylor and A. Kornev, “Protein kinases: evolution of dynamic regulatory proteins,” *Trends Biochem. Sci.*, vol. 36, no. 2, pp. 65–77, Feb. 2011.

[5] J. a Ubersax and J. E. Ferrell Jr, “Mechanisms of specificity in protein phosphorylation,” *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 7, pp. 530–541, Jul. 2007.

[6] T. Hunter, “Why nature chose phosphate to modify proteins.,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 367, no. 1602, pp. 2513–6, Sep. 2012.

[7] S. W. Cowan-Jacob, “Structural biology of protein tyrosine kinases.,” *Cell. Mol. Life Sci.*, vol. 63, no. 22, pp. 2608–25, Nov. 2006.

[8] S. K. Hanks, A. M. Quinn, and T. Hunter, “The Kinase Family : Conserved Protein Phylogeny Features and Deduced Domains of the Catalytic,” *Science (80-.).*, vol. 241, no. 4861, pp. 42–52, 1988.

[9] A. W. Oliver, S. Knapp, and L. H. Pearl, “Activation segment exchange: a common mechanism of kinase autophosphorylation?,” *Trends Biochem. Sci.*, vol. 32, no. 8, pp. 351–6, Aug. 2007.

[10] R. a Norman, D. Toader, and A. D. Ferguson, “Structural approaches to obtain kinase selectivity.,” *Trends Pharmacol. Sci.*, vol. 33, no. 5, pp. 273–8, May 2012.

[11] A. Krupa, G. Preethi, and N. Srinivasan, “Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases.,” *J. Mol. Biol.*, vol. 339, no. 5, pp. 1025–39, Jun. 2004.

[12] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, “The Protein Kinase Complement of the Human Genome,” *Science (80-.).*, vol. 298, no. 5600, pp. 1912–1934, 2002.

[13] M. Magrane and U. Consortium, "UniProt Knowledgebase: a hub of integrated protein data.," *Database (Oxford).*, vol. 2011, p. bar009, Jan. 2011.

[14] S. K. Hanrs and T. Hunter, "The eukaryotic protein kinase superfamily: kinase (catalytic) domam structure and classification.," *FASEB J.*, vol. 9, no. 8, pp. 576–96, 1995.

[15] D. R. Knighton, J. Zheng, L. F. T. E. N. Eyck, V. A. Ashford, N. Xuong, S. S. Taylor, and J. M. Sowadski, "crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase," *Science (80-.).*, vol. 253, pp. 407–414, 1991.

[16] S. S. Taylor, P. Zhang, J. M. Steichen, M. M. Keshwani, and A. P. Kornev, "PKA: lessons learned after twenty years.," *Biochim. Biophys. Acta*, vol. 1834, no. 7, pp. 1271–8, Jul. 2013.

[17] J. Kunz, M. P. Wilson, M. Kisseleva, J. H. Hurley, P. W. Majerus, and R. A. Anderson, "The activation loop of phosphatidylinositol phosphate kinases determines signaling specificity," *Mol Cell*, vol. 5, no. 1, pp. 1–11., 2000.

[18] H. N. Banavath, O. P. Sharma, M. S. Kumar, and R. Baskaran, "Identification of novel tyrosine kinase inhibitors for drug resistant T315I mutant BCR-ABL: a virtual screening and molecular dynamics simulations study.," *Sci. Rep.*, vol. 4, p. 6948, Jan. 2014.

[19] C. A. Lipinski and J. C. Loftus, "The Pyk2 FERM domain: a Novel Therapeutic Target Christopher," *Expert Opin. Ther. Targets*, vol. 14, no. 1, pp. 95–108, 2010.

[20] P. Filippakopoulos, M. Kofler, O. Hantschel, G. D. Gish, F. Grebien, E. Salah, P. Neudecker, L. E. Kay, B. E. Turk, G. Superti-Furga, T. Pawson, and S. Knapp, "Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation.," *Cell*, vol. 134, no. 5, pp. 793–803, Sep. 2008.

[21] G. M. Goñi, C. Epifano, J. Boskovic, M. Camacho-Artacho, J. Zhou, A. Bronowska, M. T. Martín, M. J. Eck, L. Kremer, F. Gräter, F. L. Gervasio, M. Perez-Moreno, and D. Lietha, "Phosphatidylinositol 4,5-bisphosphate triggers activation of focal adhesion kinase by inducing clustering and conformational changes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 31, pp. E3177–86, Aug. 2014.

[22] J. Monod, J. Wyman, and J. P. Changeux, "on the Nature of Allosteric Transitions: a Plausible Model.," *J. Mol. Biol.*, vol. 12, no. 1, pp. 88–118, 1965.

[23] H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser, “The ensemble nature of allostery.,” *Nature*, vol. 508, no. 7496, pp. 331–9, Apr. 2014.

[24] G. Bollag, J. Tsai, J. Zhang, C. Zhang, P. Ibrahim, K. Nolop, and P. Hirth, “Vemurafenib: the first drug approved for BRAF-mutant cancer.,” *Nat. Rev. Drug Discov.*, vol. 11, no. 11, pp. 873–86, Dec. 2012.

[25] Y. Shan, M. a Seeliger, M. P. Eastwood, F. Frank, H. Xu, M. Ø. Jensen, R. O. Dror, J. Kuriyan, and D. E. Shaw, “A conserved protonation-dependent switch controls drug binding in the Abl kinase.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 1, pp. 139–144, 2009.

[26] V. Bolos, J. M. Gasent, S. Lopez-Tarruella, E. Grande, V. Bolós, J. M. Gasent, S. López-Tarruella, and E. Grande, “The dual kinase complex FAK-Src as a promising therapeutic target in cancer,” *Onco Targets Ther*, vol. 3, pp. 83–97, 2010.

[27] R. Bayliss, T. Sardon, I. Vernos, and E. Conti, “Structural Basis of Aurora-A Activation by TPX2 at the Mitotic Spindle,” *Mol. Cell*, vol. 12, no. 4, pp. 851–862, Oct. 2003.

[28] C.-J. Tsai and R. Nussinov, “The molecular basis of targeting protein kinases in cancer therapeutics.,” *Semin. Cancer Biol.*, vol. 23, no. 4, pp. 235–42, Aug. 2013.

[29] I. Shchemelinin, L. Šefc, and E. Nečas, “Protein kinases, their function and implication in cancer and other diseases,” *Folia Biol. (Praha).*, vol. 52, no. 3, pp. 81–101, 2006.

[30] A. P. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, and M. R. Stratton, “A CENSUS OF HUMAN CANCER GENES,” *Nat. Rev. Cancer*, vol. 4, no. 3, pp. 177–183, 2009.

[31] F. Cavallo, C. De Giovanni, P. Nanni, G. Forni, and P.-L. Lollini, “2011: the immune hallmarks of cancer,” *Cell*, vol. 60, no. 3, pp. 319–326, 2011.

[32] L. Goodman, M. Wintrobe, W. Dameshek, M. Goodman, A. Gilman, and M. McLennan, “Nitrogen mustard therapy: Use of methyl-bis(beta-chloroethyl)amine hydrochloride and tris(beta-chloroethyl)amine hydrochloride for hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders,” *JAMA*, vol. 251, no. 17, pp. 2255–2261, May 1984.

[33] S. Farber, L. K. Diamond, R. D. Mercer, R. F. Sylvester, and J. A. Wolff, “The New England Journal of Medicine Downloaded from nejm.org at HOUSTON

ACADEMY OF MEDICINE on January 10, 2012. For personal use only. No other uses without permission. From the NEJM Archive. Copyright © 2010 Massachusetts Medical Society. All rights res," 2010.

[34] J. Zhang, P. L. Yang, and N. S. Gray, "Targeting cancer with small molecule kinase inhibitors.," *Nat. Rev. Cancer*, vol. 9, no. 1, pp. 28–39, Jan. 2009.

[35] M. Angiolini, "Targeting the DFG-in kinase conformation: a new trend emerging from a patent analysis.," *Future Med. Chem.*, vol. 3, no. 3, pp. 309–337, 2011.

[36] J. Blanc, R. Geney, and C. Menet, "Type II kinase inhibitors: an opportunity in cancer for rational design.," *Anticancer. Agents Med. Chem.*, vol. 13, no. 5, pp. 731–47, 2013.

[37] J. S. Tokarski, J. a Newitt, C. Y. J. Chang, J. D. Cheng, M. Wittekind, S. E. Kiefer, K. Kish, F. Y. F. Lee, R. Borzillerri, L. J. Lombardo, D. Xie, Y. Zhang, and H. E. Klei, "The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants.," *Cancer Res.*, vol. 66, no. 11, pp. 5790–7, Jun. 2006.

[38] H. G. Choi, J. Zhang, E. Weisberg, J. D. Griffin, T. Sim, and N. S. Gray, "Development of 'DFG-out' inhibitors of gatekeeper mutant kinases," *Bioorganic Med. Chem. Lett.*, vol. 22, no. 16, pp. 5297–5302, Aug. 2012.

[39] J. A. Wereszczynski, J., McCammon, "Accelerated Molecular Dynamics in Computational Drug Design," in *Computational Drug Discovery and Design. Methods in Molecular Biology*, 2012, pp. 515–524.

[40] F. Zuccotto, E. Ardini, E. Casale, and M. Angiolini, "Through the 'gatekeeper door': exploiting the active kinase conformation.," *J. Med. Chem.*, vol. 53, no. 7, pp. 2681–94, Apr. 2010.

[41] K. Dorey, J. R. Engen, J. Kretzschmar, M. Wilm, G. Neubauer, T. Schindler, and G. Superti-Furga, "Phosphorylation and structure-based functional studies reveal a positive and a negative role for the activation loop of the c-Abl tyrosine kinase.," *Oncogene*, vol. 20, no. 56, pp. 8075–84, Dec. 2001.

[42] H. Enright and P. B. McGlave, "Chronic myelogenous leukemia.," *Curr. Opin. Hematol.*, vol. 2, no. 4, pp. 293–299, 1995.

[43] B. Clarkson, a Strife, D. Wisniewski, C. L. Lambek, and C. Liu, "Chronic myelogenous leukemia as a paradigm of early cancer and possible curative strategies.," *Leukemia*, vol. 17, no. 7, pp. 1211–62, Jul. 2003.

[44] O. Hantschel and G. Superti-Furga, “Regulation of the c-Abl and Bcr–Abl Tyrosine Kinases,” *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 1, pp. 33–44, Jan. 2004.

[45] A. Sirvent, C. Benistant, and S. Roche, “Cytoplasmic signalling by the c-Abl tyrosine kinase in normal and cancer cells.,” *Biol. Cell*, vol. 100, no. 11, pp. 617–31, Nov. 2008.

[46] C. Gambacorti-Passerini, L. Antolini, F.-X. Mahon, F. Guilhot, M. Deininger, C. Fava, A. Nagler, C. M. Della Casa, E. Morra, E. Abruzzese, A. D’Emilio, F. Stagno, P. le Coutre, R. Hurtado-Monroy, V. Santini, B. Martino, F. Pane, A. Piccin, P. Giraldo, S. Assouline, M. a Durosinmi, O. Leeksma, E. M. Pogliani, M. Puttini, E. Jang, J. Reiffers, M. G. Valsecchi, and D.-W. Kim, “Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib.,” *J. Natl. Cancer Inst.*, vol. 103, no. 7, pp. 553–61, Apr. 2011.

[47] N. P. Shah and C. L. Sawyers, “Mechanisms of resistance to STI571 in Philadelphia chromosome-associated leukemias.,” *Oncogene*, vol. 22, no. 47, pp. 7389–95, Oct. 2003.

[48] B. Druker and F. Guilhot, “Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia,” *N. Engl. J. Med.*, vol. 355, pp. 2408–17, 2006.

[49] G. W. McLean, N. O. Carragher, E. Avizienyte, J. Evans, V. G. Brunton, and M. C. Frame, “The role of focal-adhesion kinase in cancer - a new therapeutic opportunity.,” *Nat. Rev. Cancer*, vol. 5, no. 7, pp. 505–15, Jul. 2005.

[50] G. A. Infusino and J. R. Jacobson, “Endothelial FAK as a Therapeutic Target in Disease,” *Brain Behav. Immun.*, vol. 22, no. 5, p. 4109, 2008.

[51] J.-P. Liu, J. Baker, A. S. Perkins, E. J. Robertson, and A. Efstratiadis, “Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r),” *Cell*, vol. 75, no. 1, pp. 59–72, Oct. 1993.

[52] F. J. Sulzmaier, C. Jean, and D. D. Schlaepfer, “FAK in cancer: mechanistic findings and clinical applications.,” *Nat. Rev. Cancer*, vol. 14, no. 9, pp. 598–610, Aug. 2014.

[53] D. Lietha, X. Cai, D. F. J. Ceccarelli, Y. Li, M. D. Schaller, and M. J. Eck, “Structural basis for the autoinhibition of Focal Adhesion Kinase,” *Cell*, vol. 129, no. 6, pp. 1177–1187, 2008.

[54] a Ali, K. P. Hoeflich, and J. R. Woodgett, "Glycogen synthase kinase-3: properties, functions, and regulation.," *Chem. Rev.*, vol. 101, no. 8, pp. 2527–40, Aug. 2001.

[55] M. Alonso and a Martinez, "GSK-3 inhibitors: discoveries and developments.," *Curr. Med. Chem.*, vol. 11, no. 6, pp. 755–63, Mar. 2004.

[56] P. Cohen and S. Frame, "The renaissance of GSK3.," *Nat. Rev. Mol. Cell Biol.*, vol. 2, no. 10, pp. 769–776, 2001.

[57] M. Valiev, J. Yang, J. a Adams, S. S. Taylor, and J. H. Weare, "Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations.," *J. Phys. Chem. B*, vol. 111, no. 47, pp. 13455–64, Nov. 2007.

[58] A. C. Bastidas, M. S. Deal, J. M. Steichen, Y. Guo, J. Wu, and S. S. Taylor, "Phosphoryl transfer by protein kinase A is captured in a crystal lattice.," *J. Am. Chem. Soc.*, vol. 135, no. 12, pp. 4788–98, Mar. 2013.

[59] B. Halford, "Limits of lithium," *Chem. Eng. News*, vol. 91, no. 12, pp. 15–20, 2013.

[60] W. J. Ryves, R. Dajani, L. Pearl, and A. J. Harwood, "Glycogen synthase kinase-3 inhibition by lithium and beryllium suggests the presence of two magnesium binding sites.," *Biochem. Biophys. Res. Commun.*, vol. 290, no. 3, pp. 967–72, Jan. 2002.

[61] R. S. Jope, "Lithium and GSK-3: one inhibitor, two inhibitory actions, multiple outcomes.," *Trends Pharmacol. Sci.*, vol. 24, no. 9, pp. 441–3, Sep. 2003.

[62] J. a McCubrey, L. S. Steelman, F. E. Bertrand, N. M. Davis, M. Sokolosky, S. L. Abrams, G. Montalto, a B. D'Assoro, M. Libra, F. Nicoletti, R. Maestro, J. Basecke, D. Rakus, A. Gizak, Z. N. Demidenko, L. Cocco, a M. Martelli, and M. Cervello, "GSK-3 as potential target for therapeutic intervention in cancer," *Oncotarget*, vol. 5, no. 10, pp. 2881–2911, 2014.

[63] H.-C. Cheng, R. Z. Qi, H. Paudel, and H.-J. Zhu, "Regulation and function of protein kinases and phosphatases.," *Enzyme Res.*, vol. 2011, p. 794089, Jan. 2011.

[64] J. C. Whisstock, S. Romero, R. Gurung, H. Nandurkar, L. M. Ooms, S. P. Bottomley, and C. a Mitchell, "The inositol polyphosphate 5-phosphatases and the apurinic/apyrimidinic base excision repair endonucleases share a common mechanism for catalysis.," *J. Biol. Chem.*, vol. 275, no. 47, pp. 37055–61, Nov.

2000.

[65] Y. Chi, B. Zhou, W.-Q. Wang, S.-K. Chung, Y.-U. Kwon, Y.-H. Ahn, Y.-T. Chang, Y. Tsujishita, J. H. Hurley, and Z.-Y. Zhang, "Comparative mechanistic and substrate specificity study of inositol polyphosphate 5-phosphatase Schizosaccharomyces pombe Synaptojanin and SHIP2.," *J. Biol. Chem.*, vol. 279, no. 43, pp. 44987–95, Oct. 2004.

[66] G. Di Paolo and P. De Camilli, "Phosphoinositides in cell regulation and membrane dynamics.," *Nature*, vol. 443, no. 7112, pp. 651–7, Oct. 2006.

[67] F. Nakatsu, R. M. Perera, L. Lucast, R. Zoncu, J. Domin, F. B. Gertler, D. Toomre, and P. De Camilli, "The inositol 5-phosphatase SHIP2 regulates endocytic clathrin-coated pit dynamics.," *J. Cell Biol.*, vol. 190, no. 3, pp. 307–15, Aug. 2010.

[68] H. Ishihara, T. Sasaoka, H. Hori, T. Wada, H. Hirai, T. Haruta, W. J. Langlois, and M. Kobayashi, "Molecular Cloning of Rat SH2-Containing Inositol Phosphatase 2 (SHIP2) and Its Role in the Regulation of Insulin Signaling," *Biochem. Biophys. Res. Commun.*, vol. 260, no. 1, pp. 265–272, Jun. 1999.

[69] A. Suwa, T. Yamamoto, A. Sawada, K. Minoura, N. Hosogai, A. Tahara, T. Kurama, T. Shimokawa, and I. Aramori, "Discovery and functional characterization of a novel small molecule inhibitor of the intracellular phosphatase, SHIP2.," *Br. J. Pharmacol.*, vol. 158, no. 3, pp. 879–87, Oct. 2009.

[70] G. M. Fuhler, R. Brooks, B. Toms, S. Iyer, E. a Gengo, M.-Y. Park, M. Gumbleton, D. R. Viernes, J. D. Chisholm, and W. G. Kerr, "Therapeutic potential of SH2 domain-containing inositol-5'-phosphatase 1 (SHIP1) and SHIP2 inhibition in cancer.," *Mol. Med.*, vol. 18, pp. 65–75, Jan. 2012.

[71] M. Pirruccello and P. De Camilli, "Inositol 5-phosphatases: insights from the Lowe syndrome protein OCRL.," *Trends Biochem. Sci.*, vol. 37, no. 4, pp. 134–43, Apr. 2012.

[72] P. W. Majerus, M. V Kisseleva, and F. Anderson Norris, "The role of phosphatases in inositol signaling reactions," *J. Biol. Chem.*, vol. 274, no. 16, pp. 10669–10672, 1999.

[73] C. A. Mitchell, S. Brown, J. K. Campbell, A. D. Munday, and C. J. Speed, "Regulation of second messengers by the inositol polyphosphate 5-phosphatases," *Biochem. Soc. Trans.*, vol. 24, no. 4, p. 994—1000, 1996.

[74] D. Crowfoot Hodgkin, “The X-ray analysis of complicated molecules,” *Nobel Lect.*, 1964.

[75] B. Y. L. Pauling and R. B. Corey, “A PROPOSED STRUCTURE FOR THE NUCLEIC ACIDS,” *Proc. Natl. Acad. Sci.*, vol. 39, pp. 84–97, 1953.

[76] J. C. KENDREW, G. BODO, H. M. DINTZIS, R. G. PARRISH, H. WYCKOFF, and D. C. PHILLIPS, “A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis,” *Nature*, vol. 181, no. 4610, pp. 662–666, Mar. 1958.

[77] B. Rubin and J. S. Richardson, “The simple construction of protein alpha-Carbon models,” *Biopolymers*, vol. 11, no. 11, pp. 2381–2385, 1972.

[78] R. A. Sayle and E. J. Milner-White, “RASMOL: biomolecular graphics for all,” *Trends Biochem. Sci.*, vol. 20, no. 9, pp. 374–376, Sep. 1995.

[79] S. Takamori, M. Holt, K. Stenius, E. a Lemke, M. Grønborg, D. Riedel, H. Urlaub, S. Schenck, B. Brügger, P. Ringler, S. a Müller, B. Rammner, F. Gräter, J. S. Hub, B. L. De Groot, G. Mieskes, Y. Moriyama, J. Klingauf, H. Grubmüller, J. Heuser, F. Wieland, and R. Jahn, “Molecular anatomy of a trafficking organelle.,” *Cell*, vol. 127, no. 4, pp. 831–46, Nov. 2006.

[80] N. L. Allinger, “Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms,” *J. Am. Chem. Soc.*, vol. 99, no. 25, pp. 8127–8134, Dec. 1977.

[81] C. K. Varma, “Molecular Mechanical Force fields,” *Biochem. Standford*, vol. 218, pp. 1–11, 2001.

[82] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, “Systematic validation of protein force fields against experimental data,” *PLoS One*, vol. 7, no. 2, pp. 1–6, Jan. 2012.

[83] S. L. Mayo, B. D. Olafson, and W. A. I. Goddard, “DREIDING: A Generic Force Field for Molecular Simulations,” *J. Phys. Chem.*, vol. 101, no. Suite 540, pp. 8897–8909, 1990.

[84] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules,” *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, May 1995.

[85] N. L. Allinger, “Understanding molecular structure from molecular mechanics.,” *J. Comput. Aided. Mol. Des.*, vol. 25, no. 4, pp. 295–316, Apr. 2011.

[86] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, a Aspuru-Guzik, and a G. White, “Towards quantum chemistry on a quantum computer.,” *Nat. Chem.*, vol. 2, no. 2, pp. 106–11, Feb. 2010.

[87] M. R. Hestenes and E. Stiefel, “Methods of conjugate gradients for solving linear systems,” *J. Res. Natl. Bur. Stand. (1934).*, vol. 49, no. 6, p. 409, Dec. 1952.

[88] S. S. Petrova and A. D. Solov’ev, “The Origin of the Method of Steepest Descent,” *Hist. Math.*, vol. 24, no. 4, pp. 361–375, 1997.

[89] B. J. Alder and T. E. Wainwright, “Phase Transition for a Hard Sphere System,” *J. Chem. Phys.*, vol. 27, no. 5, 1957.

[90] A. Rahman, “Correlations in the Motion of Atoms in Liquid Argon,” *Phys. Rev.*, vol. 136, no. 2A, pp. A405–A411, Oct. 1964.

[91] F. H. Stillinger and B. Laboratories, “Propagation of sound in water. A molecular-dynamics study,” *Phys. Rev.*, vol. 10, no. 1, pp. 368–378, 1974.

[92] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins.,” *Nature*, vol. 267, no. 5612, pp. 585–90, 1977.

[93] D. E. Shaw, P. Maragakis, K. Lindorff-larsen, S. Piana, Y. Shan, and W. Wriggers, “Atomic-Level Characterization of the Structural Dynamics of Proteins,” *Science (80-.).*, vol. 330, pp. 341–347, 2010.

[94] D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. M. Deneroff, M. a. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, and K. J. Bowers, “Anton, a special-purpose machine for molecular dynamics simulation,” *ACM SIGARCH Comput. Archit. News*, vol. 35, no. 2, p. 1, Jun. 2007.

[95] H. Ode, M. Nakashima, S. Kitamura, W. Sugiura, and H. Sato, “Molecular dynamics simulation in virus research.,” *Front. Microbiol.*, vol. 3, no. July, p. 258, Jan. 2012.

[96] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, a. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, no. 8, p. 3684, 1984.

[97] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Phys. Rev.*, vol. 31, no. 3, pp. 1695–1697, 1985.

[98] M. W. Karaman, S. Herrgard, D. K. Treiber, P. Gallant, C. E. Atteridge, B. T. Campbell, K. W. Chan, P. Ciceri, M. I. Davis, P. T. Edeen, R. Faraoni, M. Floyd, J. P. Hunt, D. J. Lockhart, Z. V Milanov, M. J. Morrison, G. Pallares, H. K. Patel, S. Pritchard, L. M. Wodicka, and P. P. Zarrinkar, "A quantitative analysis of kinase inhibitor selectivity.," *Nat. Biotechnol.*, vol. 26, no. 1, pp. 127–132, Jan. 2008.

[99] O. F. Lange and H. Grubmüller, "Can principal components yield a dimension reduced description of protein dynamics on long time scales?," *J. Phys. Chem. B*, vol. 110, no. 45, pp. 22842–22852, 2006.

[100] J. G. Kirkwood, "Statistical Mechanics of Fluid Mixtures," *J. Chem. Phys.*, vol. 3, no. 5, p. 300, 1935.

[101] B. Roux, "The calculation of the potential of mean force using computer simulations," *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 275–282, Sep. 1995.

[102] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *J. Comput. Phys.*, vol. 23, no. 2, pp. 187–199, Feb. 1977.

[103] J. Kästner, "Umbrella sampling," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 1, no. 6, pp. 932–942, Nov. 2011.

[104] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comput. Chem.*, vol. 13, no. 8, pp. 1011–1021, 1992.

[105] M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations," *Comput. Phys. Commun.*, vol. 135, no. 1, pp. 40–57, Mar. 2001.

[106] J. Kästner and W. Thiel, "Bridging the gap between thermodynamic integration

and umbrella sampling provides a novel analysis method: 'Umbrella integration,'" *J. Chem. Phys.*, vol. 123, no. 14, p. -, 2005.

[107] A. Laio and M. Parrinello, "Escaping free-energy minima.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 20, pp. 12562–6, Oct. 2002.

[108] A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports Prog. Phys.*, vol. 71, no. 12, p. 126601, Dec. 2008.

[109] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, "Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics.," *J. Phys. Chem. B*, vol. 110, no. 8, pp. 3533–9, Mar. 2006.

[110] C. Micheletti, A. Laio, and M. Parrinello, "Reconstructing the Density of States by History-Dependent Metadynamics," *Phys. Rev. Lett.*, vol. 92, no. 17, p. 170601, 2004.

[111] M. Bonomi and M. Parrinello, "Enhanced Sampling in the Well-Tempered Ensemble," *Phys. Rev. Lett.*, vol. 104, no. 19, pp. 1–4, May 2010.

[112] N. Go, T. Noguti, and T. Nishikawa, "Dynamics of a small globular protein in terms of low-frequency vibrational modes.," *Proc. Natl. Acad. Sci.*, vol. 80, no. 12, pp. 3696–3700, 1983.

[113] J. S. Hub and B. L. de Groot, "Detection of functional modes in protein dynamics.," *PLoS Comput. Biol.*, vol. 5, no. 8, p. e1000480, Aug. 2009.

[114] A. Manuscript, "Requirement for Mechanochemical Activity of Enzymes," vol. 13, no. 6, pp. 893–904, 2006.

[115] F. Tama and Y.-H. Sanejouand, "Conformational change of proteins arising from normal mode calculations," *Protein Eng. Des. Sel.*, vol. 14, no. 1, pp. 1–6, Jan. 2001.

[116] M. Delarue and Y.-H. Sanejouand, "Simplified Normal Mode Analysis of Conformational Transitions in DNA-dependent Polymerases: the Elastic Network Model," *J. Mol. Biol.*, vol. 320, no. 5, pp. 1011–1024, Jul. 2002.

[117] A. Bakan, L. M. Meireles, and I. Bahar, "ProDy: protein dynamics inferred from theory and experiments.," *Bioinformatics*, vol. 27, no. 11, pp. 1575–7, Jun. 2011.

[118] F. Gago, "Molecular simulations of drug–receptor complexes in anticancer research," *Future Med. Chem.*, vol. 4, no. 15, pp. 1961–1970, Oct. 2012.

[119] E. Fischer, "Einfluß der Konfiguration auf die Wirkung der Enzyme. I," in *Untersuchungen Über Kohlenhydrate und Fermente (1884–1908) SE - 101*, Springer Berlin Heidelberg, 1909, pp. 836–844.

[120] D. E. Koshland, "The Key–Lock Theory and the Induced Fit Theory," *Angew. Chemie Int. Ed. English*, vol. 33, no. 2324, pp. 2375–2378, 1995.

[121] C. J. Tsai, S. Kumar, B. Ma, and R. Nussinov, "Folding funnels, binding funnels, and protein function.," *Protein Sci.*, vol. 8, no. 6, pp. 1181–90, Jun. 1999.

[122] A. D. Vogt and E. Di Cera, "NIH Public Access," vol. 52, no. 34, pp. 5723–5729, 2014.

[123] A. C. Gibbs, "Elements and Modulation of Functional Dynamics," *J. Med. Chem.*, vol. 57, no. 19, pp. 7819–7837, Oct. 2014.

[124] T. Kenakin, "Principles: receptor theory in pharmacology.," *Trends Pharmacol. Sci.*, vol. 25, no. 4, pp. 186–92, Apr. 2004.

[125] P. R. Connelly, R. A. Aldape, F. J. Bruzzese, S. P. Chambers, M. J. Fitzgibbon, M. A. Fleming, S. Itoh, D. J. Livingston, M. A. Navia, and J. A. Thomson, "Enthalpy of hydrogen bond formation in a protein-ligand binding reaction.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 5, pp. 1964–1968, 1994.

[126] T. Hou, J. Wang, Y. Li, and W. Wang, "Assessing the performance of the MM/PBSA and MM/GBSA methods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations," *J. Chem. Inf. Model.*, vol. 51, no. 1, pp. 69–82, 2012.

[127] D. L. Mobley and K. A. Dill, "Binding of Small-Molecule Ligands to Proteins: 'What You See' Is Not Always 'What You Get,'" *NIH*, vol. 17, no. 4, pp. 489–498, 2010.

[128] D. Jiao, P. a Golubkov, T. a Darden, and P. Ren, "Calculation of protein-ligand binding free energy by using a polarizable potential.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 17, pp. 6290–6295, Apr. 2008.

[129] M. R. Reddy and M. D. Erion, "Calculation of Relative Binding Free Energy Differences for Fructose 1,6-Bisphosphatase Inhibitors Using the

Thermodynamic Cycle Perturbation Approach," *J. Am. Chem. Soc.*, vol. 123, no. 26, pp. 6246–6252, Jul. 2001.

[130] P. Ferrara, H. Gohlke, D. J. Price, G. Klebe, and C. L. Brooks, "Assessing Scoring Functions for Protein−Ligand Interactions," *J. Med. Chem.*, vol. 47, no. 12, pp. 3032–3047, Jun. 2004.

[131] B. K. Shoichet, A. R. Leach, and I. D. Kuntz, "Ligand solvation in molecular docking," *Proteins Struct. Funct. Bioinforma.*, vol. 34, no. 1, pp. 4–16, 1999.

[132] B. Munos, "Lessons from 60 years of pharmaceutical innovation.," *Nat. Rev. Drug Discov.*, vol. 8, no. 12, pp. 959–68, Dec. 2009.

[133] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders, and S. Otto, "Dynamic combinatorial chemistry.," *Chem. Rev.*, vol. 106, no. 9, pp. 3652–711, Sep. 2006.

[134] R. P. Hertzberg and A. J. Pope, "High-throughput screening: new technology for the 21st century," *Curr. Opin. Chem. Biol.*, vol. 4, no. 4, pp. 445–451, Aug. 2000.

[135] B. S. P. A. Srinivas Reddy, B. S. P. S. Priyadarshini Pati, B. S. P. P. Praveen Kumar, B. S. P. H.N. Pradeep, and B. S. P. G. Narahari Sastry, "Virtual Screening in Drug Discovery - A Computational Perspective," *Curr. Protein Pept. Sci.*, vol. 8, no. 4, pp. 329–351, Aug. 2007.

[136] J. K. Morrow and S. Zhang, "Computational Prediction of Hot Spot Residues," *Curr. Pharmacol. Des.*, vol. 18, no. 9, pp. 1255–1265, 2013.

[137] P. J. Goodford, "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules," *J. Med. Chem.*, vol. 28, no. 7, pp. 849–857, Jul. 1985.

[138] R. Brenke, D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M. R. Landon, C. Mattos, and S. Vajda, "Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques.," *Bioinformatics*, vol. 25, no. 5, pp. 621–7, Mar. 2009.

[139] C. H. Ngan, T. Bohnuud, S. E. Mottarella, D. Beglov, E. a Villar, D. R. Hall, D. Kozakov, and S. Vajda, "FTMAP: extended protein mapping with user-selected probe molecules.," *Nucleic Acids Res.*, vol. 40, pp. W271–5, Jul. 2012.

[140] G. López, A. Valencia, and M. L. Tress, "Firestar--Prediction of Functionally

Important Residues Using Structural Templates and Alignment Reliability.," *Nucleic Acids Res.*, vol. 35, pp. W573–7, Jul. 2007.

[141] J. Günther, A. Bergner, M. Hendlich, and G. Klebe, "Utilising Structural Knowledge in Drug Design Strategies: Applications Using Relibase," *J. Mol. Biol.*, vol. 326, no. 2, pp. 621–636, Feb. 2003.

[142] T. a Halgren, "Identifying and characterizing binding sites and assessing druggability.," *J. Chem. Inf. Model.*, vol. 49, no. 2, pp. 377–89, Feb. 2009.

[143] S. Mukherjee, T. E. Balius, and R. C. Rizzo, "Docking Validation Resources: Protein Family and Ligand Flexibility Experiments," *J. Chem. Inf. Model.*, vol. 50, no. 11, pp. 1986–2000, Oct. 2010.

[144] Á. Cortés Cabrera, J. Klett, H. G. Dos Santos, A. Perona, S. M. Francis, E. M. Priego, F. Gago, and A. Morreale, "CRDOCK : An Ultrafast Multipurpose Protein − Ligand Docking Tool," *J. Chem. Inf. Model.*, vol. 52, pp. 2300–2309, 2012.

[145] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *J. Mol. Biol.*, vol. 267, no. 3, pp. 727–748, Apr. 1997.

[146] G. Jones, P. Willett, and R. C. Glen, "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation," *J. Mol. Biol.*, vol. 245, no. 1, pp. 43–53, 1995.

[147] V. Namasivayam and R. Günther, "Research Article: pso@autodock: A Fast Flexible Molecular Docking Program Based on Swarm Intelligence," *Chem. Biol. Drug Des.*, vol. 70, no. 6, pp. 475–484, 2007.

[148] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010.

[149] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, "Metropolis." 1953.

[150] A. Žiemys, L. Rimkute, and J. Kulys, "Modeling the Enantioselective Enzymatic Reaction with Modified Genetic Docking Algorithm," *Nonlinear Anal.*, vol. 9, no. 4, pp. 373–383, 2004.

[151] H. Park, J. Lee, and S. Lee, "Critical assessment of the automated AutoDock as a new docking tool for virtual screening," *Proteins Struct. Funct. Bioinforma.*, vol. 65, no. 3, pp. 549–554, 2006.

[152] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, vol. 47, no. 7, pp. 1739–1749, Feb. 2004.

[153] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening," *J. Med. Chem.*, vol. 47, no. 7, pp. 1750–1759, Feb. 2004.

[154] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm.," *J. Mol. Biol.*, vol. 261, no. 3, pp. 470–89, Aug. 1996.

[155] B. Kramer, M. Rarey, and T. Lengauer, "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking.," *Proteins*, vol. 37, no. 2, pp. 228–41, 1999.

[156] X. Barril and F. Javier Luque, "Molecular simulation methods in drug discovery: a prospective outlook," *J. Comput. Aided. Mol. Des.*, vol. 26, no. 1, pp. 81–86, 2012.

[157] F. Sousa, P. A. Fernandes, and M. J. Ramos, "Protein – Ligand Docking : Current Status and Future Challenges," *Proteins Struct. Funct. Bioinforma.*, vol. 65, pp. 15–26, 2006.

[158] Y.-C. Chen, "Beware of docking!," *Trends Pharmacol. Sci.*, vol. 36, no. 2, pp. 78–95, Dec. 2014.

[159] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev, "H++: a server for estimating pKas and adding missing hydrogens to macromolecules.," *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W368–71, Jul. 2005.

[160] B. K. Shoichet, "Virtual screening of chemical libraries.," *Nature*, vol. 432, no. 7019, pp. 862–865, 2004.

[161] N. B. Vinh, J. S. Simpson, P. J. Scammells, and D. K. Chalmers, "Virtual

screening using a conformationally flexible target protein: models for ligand binding to p38α MAPK.," *J. Comput. Aided. Mol. Des.*, vol. 26, no. 4, pp. 409–23, Apr. 2012.

[162] R. A. E. Carr, M. Congreve, C. W. Murray, and D. C. Rees, "Fragment-based lead discovery: Leads by design," *Drug Discov. Today*, vol. 10, no. 14, pp. 987–992, 2005.

[163] P. Gribbon and S. Andreas, "High-throughput drug discovery: What can we expect from HTS?," *Drug Discov. Today*, vol. 10, no. 1, pp. 17–22, Jan. 2005.

[164] M. M. Hann, A. R. Leach, and G. Harper, "Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery," *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 3, pp. 856–864, Apr. 2001.

[165] C. Lipinski and A. Hopkins, "Navigating chemical space for biology and medicine," *Nature*, vol. 432, no. 7019, pp. 855–861, Dec. 2004.

[166] H. Chen, X. Zhou, A. Wang, Y. Zheng, Y. Gao, and J. Zhou, "Evolutions in fragment-based drug design: the deconstruction-reconstruction approach.," *Drug Discov. Today*, vol. 20, no. 1, pp. 105–113, Sep. 2014.

[167] S. F. Sousa, a. J. M. Ribeiro, J. T. S. Coimbra, R. P. P. Neves, S. a. Martins, N. S. H. N. Moorthy, P. a. Fernandes, and M. J. Ramos, "Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field," *Curr. Med. Chem.*, vol. 20, no. 18, pp. 2296–2314, Apr. 2013.

[168] A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, S.-J. J. Deng, R. T. Gampe, T. E. Grisard, K. P. Madauss, R. T. Nolte, W. J. Rocque, L. Wang, K. L. Weaver, S. P. Williams, G. B. Wisely, R. Xu, and L. M. Shewchuk, "Crystallization of protein-ligand complexes.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 63, no. Pt 1, pp. 72–9, Jan. 2007.

[169] O. Gerlits, J. Tian, A. Das, S. Taylor, P. Langan, and W. T. Heller, "Phosphoryl transfer reaction snapshots in crystals: Insights into the mechanism of protein kinase A catalytic subunit," *J. Biol. Chem.*, vol. 290, pp. 15538–15548, 2015.

[170] O. Gerlits, M. J. Waltman, S. Taylor, P. Langan, and A. Kovalevsky, "Insights into the phosphoryl transfer catalyzed by cAMP-dependent protein kinase: an X-ray crystallographic study of complexes with various metals and peptide substrate SP20.," *Biochemistry*, vol. 52, no. 21, pp. 3721–7, May 2013.

[171] N. Homeyer, A. C. Horn, H. Lanig, and H. Sticht, "AMBER force-field

parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine," *J. Mol. Model.*, vol. 12, no. 3, pp. 281–289, 2006.

[172] S. Frame, P. Cohen, and R. M. Biondi, "A common phosphate binding site explains the unique substrate specificity of GSK3 and its inactivation by phosphorylation," *Mol. Cell*, vol. 7, pp. 1321–1327, 2001.

[173] R. Dajani, E. Fraser, S. M. Roe, N. Young, V. Good, T. C. Dale, and L. H. Pearl, "Crystal structure of glycogen synthase kinase 3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition.," *Cell*, vol. 105, no. 6, pp. 721–32, Jun. 2001.

[174] W. J. Ryves and a J. Harwood, "Lithium inhibits glycogen synthase kinase-3 by competition for magnesium.," *Biochem. Biophys. Res. Commun.*, vol. 280, no. 3, pp. 720–5, Jan. 2001.

[175] L. Trésaugues, C. Silvander, S. Flodin, M. Welin, T. Nyman, S. Gräslund, M. Hammarström, H. Berglund, and P. Nordlund, "Structural basis for phosphoinositide substrate recognition, catalysis, and membrane interactions in human inositol polyphosphate 5-phosphatases," *Structure*, vol. 22, no. 5, pp. 744–755, 2014.

[176] Y. Tsujishita, S. Guo, L. E. Stolz, J. D. York, and J. H. Hurley, "Specificity determinants in phosphoinositide dephosphorylation: crystal structure of an archetypal inositol polyphosphate 5-phosphatase," *Cell*, vol. 105, no. 3, pp. 379–389, 2001.

[177] A. B. Jefferson and P. W. Majerus, "Mutation of the Conserved Domains of Two Inositol Polyphosphate 5-Phosphatases," *Biochemistry*, vol. 35, no. 24, pp. 7890–7894, Jan. 1996.

[178] D. Communi, R. Lecocq, and C. Erneux, "Arginine 343 and 350 Are Two Active Site Residues Involved in Substrate Binding by Human Type I D-myo-Inositol 1,4,5-Trisphosphate 5-Phosphatase," *J. Biol. Chem.*, vol. 271, no. 20, pp. 11676–11683, May 1996.

[179] J. Yuan, A. Amend, J. Borkowski, and R. DeMarco, "MULTICLUSTAL: a systematic method for surveying Clustal W alignment parameters," *Bioinformatics*, vol. 15, pp. 862–863., 1999.

[180] I. a Balabin, W. Yang, and D. N. Beratan, "Coarse-grained modeling of allosteric regulation in protein receptors.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 34, pp. 14253–8, Aug. 2009.

[181] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model.," *Biophys. J.*, vol. 80, no. 1, pp. 505–15, Jan. 2001.

[182] J. Lee, H. Yang, M. Georgescu, A. Di Cristofano, T. Maehama, Y. Shi, J. E. Dixon, P. Pandolfi, and N. P. Pavletich, "Crystal Structure of the PTEN Tumor Suppressor," *Cell*, vol. 99, no. 3, pp. 323–334, 1999.

[183] N. M. Levinson, O. Kuchment, K. Shen, M. a Young, M. Koldobskiy, M. Karplus, P. a Cole, and J. Kuriyan, "A Src-like inactive conformation in the abl tyrosine kinase domain.," *PLoS Biol.*, vol. 4, no. 5, p. e144, May 2006.

[184] J. M. Steichen, G. H. Iyer, S. Li, S. A. Saldanha, M. S. Deal, V. L. Woods, and S. S. Taylor, "Global consequences of activation loop phosphorylation on protein kinase A.," *J. Biol. Chem.*, vol. 285, no. 6, pp. 3825–32, Feb. 2010.

[185] L. Lin, R. Czerwinski, K. Kelleher, M. M. Siegel, P. Wu, R. Kriz, A. Aulabaugh, and M. Stahl, "Activation Loop Phosphorylation Modulates Bruton's Tyrosine Kinase (Btk) Kinase Domain Activity," *Biochemistry*, vol. 48, no. 9, pp. 2021–2032, Feb. 2009.

[186] Y. Meng and B. Roux, "Locking the active conformation of c-Src kinase through the phosphorylation of the activation loop.," *J. Mol. Biol.*, vol. 426, no. 2, pp. 423–35, Jan. 2014.

[187] D. Branduardi, F. L. Gervasio, and M. Parrinello, "From A to B in free energy space.," *J. Chem. Phys.*, vol. 126, no. 5, p. 054103, Mar. 2007.

[188] K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, and D. E. Shaw, "Structure and dynamics of an unfolded protein examined by molecular dynamics simulation," *J. Am. Chem. Soc.*, vol. 134, no. 8, pp. 3787–3791, 2012.

[189] K. Dorey, J. R. Engen, J. Kretzschmar, M. Wilm, G. Neubauer, T. Schindler, and G. Superti-Furga, "Phosphorylation and structure-based functional studies reveal a positive and a negative role for the activation loop of the c-Abl tyrosine kinase.," *Oncogene*, vol. 20, no. 56, pp. 8075–84, 2001.

[190] A. Fiser and A. Sali, "ModLoop: automated modeling of loops in protein structures," *Bioinformatics*, vol. 19, no. 18, pp. 2500–2501, Dec. 2003.

[191] E. Lindahl, C. Azuara, P. Koehl, and M. Delarue, "NOMAD-Ref: visualization,

deformation and refinement of macromolecular structures based on all-atom normal mode analysis.," *Nucleic Acids Res.*, vol. 34, no. Web Server issue, pp. W52–6, Jul. 2006.

[192] S. Hayward and B. L. de Groot, *Normal modes and essential dynamics.*, vol. 443. 2008.

[193] X. Cai, D. Lietha, D. F. Ceccarelli, A. V Karginov, Z. Rajfur, K. Jacobson, K. M. Hahn, M. J. Eck, and M. D. Schaller, "Spatial and temporal regulation of focal adhesion kinase activity in living cells.," *Mol. Cell. Biol.*, vol. 28, no. 1, pp. 201–14, Jan. 2008.

[194] A. L. Skinner and J. S. Laurence, "High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands," *J. Pharm. Sci.*, vol. 97, no. 11, pp. 4670–4695, 2008.

[195] E. Helmerhorst, D. J. Chandler, M. Nussio, and C. D. Mamotte, "Real-time and label-free bio-sensing of molecular interactions by surface plasmon resonance: A laboratory medicine perspective," *Clin. Biochem. Rev.*, vol. 33, no. November 2012, pp. 161–173, 2012.

[196] P. Dao, R. Jarray, N. Smith, Y. Lepelletier, J. Le Coq, D. Lietha, R. Hadj-Slimane, J.-P. Herbeuval, C. Garbay, F. Raynaud, and H. Chen, "Inhibition of both focal adhesion kinase and fibroblast growth factor receptor 2 pathways induces anti-tumor and anti-angiogenic activities.," *Cancer Lett.*, vol. 348, no. 1–2, pp. 88–99, Jun. 2014.

[197] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility," *J Comput Chem*, vol. 30, no. 16, pp. 2785–2791, 2010.

[198] S. M. Weis, S.-T. Lim, K. M. Lutu-Fuga, L. a Barnes, X. L. Chen, J. R. Göthert, T.-L. Shen, J.-L. Guan, D. D. Schlaepfer, and D. a Cheresh, "Compensatory role for Pyk2 during angiogenesis in adult mice lacking endothelial cell FAK.," *J. Cell Biol.*, vol. 181, no. 1, pp. 43–50, Apr. 2008.

[199] D. Lietha and M. J. Eck, "Crystal structures of the FAK kinase in complex with TAE226 and related bis-anilino pyrimidine inhibitors reveal a helical DFG conformation.," *PLoS One*, vol. 3, no. 11, p. e3800, Jan. 2008.

[200] S. Han, A. Mistry, J. S. Chang, D. Cunningham, M. Griffor, P. C. Bonnette, H. Wang, B. a Chrunyk, G. E. Aspnes, D. P. Walker, A. D. Brosius, and L. Buckbinder, "Structural characterization of proline-rich tyrosine kinase 2 (PYK2)

reveals a unique (DFG-out) conformation and enables inhibitor design.," *J. Biol. Chem.*, vol. 284, no. 19, pp. 13193–201, May 2009.

[201] S. K. Bhattacharya, G. E. Aspnes, S. W. Bagley, M. Boehm, A. D. Brosius, L. Buckbinder, J. S. Chang, J. Dibrino, H. Eng, K. S. Frederick, D. a Griffith, M. C. Griffor, C. R. W. Guimarães, A. Guzman-Perez, S. Han, A. S. Kalgutkar, J. Klug-McLeod, C. Garcia-Irizarry, J. Li, B. Lippa, D. a Price, J. a Southers, D. P. Walker, L. Wei, J. Xiao, M. P. Zawistoski, and X. Zhao, "Identification of novel series of pyrazole and indole-urea based DFG-out PYK2 inhibitors.," *Bioorg. Med. Chem. Lett.*, vol. 22, no. 24, pp. 7523–9, Dec. 2012.

[202] U. Grädler, J. Bomke, D. Musil, V. Dresing, M. Lehmann, G. Hölzemann, H. Greiner, C. Esdar, M. Krier, and T. Heinrich, "Fragment-based discovery of focal adhesion kinase inhibitors.," *Bioorg. Med. Chem. Lett.*, vol. 23, no. 19, pp. 5401–9, Oct. 2013.

[203] M. Garavís, B. López-Méndez, A. Somoza, J. Oyarzabal, C. Dalvit, A. Villasante, R. Campos-Olivas, and C. González, "Discovery of Selective Ligands for Telomeric RNA G-quadruplexes (TERRA) through 19F-NMR Based Fragment Screening," *ACS Chem. Biol.*, vol. 9, no. 7, pp. 1559–1566, May 2014.

[204] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.

[205] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[206] B. R. Brooks, C. L. B. Jr., A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, Q. C. L. Caves, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, J. M. T. Lazaridis, V. Ovchinnikov, R. W. P. E. Paci, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "CHARMM: The Biomolecular Simulation Program," *J Comput Chem*, vol. 30, no. 10, pp. 1545–1614, 2010.

[207] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman, "ZINC: a free tool to discover chemistry for biology.," *J. Chem. Inf. Model.*, vol. 52, no. 7, pp. 1757–68, Jul. 2012.

[208] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, and J. C. Gumbart, "Rapid parameterization of small molecules using the Force Field Toolkit.," *J. Comput. Chem.*, vol. 34, no. 32, pp. 2757–70, Dec. 2013.

[209] J. L. Knight, J. D. Yesselman, and C. L. Brooks, “Assessing the quality of absolute hydration free energies among CHARMM-compatible ligand parameterization schemes,” *J. Comput. Chem.*, vol. 34, no. 11, pp. 893–903, 2013.

[210] A. Hillisch, N. Heinrich, and H. Wild, “Computational Chemistry in the Pharmaceutical Industry: From Childhood to Adolescence.,” *ChemMedChem*, pp. 1–6, Sep. 2015.

[211] “Kinase Knowledge Base (Q4 2009). http://eidogen-sertanty. com/kinasekb.php (accessed Nov 30, 2012).” .

Chapter 7

7. Annexes

7.1 Annex 1

CrossMark click for updates

Phosphatidylinositol 4,5-bisphosphate triggers activation of focal adhesion kinase by inducing clustering and conformational changes

Guillermina M. Goñi[a], Carolina Epifano[b], Jasminka Boskovic[a], Marta Camacho-Artacho[a], Jing Zhou[c], Agnieszka Bronowska[c], M. Teresa Martín[d], Michael J. Eck[e,f], Leonor Kremer[d], Frauke Gräter[c], Francesco Luigi Gervasio[g,h], Mirna Perez-Moreno[b], and Daniel Lietha[a,1]

[a]Structural Biology and Biocomputing Program and [b]BBVA Foundation Cancer Cell Biology Program, Spanish National Cancer Research Centre, 28029 Madrid, Spain; [c]Heidelberg Institute for Theoretical Studies, 69118 Heidelberg, Germany; [d]Immunology and Oncology Department, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain; [e]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; [f]Department of Cancer Biology, Dana–Farber Cancer Institute, Boston, MA 02115; and [g]Institute of Structural and Molecular Biology and [h]Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom

Edited* by Timothy A. Springer, Immune Disease Institute, Program in Cellular and Molecular Medicine, Children's Hospital Boston, Boston, MA, and approved May 19, 2014 (received for review September 12, 2013)

Focal adhesion kinase (FAK) is a nonreceptor tyrosine kinase (NRTK) with key roles in integrating growth and cell matrix adhesion signals, and FAK is a major driver of invasion and metastasis in cancer. Cell adhesion via integrin receptors is well known to trigger FAK signaling, and many of the players involved are known; however, mechanistically, FAK activation is not understood. Here, using a multidisciplinary approach, including biochemical, biophysical, structural, computational, and cell biology approaches, we provide a detailed view of a multistep activation mechanism of FAK initiated by phosphatidylinositol-4,5-bisphosphate [PI(4,5)P_2]. Interestingly, the mechanism differs from canonical NRTK activation and is tailored to the dual catalytic and scaffolding function of FAK. We find PI(4,5)P_2 induces clustering of FAK on the lipid bilayer by binding a basic region in the regulatory 4.1, ezrin, radixin, moesin homology (FERM) domain. In these clusters, PI(4,5)P_2 induces a partially open FAK conformation where the autophosphorylation site is exposed, facilitating efficient autophosphorylation and subsequent Src recruitment. However, PI(4,5)P_2 does not release autoinhibitory interactions; rather, Src phosphorylation of the activation loop in FAK results in release of the FERM/kinase tether and full catalytic activation. We propose that PI(4,5)P_2 and its generation in focal adhesions by the enzyme phosphatidylinositol 4-phosphate 5-kinase type Iγ are important in linking integrin signaling to FAK activation.

cell signaling | phosphoinositides

Cell attachment to the ECM is mediated via integrin transmembrane receptors on the cell surface. Integrin engagement to ECM components results in activation and clustering of integrins. In response to integrin activation, a large number of proteins are recruited to their cytoplasmic tails, resulting in the formation of focal adhesions (FAs) (1). On the one side, FAs are anchoring points for actomyosin stress fibers, which allow tension forces to build up when contracting fibers exert their pulling force via FAs against the ECM. On the other hand, integrin activation and the generation of tension trigger intricate signaling cascades. A central signaling component in FAs is the nonreceptor tyrosine kinase (NRTK) focal adhesion kinase (FAK). FAK is activated downstream of integrins, and its signaling is important for cell migration, proliferation, and survival (2, 3). FAK contains numerous binding sites for other signaling and adaptor proteins, and it has been identified as a hub in the focal adhesion (FA) interactome (4). In addition to its role as a signaling kinase, FAK is therefore thought to function as a signaling scaffold. FAK is required for diverse processes in development, wound healing, and disease (5–7). FAK KO mice are not viable due to mesodermal defects (8), and early studies with FAK KO cells indicated that FAK is important in FA turnover by inhibiting Rho activity (9, 10). Subsequent studies portray a more complex picture (11) and indicate that FAK is also involved in increasing adhesion strength, particularly in response to tension forces (12, 13). FAK is frequently overexpressed in various human cancers (14). Its overexpression highly correlates with tumor invasiveness; hence, FAK is widely pursued as a drug target for cancer therapy (15, 16).

FAK is a 120-kDa multidomain protein containing an N-terminal 4.1, ezrin, radixin, moesin homology (FERM) domain, followed by a 50-residue linker, a central kinase domain, an ∼220 residue low-complexity proline-rich region, and a C-terminal focal adhesion targeting (FAT) domain (Fig. 1*A*). Whereas the FAT domain is important for targeting FAK to FAs through interactions with paxillin (17–19), the FERM domain is responsible for regulating catalytic activity (20). In the autoinhibited state, the FERM domain docks onto the kinase domain, which results in catalytic inhibition and sequesters

Significance

Nonreceptor tyrosine kinases are major players in cell signaling. Among them, focal adhesion kinase (FAK) is the key integrator of signals from growth factors and cell adhesion. In cancer, FAK is frequently overexpressed, and by promoting adhesion to the tumor stroma and ECM, FAK provides important signals for tumor invasion and metastasis. Although autoinhibitory mechanisms have previously been described and the players involved in FAK regulation are largely known, on a mechanistic level, FAK activation is currently not understood. Here, we present a multidisciplinary approach demonstrating a multistep mechanism resulting in FAK activation. This mechanistic insight enables the design of alternative strategies for the discovery of potential anticancer drugs that inhibit both catalytic and scaffolding functions of FAK with high specificity.

Author contributions: G.M.G., M.J.E., L.K., F.G., F.L.G., M.P.-M., and D.L. designed research; G.M.G., C.E., J.B., M.C.-A., J.Z., A.B., M.T.M., and D.L. performed research; D.L. contributed new reagents/analytic tools; G.M.G., C.E., J.B., M.C.-A., J.Z., A.B., M.T.M., M.J.E., L.K., F.G., F.L.G., M.P.-M., and D.L. analyzed data; and G.M.G., C.E., J.B., M.J.E., F.G., F.L.G., M.P.-M., and D.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3ZDT and 4CYE).

[1]To whom correspondence should be addressed. Email: dlietha@cnio.es.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1317022111/-/DCSupplemental.

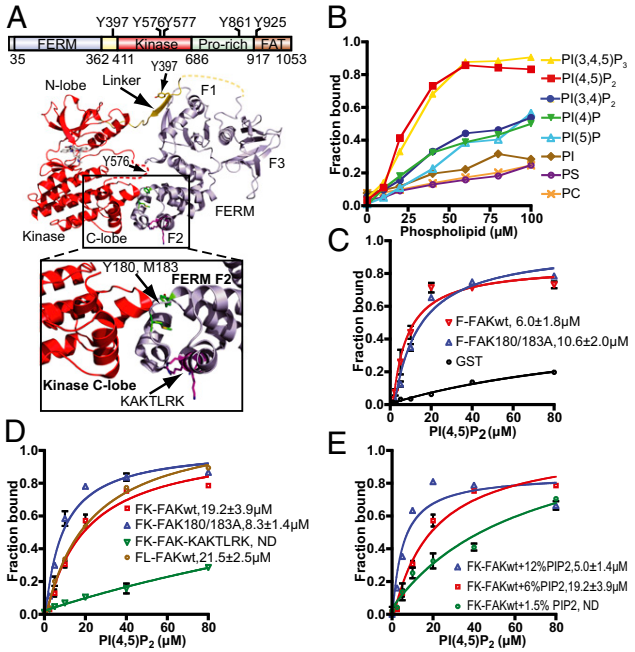


Fig. 1. FAK interacts with PI(4,5)P_2 via the basic patch in the FERM domain. (*A*) Domain structure of FAK with the main phosphorylation sites indicated and a ribbon diagram of the FK-FAK crystal structure as reported by Lietha et al. (21) (PDB ID code 2J0J). In the zoom window, the interaction between the FERM F2 lobe and the kinase C-lobe is shown, with residues Y180 and M183 at the interface colored green and the basic KAKTLRK residues (K216, K218, R221, and K222) colored magenta. (*B*) Lipid binding specificity of FK-FAK was studied using vesicle pull-down assays with PC vesicles containing 6% (mol/mol) of the indicated phospholipids. Phosphorylation of the D4 and D5 positions of the inositol head group confers full binding affinity. PS, phosphatidyl serine. (*C*) Vesicle pull-downs with 6% (mol/mol) PI(4,5)P_2 vesicles and GST-fused F-FAKwt (FERM = FAK31-405) or F-FAK180/183A (Y180 and M183 mutated to alanine). The mutations do not affect PI(4,5)P_2 binding. GST fusions were used to obtain higher readouts. (*D*) PI(4,5)P_2 vesicle pull-downs with FK-FAKwt, FK-FAK180/183A, FK-FAK-KAKTLRK (all KAKTLRK basic residues are mutated to alanine), or FL-FAKwt. Y180/M183A mutations result in ∼2.5-fold higher affinity to PI(4,5)P_2, whereas the KAKTLRK mutations abolish binding. FL-FAK binds with similar affinity to FK-FAK. ND, not determined. (*E*) Vesicle pull-downs with 1.5% (mol/mol), 6% (mol/mol), or 12% (mol/mol) PI(4,5)P_2 and FK-FAKwt. Increasing the PI(4,5)P_2 density on vesicles results in higher affinity for FK-FAKwt, indicating an avidity effect. (*C*–*E*) Error bars represent SD from three independent experiments and are shown if larger than the symbol. K_d values are determined by fitting a one-site binding model (cooperative fitting is shown in Fig. S1 *B* and *C*).

regulatory phosphorylation sites (21). FAK activation initially results in autophosphorylation of Y397 in the linker between the FERM and kinase domains (22), a process reported to occur *in trans* (23). Phosphorylated Y397 provides a docking site for the Src homology 2 (SH2) domain of the Src kinase, and recruited Src phosphorylates several tyrosines in FAK. Two of them (Y576 and Y577) are located in the activation loop of FAK, and their phosphorylation confers full catalytic activity (24). The activated FAK/Src complex phosphorylates several FA proteins, including paxillin (25) and p130Cas (26).

Many stimuli have been reported to initiate FAK activation (reviewed in ref. 27), such as integrin signaling and engagement of growth factor receptors (28–30), but molecular details remain elusive. Further, we showed that acidic phosphoinositides, such as phosphatidylinositol-4,5-bisphosphate [PI(4,5)P_2], interact with FAK and play a role in FAK activation (31). PI(4,5)P_2 is well established as a modulator of FA maturation and adhesion strength, but its role in adhesion signaling is less clear. PI(4,5)P_2 promotes formation of mature FAs by binding talin and vinculin and inducing their open state, where binding sites to other FA proteins and actin are exposed (32, 33). PI(4,5)P_2 is locally generated in FAs by the enzyme phosphatidylinositol 4-phosphate 5-kinase type Iγ (PIP5KIγ) (34, 35), which adds the 5-phosphate to PI(4)P. PIP5KIγ exists as two splice variants (PIP5KIγ661 and PIP5KIγ635 in mice), where the longer form (PIP5KIγ661, PIP5KIγ668 in humans) contains extra C-terminal residues that target the enzyme to FAs by interacting with the talin head domain (36, 37).

Although the main players involved in FAK activation have been identified at a mechanistic level, it is not known how their concerted action is orchestrated to achieve FAK activation. Here, we probe the molecular mechanism of FAK activation using a multidisciplinary approach, including biochemical, structural, in vitro FRET, molecular dynamics (MD) simulation, and cell biology studies, and we present evidence that PIP5KIγ and its product, PI(4,5)P_2, are important mediators of the integrin/FAK signaling link. Intriguingly, we find that PI(4,5)P_2 binding to a basic region on the FAK FERM domain results in clustering of FAK on the lipid membrane, a process likely to enhance integrin clustering as well as the scaffolding function of FAK. PI(4,5)P_2 binding further induces conformational changes between FERM and kinase domains without causing domain dissociation. In vitro FRET experiments, together with MD simulations, support a scenario where distal changes at the PI(4,5)P_2 binding site result in domain opening and exposure of linker regions, which, together with clustering, promote efficient autophosphorylation of Y397 within the linker. FAK autophosphorylation recruits Src, and, in turn, Src is responsible for full activation of FAK and FERM release by phosphorylating the FAK activation loop. The mechanistic insight we present here can aid the design of novel classes of therapeutics targeting both catalytic and scaffolding functions of FAK.

Results

FAK Binds PI(4,5)P_2 via Basic Residues in the FERM F2 Lobe. To characterize the FAK–PI(4,5)P_2 interaction in detail, we performed in vitro binding studies. Using vesicle pull-down experiments, we initially analyzed the phosphoinositide specificity and find that bis-phosphorylation on the D4 and D5 positions of the inositol ring is required for full binding affinity, whereas additional phosphorylation on the D3 position has no effect (Fig. 1*B*). This finding suggests the enzyme PIP5KIγ as a key enzyme regulating FAK signaling because it is the enzyme generating PI(4,5)P_2 in FAs. Other phospholipids, such as PI, phosphatidyl serine, or phosphatidyl choline (PC), only display background levels of binding. The FERM domain of FAK (F-FAK) is sufficient for full PI(4,5)P_2 binding affinity (Fig. 1*C*), and binding is mediated via a conserved basic region in the FERM F2 lobe, as shown by mutation of the basic KAKTLRK sequence (all K/R to A = FAK-KAKTLRK; Fig. 1 *A* and *D*). Interestingly, the FERM + kinase fragment of FAK (FK-FAK) interacts with approximately twofold lower affinity than F-FAK (Fig. 1 *C* and *D*). In addition, a mutant form of FK-FAK, where residues Y180 and M183 in the FERM F2 lobe are mutated to alanine (FK-FAK180/183A), displays, like F-FAK, higher PI(4,5)P_2 affinity than WT (FK-FAKwt) (Fig. 1 *A* and *D*). This observation is supported by the differential association of FK-FAKwt and FK-FAK180/183A to PI(4,5)P_2 vesicles in surface plasmon resonance experiments (Fig. S1*A*). The residues Y180 and M183 are located at the region of the FERM F2 lobe that is responsible for kinase binding and autoinhibition (21) (Fig. 1*A*). Using small-angle X-ray scattering, we show that the 180/183A mutant of FK-FAK adopts a monomeric and open conformation with FERM and kinase domains dissociated (Fig. S2 and Table S1), whereas FK-FAKwt adopts a closed conformation (Fig. S2 and ref. 21). We therefore conclude that the reduced affinity

of closed FK-FAKwt (Fig. 1*D* and Fig. S1*A*) is due to an energetically costly conformational change required to bind PI(4,5)P_2. Further, we find that C-terminal regions of FAK do not affect PI(4,5)P_2 binding, because full-length FAK (FL-FAK) exhibits a similar PI(4,5)P_2 affinity as FK-FAK (Fig. 1*D*). Importantly, the PI(4,5)P_2 affinity of FAK is altered by the PI(4,5)P_2 density on lipid vesicles, indicating an avidity effect (Fig. 1*E*). For simplicity, we use a one-site model to fit our pull-down data presented in Fig. 1 *C*–*E*; however, the data fit well with a cooperative model, indicating positive cooperatively with a Hill coefficient of ~2 (Fig. S1 *B* and *C*).

PI(4,5)P_2 Enhances FAK Autophosphorylation Without Increasing Catalytic Turnover. Next, we tested the effect of PI(4,5)P_2 on FAK activity. Using an autophosphorylation assay, we find that PI(4,5)P_2 vesicles strongly increase the autophosphorylation efficiency, as shown by immunoblotting using an antibody against the autophosphorylation site Y397 (Fig. 2 *A* and *B*), whereas PI(4)P, PI(5)P, and PI(3,4)P_2 have no effect (Fig. S3*A*). As is the case for PI(4,5)P_2 binding, enhanced autophosphorylation requires the basic KAKTLRK region. Using soluble PI(4,5)P_2 [with eight carbons in the acyl chain, C8-PI(4,5)P_2], we measured autophosphorylation by ELISA and find that C8-PI(4,5)P_2 increases the autophosphorylation efficiency of FK-FAKwt to similar levels as observed for FK-FAK180/183A (Fig. 2*C*). Although the PI(4,5)P_2 head group is necessary for this effect [compare C8-PI(4,5)P_2 vs. the C8-PC plot in Fig. 2*C*], the head group alone [Ins(1,4,5)P_3] is not sufficient. Remarkably, neither soluble nor vesicle-embedded PI(4,5)P_2 affects the catalytic activity, as measured by ATP turnover in a kinetic assay with an exogenous substrate (Fig. 2*D* and Fig. S3 *C* and *D*). The fact that dissociation of FERM and kinase domains by mutation does increase ATP turnover, as shown with the FK-FAK180/183A mutant (Fig. 2*D* and Fig. S3 *C* and *D*), suggests that PI(4,5)P_2 binding to FAK does not dissociate the FERM from the kinase domain.

PI(4,5)P_2 Induces FAK Clustering. Because FAK has been shown to autophosphorylate efficiently *in trans* (23), we considered the possibility that PI(4,5)P_2-induced autophosphorylation might be mediated via FAK oligomers. Using negative-stain transmission EM, we show that FAK forms clusters when bound to PI(4,5)P_2 vesicles, as well as bound to soluble C8-PI(4,5)P_2 (Fig. 3*A*). We performed reference-free 2D class averaging of 574 selected FL-FAK/C8-PI(4,5)P_2 particles, which likely represent a main cluster population (Fig. 3*B*). Based on 3D volumes generated from 2D averages or particle dimensions, we determined an approximate particle size of 900 kDa, from which we estimate the presence of approximately six to eight FL-FAK molecules per cluster. Like PI(4,5)P_2 binding, formation of clusters requires the basic KAKTLRK region on the F2 lobe (Fig. 3*C*, *Right*). We further analyzed clustering of F-FAK and FK-FAK by dynamic light scattering (DLS), which confirms a clear increase in molecular weight in the presence of PI(4,5)P_2, both for F-FAK and FK-FAK (Fig. S4 and Table S2). DLS also shows that PI(4,5)P_2 induces significant polydispersity, indicating that the determined size of clusters by EM likely represents a main population induced by C8-PI(4,5)P_2 but that the relevant FAK-PI(4,5)P_2 complex possibly does not adopt a defined oligomerization state. We note that C8-PI(4,5)P_2–induced clusters do not form as a consequence of micelle formation, because we determined a critical micelle concentration (CMC) for C8-PI(4,5)P_2 of 2 mM, an order of magnitude higher than used in our study.

PI(4,5)P_2 Prevents Formation of a Fully Closed Conformation. As described above, our data indicate that PI(4,5)P_2 does not induce dissociation of FERM and kinase domains. However, the fact that the open FK-FAK180/183A mutant exhibits higher affinity for PI(4,5)P_2 than closed FK-FAKwt (Fig. 1*D* and Fig. S1*A*) suggests that closed FK-FAK requires a rearrangement of

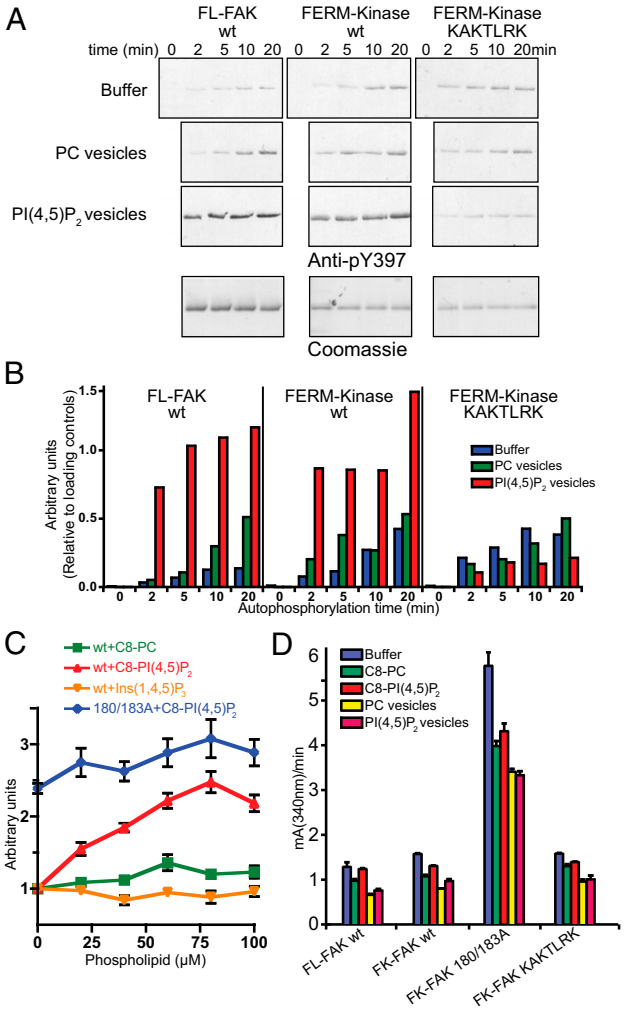


Fig. 2. PI(4,5)P_2 mediates FAK autophosphorylation but not catalytic turnover. (*A*) Autophosphorylation time course of FL-FAK, FK-FAKwt, and FK-FAK-KAKTLRK in the absence (buffer) or presence of PI(4,5)P_2 or PC vesicles was monitored by immunoblotting using an anti-pY397 antibody. (*Lower*) Loading controls stained by Coomassie blue. Note that FL-FAK stains stronger because of its higher molecular weight. (*B*) Quantifications of blots from *A* relative to loading controls [using ImageJ (National Institutes of Health)]. For FL-FAK and FK-FAKwt, autophosphorylation is significantly faster in the presence of PI(4,5)P_2 vesicles, whereas mutations in the basic patch of FK-FAK-KAKTLRK abrogate this effect. (*C*) Autophosphorylation efficiency of FK-FAK (wt or 180/183A mutant) was assessed using an ELISA method. The presence of C8-PI(4,5)P_2, but not C8-PC or the head group Ins(1,4,5)P_3, enhances autophosphorylation of FK-FAKwt to levels similar to FK-FAK180/183A, which was not affected by PI(4,5)P_2. (*D*) Catalytic steady-state activity was assayed for the indicated FAK proteins using a kinase assay, which couples ADP production to NADH consumption (*Methods*). Whereas dissociation of FERM/kinase domains by mutation (180/183A) activates FAK, none of the tested lipids increase catalytic turnover. (*C* and *D*) Error bars represent SD from three experiments.

FERM and kinase domains to allow PI(4,5)P_2 binding. To monitor conformational changes in a controlled environment, we used in vitro FRET experiments utilizing a conformational sensor of FAK. A similar sensor was used previously in cellular studies to demonstrate that FAK undergoes conformational changes in FAs (31). The sensor is based on intramolecular FRET by fusing CFP and citrine N-terminal to the FERM and kinase domains, respectively, and it is designed to report relative domain positions, with autoinhibited (closed) FAK exhibiting

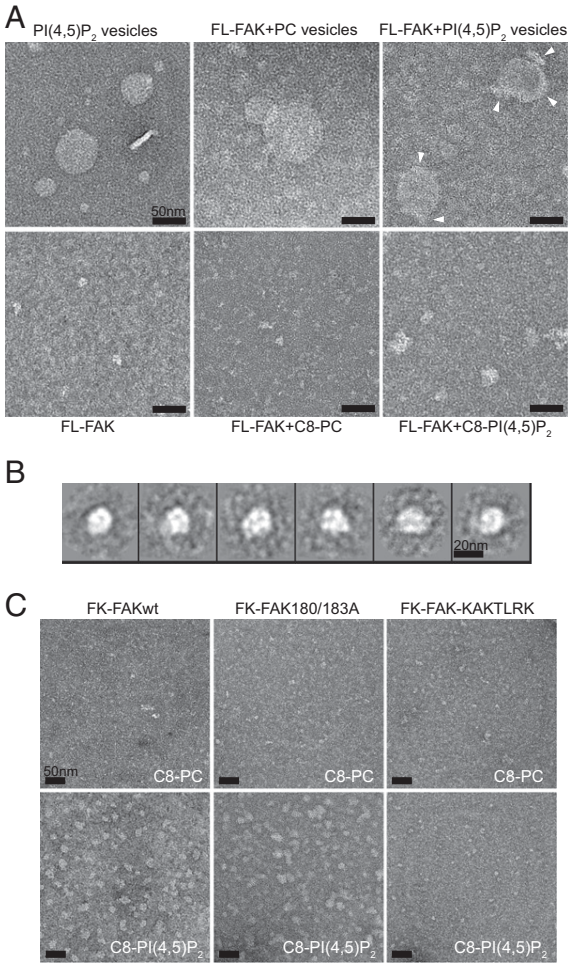


Fig. 3. EM reveals PI(4,5)P_2-induced FAK clustering. (*A*) Transmission electron micrographs of FL-FAK with lipid vesicles (*Upper*) or soluble lipids (*Lower*) imaged by negative staining. PI(4,5)P_2 vesicles and soluble C8-PI(4,5)P_2 mediate the formation of FAK clusters. Clusters on vesicles are indicated by arrowheads. (*B*) Reference-free 2D class averages of 574 FL-FAK/C8-PI(4,5)P_2 clusters, representing a main cluster population, suggest a circular arrangement of FAK molecules in clusters. Volume calculations suggest that clusters consist of six to eight FL-FAK molecules per cluster. (*C*) FK-FAKwt and the 180/183A and KAKTLRK mutants were imaged in the presence of C8-PI(4,5)P_2 or C8-PC. FK-FAKwt and FK-FAK180/183A, but not the KAKTLRK mutant, display clustering in the presence of PI(4,5)P_2. (Scale bars: *A* and *C*, 50 nm.)

high FRET and open forms displaying lower FRET signals. We used three purified versions of the sensor that contain the FERM and kinase regions of FAK (Fig. 4*A*): (*i*) a sensor with CFP and citrine inserted into the WT FAK sequence (CYFAKwt); (*ii*) a sensor that is mutated to adopt an open conformation (CYFAK180/183); and (*iii*) a high-FRET control sensor (CYFAK-HFC) that has the FRET labels in tandem at the N terminus, and hence displays high-FRET signals independent of the intramolecular FAK conformation. Changes in FRET levels of the CYFAK-HFC sensor indicate nonconformational effects, such as trans-FRET or fluorescence quenching (which, for example, could be caused by clustering). As expected, we find lower FRET levels for the open CYFAK180/183A sensor than for CYFAKwt and high-FRET levels for CYFAK-HFC [Fig. 4*B* (leftmost bars in each graph) and Fig. S5*A*]. The addition of C8-PC does not affect FRET levels; however, C8-PI(4,5)P_2 induces a reduction in FRET signals for all three sensors (Fig. 4 *B* and *C*). Because PI(4,5)P_2 also affects CYF-HFC, which reports nonconformational effects, we conclude that under these conditions (without ATP, see below), PI(4,5)P_2 does not significantly alter the conformation of FAK.

Interestingly, when we added ATP and Mg^{2+} to CYFAKwt, the FRET signal increased almost to the level of CYF-HFC (Fig. 4 *B* and *C*). This effect is not seen with CYFAK-HFC, and is therefore likely conformational, with higher FRET signals suggesting a more closed conformation in the presence of ATP. This increase in FRET is not due to a phosphorylation event, because the same effect is seen with the nonhydrolyzable ATP analog 5′-adenylylimidodiphosphate (AMP-PNP) (Fig. S5*B*). Importantly, the high-FRET state with ATP is not observed in the presence of PI(4,5)P_2 (Fig. 4 *B* and *C*, *Left*) and, in fact, is reverted in a concentration-dependent manner by PI(4,5)P_2 if added after ATP (Fig. S5*D*). This indicates a conformational effect of PI(4,5)P_2 in the presence of ATP. Because ATP is present in the cell at similar concentrations that we used in our experiments (1 mM), the high-FRET state observed in the presence of ATP likely represents the basal conformation of FAK, which is affected by PI(4,5)P_2. When comparing FRET levels of CYFAKwt in the presence of ATP with and without PI(4,5)P_2, we can partition the FRET change into a nonconformational effect of PI(4,5)P_2, which is also seen without ATP (NC in Fig. 4*B*), and a larger conformational effect (C in Fig. 4*B*). The nonconformational effect is linear (CYFAK-HFC plot in Fig. S5*D*, *Left*), which can be corrected to observe only the conformational contribution of PI(4,5)P_2 (Fig. S5*D*, *Right*). At high PI(4.5)P_2 concentrations, CYFAKwt in the presence of ATP approaches FRET levels without ATP. The FRET change even at high PI(4,5)P2 concentrations does, however, not correspond to full domain dissociation (see below). Together, these data suggest that in the presence of ATP, PI(4,5)P_2 induces partial domain opening.

In experiments where Src was added to CYFAKwt and ATP, the initial increase in FRET is followed by a switch to an open conformation (lower FRET signals; Fig. 4*C*, *Left*), whereas only a minor effect is observed for the CYFAK-HFC control. The kinase-dead mutant SrcK298M has a small effect on FRET levels in presence of ATP (Fig. S5*C*), indicating that Src phosphorylation is mainly responsible for switching CYFAKwt to the open conformation and that Src binding has a minor effect. This is consistent with observations that Src phosphorylation of the FAK activation loop is incompatible with FERM inhibition (21). Comparing FRET signals of different states corroborates that PI(4,5)P_2 does not induce full FERM/kinase dissociation, which is observed only upon Src phosphorylation or 180/183A mutation.

Allosteric Effects of PI(4,5)P_2 on the FERM/Kinase Interface. Because the two opposite effects of PI(4,5)P_2 and ATP appear to be related, we first proceeded to understand the effect of ATP better. Using MD simulations, we monitored backbone fluctuations of the solvated FAK domain in the presence or absence of ATP over 1,500 ns. We find that the presence of ATP in the active site has the largest stabilizing effect in the αC- and αG-helices of the FAK domain (Fig. 5*A*). Strikingly, these two sites exactly map the autoinhibitory interaction sites with the FERM domain, as seen in the crystal structure of FK-FAK (21) (Fig. 5*B*). In accordance with FRET and the simulation data, we therefore propose that binding of ATP to FAK induces a tightly closed FERM/kinase conformation by rigidifying the interaction interfaces. Moreover, by performing an elastic network-based allosteric connectivity analysis of the FAK domain, we find, by a completely independent computational approach (38), a strong allosteric coupling between αC-and αG-helices despite their distal locations (Fig. S6).

We then proceeded to probe conformational effects of PI(4,5)P_2 by performing MD simulations of (*i*) FK-FAK alone, (*ii*) FK-FAK bound to soluble PI(4,5)P_2 [C2-PI(4,5)P_2], or (*iii*) FK-FAK with the basic patch mutant FK-FAK-KAKTLRK. Interestingly, we

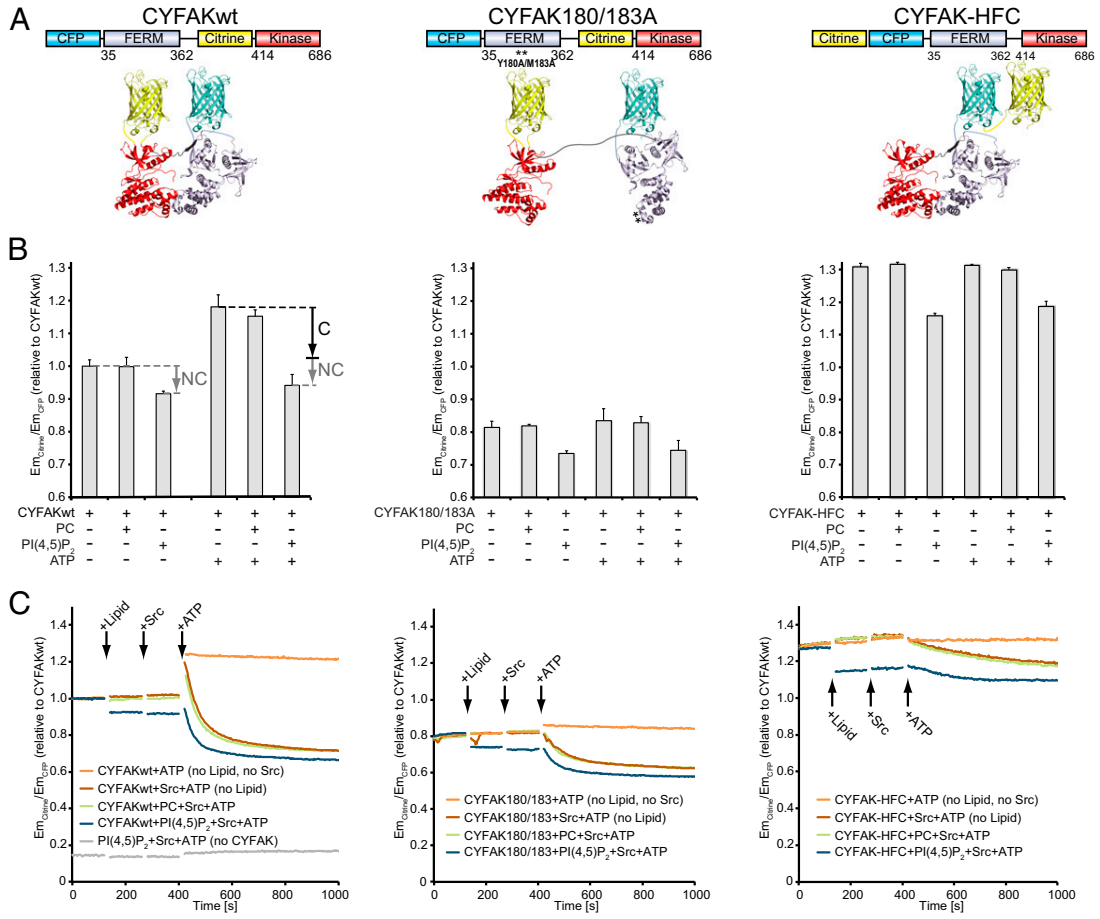


Fig. 4. $PI(4,5)P_2$ induces partial and Src induces full domain opening of the FAK FRET sensor in the presence of ATP. (*A*) Schematic illustration of the domain structure (*Upper*) and expected structural arrangement (*Lower*) of the conformational FRET sensors used. (*B*) Emission ratios of citrine ($Em_{Citrine}$) and CFP (Em_{CFP}) are plotted relative to CYFAKwt as a measure of FRET. The presence of C8-$PI(4,5)P_2$ causes a reduction in FRET levels for all three sensors, suggesting that this effect is not conformational (labeled NC). The presence of ATP significantly increases FRET levels for CYFAKwt; however, this increase is reversed in the presence of $PI(4,5)P_2$. This effect of ATP and $PI(4,5)P_2$ is not seen for CYF-HFC, and is therefore likely conformational (labeled C). Error bars represent SD from a minimum of three experiments. (*C*) Relative FRET levels were monitored in real time, initially of the sensors alone and then following addition of (*i*) lipid [C8-PC (green plots) and C8-$PI(4,5)P_2$ (blue plots)] or no lipid (brown and orange plots), (*ii*) Src (brown, green, and blue plots) or no Src (orange plots), and (*iii*) ATP/Mg^{2+} (all plots). As in *B*, $PI(4,5)P_2$ reduces FRET levels of all sensors in all states (before/after phosphorylation), indicating a nonconformational effect. ATP results in a FRET spike only with CYFAKwt. If active Src is present, the spike is followed by a switch to the open conformation (lower FRET levels). In contrast, inactive SrcK298M induces only a modest FRET decrease (Fig. S5*C*). (*Left*) Gray control plot is without the CYFAK sensor to verify that $PI(4,5)P_2$, Src, and ATP do not exhibit intrinsic fluorescence.

find that $PI(4,5)P_2$ binding results in altered interresidue forces within the FERM F2 lobe that originate at the basic patch and propagate to the interface with the kinase C-lobe (Fig. S7*A*, *Upper*). In particular, $PI(4,5)P_2$ binding interferes with a set of stabilizing salt bridges between the basic patch (K218/R221) and the adjacent helix at the domain interface (E195/E198, Fig. 5 *C* and *D*). These $PI(4,5)P_2$-induced changes also affect the loop containing the autoinhibitory residues Y180 and M183. This effect was reproduced in MD simulations of FK-FAK-KAKTLRK [Fig. 5*D* and Fig. S7 *A* (*Lower*) and *B*], suggesting that neutralizing the charges of the basic KAKTLRK residues by mutation has a similar effect on the FERM F2 lobe interatomic forces as $PI(4,5)P_2$ binding. We validated our model for the $PI(4,5)P_2$–FERM F2 lobe interaction by performing MD simulations with $PI(4,5)P_2$ embedded in a lipid bilayer, and we find that the mode of the FERM–$PI(4,5)P_2$ interaction is highly similar. $PI(4,5)P_2$ remains embedded in the bilayer (at least within the time scale of the simulations) such that the $PI(4,5)P_2$–FAK interaction is restricted to the head group; therefore, like with short-chain lipids, MD simulations primarily provide details on a charge neutralization effect.

In the absence of a FAK-$PI(4,5)P_2$ cocrystal structure, which is complicated by clustering, as seen in Fig. 3 and Fig. S4, we pursued a crystal structure of the basic patch mutant FERM domain (F-FAK-KAKTLRK, crystallographic Table S3). Although the FERM domain retains a very similar conformation to F-FAKwt overall, in both independent molecules in the asymmetrical unit cell, no electron density is observed for the loop Y180-K190 (containing the autoinhibitory residues Y180/M183), indicating that the loop is disordered (Fig. 5*E* and Fig. S7*C*). This is in striking agreement with observations from the MD simulations described above, which find a destabilized F2 lobe upon neutralization of the basic region in the FERM F2 lobe (Fig. 5 *C* and *D* and Fig. S7 *A* and *B*). We note that the disorder could potentially be due to less stable crystal contacts, and, in fact, one F-FAKwt structure [Protein Data Bank (PDB) ID code 2AL6] does exhibit high B-factors in this region; however, the most similar packing environment is observed in F-FAKwt with PDB ID codes 4CYE and 2AEH (chain A),

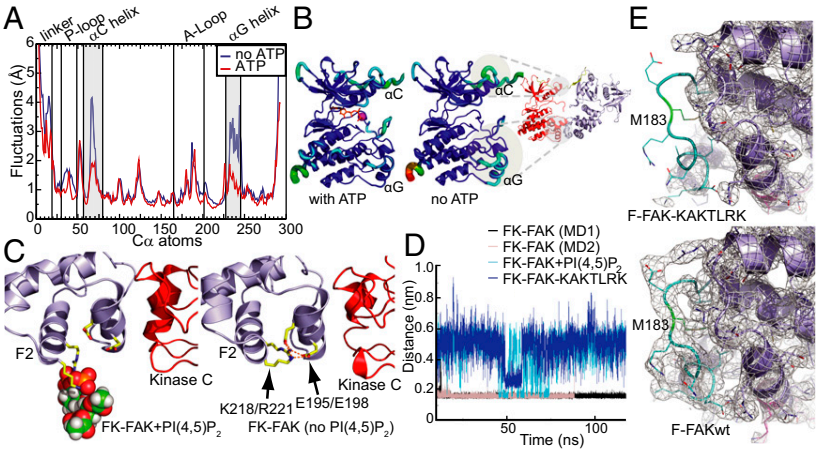


Fig. 5. Effects of ATP and PI(4,5)P_2 on FAK conformation. (*A*) Root mean square fluctuations (RMSFs) from unbiased MD simulations of the FAK kinase domain are shown for a window of 700 ns (discarding the first 200 ns of equilibration) for FAK alone (blue plot) and for FAK with ATP/Mg^{2+} (red plot). ATP binding stabilizes the αC- and αG-helices. (*B*) RMSF values from *A* are color-mapped onto the FAK structure. RMSFs range from blue (low values) to red (high values). Stabilizing ATP effects on αC- and αG-helices map autoinhibitory interaction sites as seen in the FK-FAK crystal structure (21). Neutralization of the basic patch by PI(4,5)P_2 binding (*C*) or by mutation (*D*) interferes with a set of FERM F2 lobe-stabilizing salt bridges. MD simulations suggest that salt bridges between K218/R221 in the basic patch and E195/E198 are formed in the absence of PI(4,5)P_2 (*C*, *Right*) but not in the presence of PI(4,5)P_2 (*C*, *Left*), leading to a partial destabilization of the FERM F2 lobe and an altered force distribution, as shown in Fig. S7*A*. (*D*) Minimum distances for the residues K218/R221 and residue pair E195/E198 are shown during MD simulations with FK-FAK alone (MD1 and MD2), FK-FAK bound to PI(4,5)P_2, or the basic patch mutant FK-FAK-KAKTLRK. In the two independent MD simulations of FK-FAKwt, these two pairs of oppositely charged residues strongly interact with each other. PI(4,5)P_2 binding or KAKTLRK mutations to alanines significantly increase the minimum distances and fluctuations. (*E*) Crystal structure of the basic patch mutant FERM domain (F-FAK-KAKTLRK; PDB ID code 3ZDT; full structure is shown in Fig. S7*C*). In contrast to F-FAKwt (*Lower*, PDB ID code 4CYE), the structure of F-FAK-KAKTLRK (*Upper*) exhibits no electron density for the loop (cyan) containing the autoinhibitory residues Y180/M183 (green), indicating that this loop is disordered. The 2Fo-Fc electron density maps are shown as gray mesh countered at 1σ.

where the loop is ordered (Fig. S7*D*). In conclusion, we propose that neutralization of the basic patch by specific binding of PI(4,5)P_2 destabilizes the FERM F2 lobe, resulting in an altered FERM F2/kinase C-lobe interface, which allows partial domain opening.

PI(4,5)P_2 Enhances Src-Mediated FAK Activation. FAK activation proceeds through a multistep mechanism. PI(4,5)P_2-induced FAK autophosphorylation is followed by Src recruitment to the autophosphorylation site and phosphorylation of the activation loop of FAK (residues Y576 and Y577) by Src. Because Src recruitment requires FAK autophosphorylation, it can be expected that PI(4,5)P_2 should also enhance Src-mediated phosphorylation of the activation loop of FAK. To test this, we monitored Src phosphorylation of Y576 in the presence or absence of PI(4,5)P_2 by immunoblotting, using an antibody against phospho-Y576, following an autophosphorylation step (Fig. 6). To uncouple autophosphorylation from Src phosphorylation, autophosphorylation was stopped after 2 min with the specific FAK inhibitor TAE226 before starting Src reactions (details are provided in *Methods*). As expected, we find that Y576 phosphorylation is more efficient in the presence of PI(4,5)P_2, whereas other tested phosphoinositides have no significant effect (Fig. S3*B*). In part, this effect is due to enhanced FAK autophosphorylation, because in the absence of autophosphorylation (for FK-FAK Y397F), the Y576 phosphorylation efficiency of Src is reduced. However, PI(4,5)P_2 also enhances Src phosphorylation of the FK-FAK Y397F mutant, suggesting an autophosphorylation-independent effect. These data are consistent with PI(4,5)P_2 inducing a partially open conformation that allows easier access to Y576.

PI(4,5)P_2 in FAs Promotes FAK Signaling and Cell Adhesion. We have previously proposed acidic phospholipids to be involved in FAK activation in cells (31); however, the lipid specificity (and hence the relevant upstream components) was not defined. Because we demonstrate the requirement of D4 and D5 phosphorylation of phosphoinositides (Fig. 1*B*), and because PIP5KIγ is known to be responsible for PI(4,5)P_2 generation in FAs, we performed PIP5KIγ loss-of-function experiments in HeLa cells to establish the role of PI(4,5)P_2 in FAK signaling in FAs. The expression of PIP5KIγ was knocked down using a specific shRNA against isoform 2 of PIP5KIγ (the longer PIP5KIγ668 isoform), and two different stable knockdown (KD) clones were selected (KD1 and KD2). The KD efficiency was very high, as observed by Western blotting and immunofluorescence (Fig. 7*A* and Fig. S8*A*). The total levels of FAK are not changed in the absence of PIP5KIγ (Fig. 7*B*). However, Y397 autophosphorylation and Y576/Y577 Src phosphorylation are significantly reduced at early stages of focal adhesion formation, compared with the scramble controls (Fig. 7*B*). Overall, these data suggest that PIP5KIγ may function in the proper activation of FAK. To validate the causal role of

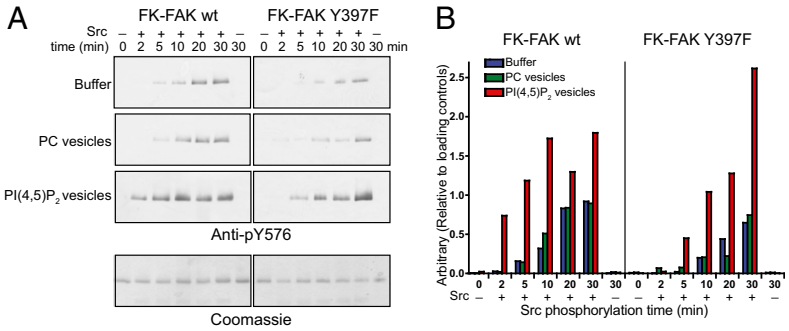


Fig. 6. PI(4,5)P_2 enhances Src phosphorylation of Y576 in FAK. (*A*) Time course of Src phosphorylation of FK-FAK in the absence or presence of PI(4,5)P_2 or PC vesicles as monitored by Western blotting using an anti-pY576 antibody following a 2-min autophosphorylation reaction (*Methods*). Mutation of the autophosphorylation site (Y397F) significantly reduces the Y576 phosphorylation rate, whereas PI(4,5)P_2 enhances Y576 phosphorylation for FK-FAKwt and FK-FAK-Y397F, indicating a combined effect of more efficient Src recruitment to the FAK autophosphorylation site and PI(4,5)P_2-induced conformational changes. (*Lower*) Coomassie blue-stained loading controls are shown. (*B*) Quantifications of blots in *A*.

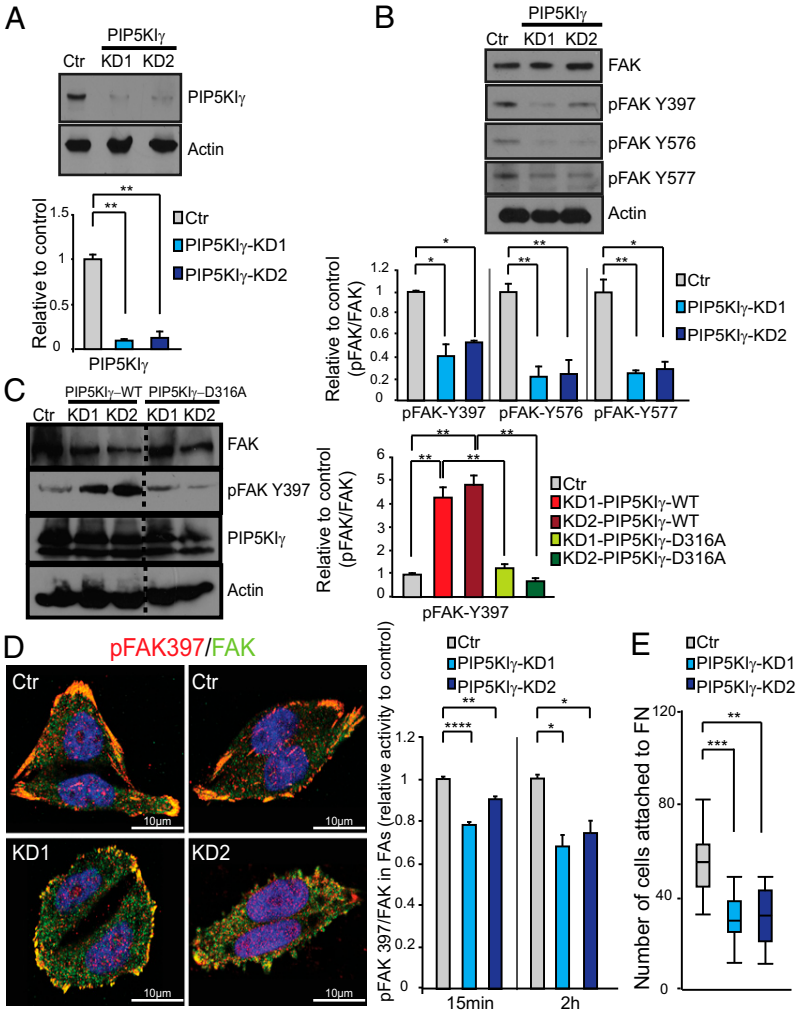


Fig. 7. PIP5KIγ KD decreases cellular FAK activity and cell attachment. (*A*) Expression levels of PIP5KIγ668 (isoform 2) in HeLa cells after KD are shown for two different clones (KD1 and KD2) and scramble control (Ctr). (*Upper*) Representative immunoblot is shown. (*Lower*) Quantification of PIP5KIγ levels observed by immunoblotting is depicted in the histogram. Quantifications are from four blots (two independent experiments, each in duplicate). (*B*) Effect of PIP5KIγ reductions on the total level of FAK and on the total cellular levels of pFAK Y397, pFAK Y576, and pFAK Y577. A representative immunoblot and quantifications from four blots (two independent experiments, each in duplicate) are shown. (*C*) pFAKY397 levels in HeLa cells of Ctr or KD1 and KD2 cells transfected with WT (PIP5KIγ-WT) or a kinase-dead (PIP5Kγ-D316A) form of PIP5KIγ. Immunoblots and quantifications of two independent experiments are shown, with each performed twice. Note that PIP5KIγ-WT more than rescues but the kinase-dead mutant does not fully rescue FAK phosphorylation levels in KD cells compared with controls. The dashed line merges different lanes of the same immunoblot experiment. (*D*) Immunofluorescence staining of pFAK397 (red) and total FAK (green) in PIP5KIγ KD1 and KD2 clones and Ctr, and quantification of the immunofluorescence intensity of pFAK397 relative to total FAK signals specifically in FAs at 15 min and 2 h after stimulation with serum and fibronectin. (*E*) Functional cell adhesion assay to determine the ability of PIP5KIγ-deficient cells to attach to fibronectin (FN) compared with Ctr. Quantification of the number of attached cells for three independent experiments in triplicate is shown. Data represent the mean value ± SEM. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$ (unpaired Student *t* test).

PIP5KIγ in the activation of FAK, we performed rescue experiments using vectors encoding either a WT (PIP5KIγ-WT) or kinase-dead (PIP5Kγ-D316A) form of PIP5KIγ. As shown in Fig. 7*C*, on a PIP5Kγ-deficient background, expression of PIP5KIγ-WT not only restores but increases the phosphorylation levels of pFAKY397 compared with the scramble controls. By contrast, the expression of the PIP5Kγ-D316A kinase-dead mutant in PIP5Kγ-deficient cells is not able to rescue pFAKY397 levels fully (Fig. 7*C*). Overall, these data provide evidence that PIP5Kγ activity is required for the activation of FAK, at least in certain cellular contexts.

Immunofluorescence analysis revealed that the effect of PIP5Kγ KD on the activation of FAK is not due to impaired recruitment of FAK to FAs (Fig. S8*B*). FAK is recruited to FAs in KD cells from early time points of stimulation, although the pFAKY397 levels are reduced compared with controls (Fig. 7*D*). Interestingly, KD cells display smaller FAs (Fig. S8*C*). To analyze the functional relevance of PIP5Kγ in FAs, we further evaluated the ability of PIP5Kγ-deficient cells to attach to fibronectin-coated plates at early stages of cell adhesion and spreading. In comparison to controls, fewer PIP5Kγ-deficient cells attach to the plates (Fig. 7*E*), indicating that PIP5Kγ is required to promote the adequate adhesive properties of FAs. Globally, these data establish PIP5KIγ and its product $PI(4,5)P_2$ as upstream activators of FAK in FAs and suggest they play important roles in FA maturation, cell attachment, and spreading.

Discussion

All NRTKs use remarkably similar concepts to switch kinase activity off. In all known cases, regulatory domains N-terminal to the tyrosine kinase domain bind the kinase to induce autoinhibition, and this sequesters regulatory phosphorylation sites in many cases. Also, initial activation steps often follow similar principles. For Src, Abl, and Syk family tyrosine kinases, activation is initiated by activator binding to their regulatory domains, which induces release of the regulatory domain from the kinase, resulting in activation. In the case of FAK, autoinhibition is achieved, as for other NRTKs, by the regulatory FERM domain docking onto the kinase domain (21). It was widely assumed that activation would also follow the canonical path of ligand-induced release of the FERM domain. This was partially based on studies showing that FAK undergoes large conformational changes in FAs (31). Here, we show that with respect to the activating $PI(4,5)P_2$ ligand, the situation is different for FAK. FAK is in many ways an outsider among NRTKs, in that it is localized to a highly dense and clustered environment (FAs), where it has been proposed to act not only as an active kinase but also as a scaffolding protein. Our data indicate that $PI(4,5)P_2$-induced activation of FAK follows concepts distinct from other NRTKs that are adapted to a highly crowded environment and allows concerted scaffolding and catalytic function.

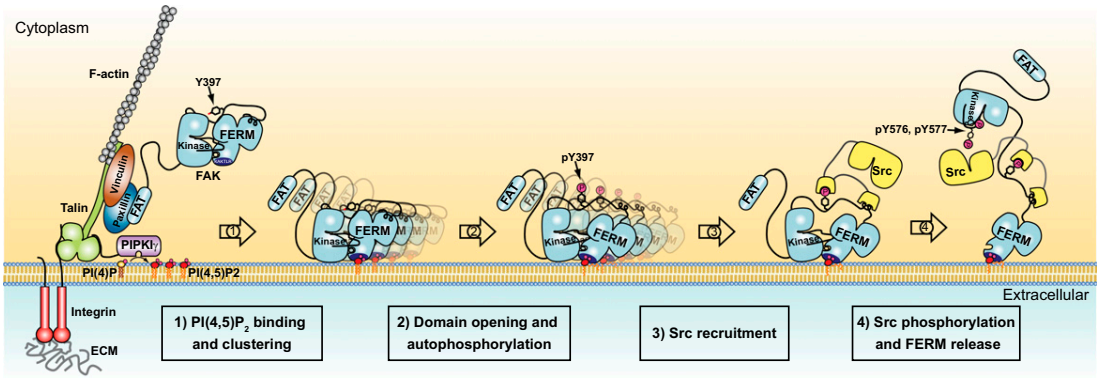


Fig. 8. Schematic model for integrin-mediated FAK activation by PI(4,5)P_2. (*Left*) Cell adhesion via integrin receptors to the ECM results in integrin clustering and the recruitment of FA proteins (as illustrated here for talin, vinculin, paxillin, FAK, and PIP5KIγ) to form adhesion structures that link integrins to the actin cytoskeleton. Recruitment of PIP5KIγ results in a local increase of PI(4,5)P_2 levels in FAs. PI(4,5)P_2 in FAs binds FAK via the basic patch (dark blue) in the FERM domain of FAK, resulting in FAK clustering at the cell membrane (step 1). PI(4,5)P_2-induced FAK clustering results in a relaxed FERM/kinase conformation, with the kinase N-lobe dissociated from the linker and FERM F1 lobe. PI(4,5)P_2-induced clustering and conformational relaxation allow efficient autophosphorylation of Y397 (step 2) and Src recruitment via SH2 and SH3 domains (step 3). Recruited Src phosphorylates the activation loop residues Y576/Y577 of FAK, which results in full activation and release of the kinase from the membrane-clustered FERM domain (step 4).

Our findings support a rather complex model of sequential FAK activation, presented in Fig. 8. First, we establish PI(4,5)P_2, generated in FAs by PIP5KIγ, as an important signaling messenger upstream of FAK activation. We had previously shown that acidic phospholipids, including PI(4,5)P_2, bind FAK (31); however, the phospholipid specificity was not established. Here, we show that both the D4 and D5 phosphates in PI(4,5)P_2 are required for full binding affinity (Fig. 1*B*). Although PI(3,4,5)P_3 exhibits a similar affinity, PI(4,5)P_2 is likely the relevant signaling lipid in FAs due to its abundance. It has been a long-standing question in the field as to how integrin signaling links to FAK activation. Because integrin signaling specifically results in PIP5KIγ recruitment and local PI(4,5)P_2 production (34, 35), and we show that PIP5KIγ is an upstream activator of FAK in HeLa cells (Fig. 7), we propose that the path of the integrin-FAK signaling link leads via PIP5KIγ and generation of PI(4,5)P_2 (Fig. 8, *Left*). However, we note that due to the complexity of FAK signaling, the exact upstream signaling events likely vary in different tissue contexts. In platelets, for example, FAK signaling does not depend on PIPKIγ (39). We note that PI(4,5)P_2 levels are generally not as tightly regulated as, for example, PI(3,4,5)P_3 levels; however, the property of the FAK–PI(4,5)P_2 interaction appears well suited to the characteristics of PI(4,5)P_2 levels in the cell and the localized generation of PI(4,5)P_2. The avidity effect we observe with increased PI(4,5)P_2 density (Fig. 1*E* and Fig. S1*C*) will prevent responding to basal PI(4,5)P_2 levels (average estimates in the cell are 1% of total lipid) but allows a sharp and specific response to highly localized PI(4,5)P_2 production in FAs by PIP5KIγ. On the other hand, Legate et al. (34) propose that bulk PI(4,5)P_2 can partially substitute for local PI(4,5)P_2 production, which may explain the differential effects of PIP5KIγ depletion in different cell types (39). We note that lipid involvement in FAK activation fits well with superresolution optical microscopy studies showing that FAK in FAs resides in close proximity to the cell membrane (40).

We demonstrate that binding to PI(4,5)P_2 induces FAK clustering [Fig. 3 and Figs. S4 and S8 (step 1)]. Together with findings showing FAK autophosphorylation to occur *in trans* (23), this suggests FAK colocalization as an important mechanism promoting FAK transautophosphorylation, similar to what is proposed for receptor tyrosine kinases (41). Further, because FAK oligomers are multivalent PI(4,5)P_2 binders, clustering likely explains the avidity effect we observe (Fig. 1*E* and Fig. S1*C*), although we cannot rule out multiple PI(4,5)P_2 binding sites on a single FAK molecule. Clustering together with observed conformational changes raises the possibility that PI(4,5)P_2 binding occurs in a cooperative mode. Indeed, our vesicle pull-down data fit well with a cooperative model, indicating positive cooperativity with a Hill coefficient close to 2 (Fig. S1 *B* and *C*). The observed cooperativity could be due to induced conformational changes that increase PI(4,5)P_2 affinity or could occur because the induced clustering increases affinity due to the avidity effect. FAK clusters with soluble PI(4,5)P_2 appear globular in shape (Fig. 3*B*), and we note that 2D averaged images suggest that FAK molecules are circularly arranged with less density at the center and an opening at one side of the circle. The size of FAK clusters on lipid vesicles appears more extended. Possibly, in the case of FAK clustering on a 2D membrane, the circular FAK arrangement seen with soluble PI(4,5)P_2 is opened, allowing further propagation of FAK clusters along the membrane. It is currently unclear how FAK clustering is mediated, but because we observe clustering with soluble PI(4,5)P_2 below its CMC concentration, direct protein–protein interactions among FAK molecules are likely involved. A potential site is the conserved W266 and a neighboring pocket involving L327, which are involved in lattice contacts in almost all FAK crystals containing the FERM domain. Interestingly, a recent report suggests W266 to be involved in activating FAK dimer formation (42). Potentially, this dimer could be stabilized in PI(4,5)P_2-induced clusters; however, it is likely that additional contacts are required to form the more extended clusters we observe.

As suggested from the multiple binding partners of FAK, the clustering also underlines the importance of the scaffolding function of FAK and suggests that FAK may play an important role in FA architecture, and perhaps aids integrin clustering. Consistent with this notion, we observe that PIP5KIγ KD results in smaller FAs (Fig. S8*C*) and weaker cell adhesion (Fig. 7*E*), although this effect is not necessarily related to FAK, because PI(4,5)P_2 affects several other focal adhesion proteins, such as talin. A role of FAK in FA maturation and adhesion strength would seem to contradict early studies that implicated FAK in FA turnover (8, 10). However, it is becoming clear that adhesion dynamics are highly complex, and more recent studies find an important role for FAK in increasing adhesion strength (43), particularly in response to rigid substrates and increased tension forces (13).

The PI(4,5)P_2-induced reduction in FRET of CYFAKwt from a closed state in the presence of ATP to an intermediate FRET level (Fig. 4 and Fig. S5*D*) indicates that PI(4,5)P_2, in addition to

clustering, induces partial FERM/kinase domain opening. These FRET data can be explained by two possible scenarios: (*i*) $PI(4,5)P_2$ induces an alternative conformation, which is partially open but with FERM and kinase domains still associated, or (*ii*) $PI(4,5)P_2$ partially shifts the equilibrium between the fully open and fully closed states. Our activity data suggests that model *i* is correct, because a shift in equilibrium toward the open state should result in increased catalytic turnover, which we do not observe (Fig. 2*D* and Fig. S3 *C* and *D*). Because domain association is controlled by the FERM F2/kinase C-lobe interface (Fig. S2), we propose that this interface is intact upon $PI(4,5)P_2$ binding. Taking all our data together, we propose that domain opening occurs at the FERM F1/linker/kinase N-lobe linkage. Opening at this site is expected to increase the distance of the FRET labels, without inducing dissociation of the FERM and kinase domains. Further, it will expose the linker for efficient autophosphorylation (Fig. 8, step 2). A recent study by Ritt et al. (44) finds that modifying the linker length greatly affects Y397 autophosphorylation without having a large effect on catalytic activity. This illustrates that similar to what we find for $PI(4,5)P_2$, autophosphorylation and catalytic activity can be decoupled and that the linker conformation has an important effect on autophosphorylation efficiency. An intriguing question is how binding of $PI(4,5)P_2$ to the distal basic patch on the FERM F2 lobe could affect the FERM F1/linker/kinase N-lobe interface. Our studies support a scenario where neutralization of the KAKTLRK region by $PI(4,5)P_2$ binding results in partial destabilization of the FERM F2 lobe (Fig. 5 *C*–*E* and Fig. S7), allowing higher mobility at the FERM F2/kinase C-lobe interface, which, affects the kinase N-lobe linkage to the linker and FERM F1 lobe through increased domain movement or an allosteric coupling (Fig. S6). We note that because neither KAKTLRK mutation nor head group binding alone induces enhanced autophosphorylation (Fig. 2 *A*–*C*), it appears that, although necessary, charge neutralization is not sufficient, and it is likely that other parts of the lipid, such as triglyceride parts or acyl chains, are involved in clustering and/or promoting conformational changes. Therefore, our MD simulations with mainly charge-based $PI(4,5)P_2$ interactions appear not to capture the full effect of $PI(4,5)P_2$ but rather an initial charge neutralization of $PI(4,5)P_2$ phosphates binding to the basic KAKTLRK residues. We do currently not understand how other parts of the $PI(4,5)P_2$ lipid are involved in FAK conformational changes and clustering.

If the FERM F2/kinase C-lobe interface stays intact upon $PI(4,5)P_2$ binding, this raises the question of how autophosphorylation can occur efficiently in a state where steady-state kinetics are inhibited. As described by Grant and Adams (45), in kinase–substrate complexes and in scenarios where the enzyme/substrate ratio is close to 1 (both are the case here), phosphorylation rates are not determined by steady-state kinetics but rather by pre-steady-state kinetics [reviewed by Taylor et al. (46)]. In such conditions, each enzyme only has to phosphorylate one substrate and can then stay attached to the substrate (off rate is irrelevant). Accordingly, we propose that in $PI(4,5)P_2$-induced clusters, FAK molecules arrange themselves in an ideal position to phosphorylate their neighboring FAK molecule on the linker Y397 site.

Once FAK is autophosphorylated, Src is recruited via its SH2 and SH3 domains (Fig. 8, step 3). Our data presented in Fig. 6 show that $PI(4,5)P_2$, in addition to mediating FAK autophosphorylation, enhances Src phosphorylation of Y576 in the activation loop of FAK. Our data indicate that this is achieved by a combined effect of increased recruitment through autophosphorylation and easier access to Y576 due to conformational changes. Using our FRET sensor, we show that Src-mediated phosphorylation of the FAK activation loop results in full dissociation of the FERM domain [Figs. 4*C* and 8 (step 4)], which is fully consistent with the observed structural incompatibility of FERM inhibition and Y576 phosphorylation (21). Therefore, $PI(4,5)P_2$, via promoting autophosphorylation and Src phosphorylation, can trigger full FAK activation. However, it is clear that additional FAK activators are likely to be important, perhaps depending on the cellular setting. Cytoplasmic portions of several growth factor receptors have been reported to activate FAK (28–30), and direct mechanisms have been proposed for some of them that involve the KAKTLRK region of FAK (28, 29). Phosphorylated tails of the c-Met receptor are reported to interact with the KAKTLRK region, resulting in phosphorylation of Y194 in FAK (47). Intriguingly, Y194 is completely buried within the FERM F2 lobe, indicating that perhaps phospho-Met tail interactions could induce similar destabilization of the FERM F2 lobe as we describe for $PI(4,5)P_2$ (Fig. 5 *C*–*E* and Fig. S7), hence exposing Y194 for phosphorylation. However, details of growth factor-mediated FAK activation remain to be determined, particularly whether they also involve $PI(4,5)P_2$. Interestingly, EGF-stimulated cell migration was reported to depend on PIP5KIγ, indicating an important role for $PI(4,5)P_2$ generation in the case of EGF receptor signaling (48). On the other hand, using FRET sensors similar to the ones used in our study, Ritt et al. (44) demonstrate that FAK can be conformationally activated by increased pH, which could be a route by which cancer cells could activate FAK ligand independently. Yet another signal involved in FAK activation could be tension forces generated by actomyosin contraction, because FAK was described as a force sensor (12, 13). It is probable that different routes resulting in FAK activation are active, depending on the cellular context. The mechanism we propose is likely restricted to situations where basal $PI(4,5)P_2$ concentrations are limiting FAK signaling, and activation can therefore be triggered by local $PI(4,5)P_2$ production.

In conclusion, the combination of our experimental approach with computational simulations has allowed us to propose a complex and detailed multistep activation mechanism for FAK. We propose that $PI(4,5)P_2$ activates FAK via a combination of clustering and conformational changes that promote efficient autophosphorylation; Src recruitment; and, in turn, full activation by Src phosphorylation. Our findings fit well with observations from numerous cell biology studies that have described FAK signaling at the cell membrane in a highly crowded environment with dual catalytic and scaffolding functions. Our insights into the allosteric regulation and intramolecular communication of FAK could have major implications for the development of new classes of allosteric cancer therapeutics targeting FAK regulatory mechanisms that affect both catalytic and scaffolding functions of FAK.

Methods

Additional methods, including structure solution, MD simulations, docking of $P(4,5)P_2$, and cell biology, are provided in *SI Methods*.

Vesicle Pull-Downs. Lipid vesicles were prepared by mixing dissolved phospholipids (Avanti Polar Lipids) to obtain 6% (mol/mol) phosphoinositide in PC (or as indicated). Solvent was removed on a rotary evaporator; lipid films were hydrated with 20 mM Hepes (pH 7.5), 150 mM NaCl, and 1 mM Tris (2-carboxyethyl) phosphine (TCEP), and were sonicated until resuspended. Vesicle pull-downs were performed with 2 μM FAK proteins and serial dilution of vesicles. After 2 h of incubation, vesicles were centrifuged for 30 min at 16,100 × *g* at 4 °C. Supernatants were collected, and pelleted vesicles were washed and resuspended. Proteins in pellets and supernatants were quantified by the Bradford method (Biorad), reading absorption at 590 nm. Because F-FAK proteins at 2 μM yield low readings, we used GST fusion proteins of F-FAK. For all samples (including GST fusions), buffer blanks were subtracted.

Kinase Activity Assays. Autophosphorylation was monitored by Western blotting using an anti-pY397 antibody and colorimetric staining with 4CN (Biorad). Alternatively, autophosphorylation levels were measured by ELISA using a PY20 antibody conjugated with HRP and development with TMB (Calbiochem). An enzyme-coupled assay using poly(E_4Y) as a substrate was used to determine ATP turnover, as described by Lietha et al. (21). Plotted in Fig. 2*D* are negative slopes of NADH depletion. Lipid concentrations are at

100 μM C8-lipids or 1.5 mM lipid vesicles [only PC or 6% (mol/mol) $PI(4,5)P_2$, equaling 45 μM $PI(4,5)P_2$ per leaflet]. Src phosphorylation of Y576 was monitored by Western blotting. To prevent confounding results due to low levels of Y576 autophosphorylation, autophosphorylation and Src phosphorylation reactions were performed in two separate steps. Initially, FAK was allowed to autophosphorylate for 2 min. Autophosphorylation was stopped with the specific FAK inhibitor TAE226, and Src phosphorylation was started by adding Src. Blots were developed using an anti-pY576 antibody and colorimetric staining with 4CN.

EM. Proteins were prepared in the presence of 200 μM C8-PC, C8-$PI(4,5)P_2$, or polymerized PC liposomes containing 0% or 5% (mol/mol) $PI(4,5)P_2$ PolyPIPosomes (Echelon Biosciences). Samples were negatively stained on carbon-coated grids with uranyl acetate and imaged using a Tecnai G2 Spirit electron microscope (FEI) operated at 120 kV and a Slow Scan CCD (Gatan) or TemCam-F416 4,096 × 4,096-pixel camera (TVIPS GmbH).

FRET Measurements. FRET measurements were performed using a CFP excitation wavelength of 410 nm and recording emissions at 475 nm (CFP) and 525 nm (citrine), and relative FRET levels were obtained by the 525-nm/475-nm emission ratio.

ACKNOWLEDGMENTS. We thank staff at European Synchrotron Radiation Facilities beamline ID14-4 and Swiss Light Source beamline X06SA and Liang Guo at Advanced Photon Source beamline 18ID for assistance. We also thank Diego Megias, head of the Confocal Microscopy Unit at the Centro Nacional de Investigaciones Oncológicas, for his help with FA quantifications. The research was supported by Spanish Ministerio de Ciencia e Innovacion (now Ministerio de Economía y Competitividad) Grants BFU2010-15923 (to D.L.) and BFU2009-11885 (to M.P.-M.), Comunidad Autónoma de Madrid Grant S2010/BMD-2457 (to D.L.), Bundesministerium für Bildung und Forschung Grant, New Methods in Systems Biology (to F.G.), and a grant from the Klaus Tschira Foundation (to F.G.). D.L. is also the recipient of awards from the Ramón y Cajal program (Grant RYC-2010-06948) and the Volkswagen Foundation.

1. Schiller HB, Fässler R (2013) Mechanosensitivity and compositional dynamics of cell-matrix adhesions. *EMBO Rep* 14(6):509–519.
2. Schaller MD (2010) Cellular functions of FAK kinases: Insight into molecular mechanisms and novel functions. *J Cell Sci* 123(Pt 7):1007–1013.
3. Zhao X, Guan JL (2011) Focal adhesion kinase and its signaling pathways in cell migration and angiogenesis. *Adv Drug Deliv Rev* 63(8):610–615.
4. Zamir E, Geiger B (2001) Molecular complexity and dynamics of cell-matrix adhesions. *J Cell Sci* 114(Pt 20):3583–3590.
5. Ashton GH, et al. (2010) Focal adhesion kinase is required for intestinal regeneration and tumorigenesis downstream of Wnt/c-Myc signaling. *Dev Cell* 19(2):259–269.
6. Lim ST, et al. (2010) Knock-in mutation reveals an essential role for focal adhesion kinase activity in blood vessel morphogenesis and cell motility-polarity but not cell proliferation. *J Biol Chem* 285(28):21526–21536.
7. Peng X, et al. (2008) Cardiac developmental defects and eccentric right ventricular hypertrophy in cardiomyocyte focal adhesion kinase (FAK) conditional knockout mice. *Proc Natl Acad Sci USA* 105(18):6638–6643.
8. Ilić D, et al. (1995) Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice. *Nature* 377(6549):539–544.
9. Chen BH, Tzen JT, Bresnick AR, Chen HC (2002) Roles of Rho-associated kinase and myosin light chain kinase in morphological and migratory defects of focal adhesion kinase-null cells. *J Biol Chem* 277(37):33857–33863.
10. Ren XD, et al. (2000) Focal adhesion kinase suppresses Rho activity to promote focal adhesion turnover. *J Cell Sci* 113(Pt 20):3673–3678.
11. Tomar A, Schlaepfer DD (2009) Focal adhesion kinase: Switching between GAPs and GEFs in the regulation of cell motility. *Curr Opin Cell Biol* 21(5):676–683.
12. Tilghman RW, Parsons JT (2008) Focal adhesion kinase as a regulator of cell tension in the progression of cancer. *Semin Cancer Biol* 18(1):45–52.
13. Wang HB, Dembo M, Hanks SK, Wang Y (2001) Focal adhesion kinase is involved in mechanosensing during fibroblast migration. *Proc Natl Acad Sci USA* 98(20):11295–11300.
14. Zhao J, Guan JL (2009) Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev* 28(1-2):35–49.
15. Parsons JT, Slack-Davis J, Tilghman R, Roberts WG (2008) Focal adhesion kinase: Targeting adhesion signaling pathways for therapeutic intervention. *Clin Cancer Res* 14(3):627–632.
16. Schultze A, Fiedler W (2010) Therapeutic potential and limitations of new FAK inhibitors in the treatment of cancer. *Expert Opin Investig Drugs* 19(6):777–788.
17. Arold ST, Hoellerer MK, Noble ME (2002) The structural basis of localization and signaling by the focal adhesion targeting domain. *Structure* 10(3):319–327.
18. Gao G, et al. (2004) NMR solution structure of the focal adhesion targeting domain of focal adhesion kinase in complex with a paxillin LD peptide: Evidence for a two-site binding model. *J Biol Chem* 279(9):8441–8451.
19. Hayashi I, Vuori K, Liddington RC (2002) The focal adhesion targeting (FAT) region of focal adhesion kinase is a four-helix bundle that binds paxillin. *Nat Struct Biol* 9(2):101–106.
20. Cooper LA, Shen TL, Guan JL (2003) Regulation of focal adhesion kinase by its amino-terminal domain through an autoinhibitory interaction. *Mol Cell Biol* 23(22):8030–8041.
21. Lietha D, et al. (2007) Structural basis for the autoinhibition of focal adhesion kinase. *Cell* 129(6):1177–1187.
22. Schaller MD, et al. (1994) Autophosphorylation of the focal adhesion kinase, pp125FAK, directs SH2-dependent binding of pp60src. *Mol Cell Biol* 14(3):1680–1688.
23. Toutant M, et al. (2002) Alternative splicing controls the mechanisms of FAK autophosphorylation. *Mol Cell Biol* 22(22):7731–7743.
24. Calalb MB, Polte TR, Hanks SK (1995) Tyrosine phosphorylation of focal adhesion kinase at sites in the catalytic domain regulates kinase activity: A role for Src family kinases. *Mol Cell Biol* 15(2):954–963.
25. Schaller MD, Parsons JT (1995) pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. *Mol Cell Biol* 15(5):2635–2645.
26. Tachibana K, et al. (1997) Tyrosine phosphorylation of Crk-associated substrates by focal adhesion kinase. A putative mechanism for the integrin-mediated tyrosine phosphorylation of Crk-associated substrates. *J Biol Chem* 272(46):29083–29090.
27. Frame MC, Patel H, Serrels B, Lietha D, Eck MJ (2010) The FERM domain: Organizing the structure and function of FAK. *Nat Rev Mol Cell Biol* 11(11):802–814.
28. Chen SY, Chen HC (2006) Direct interaction of focal adhesion kinase (FAK) with Met is required for FAK to promote hepatocyte growth factor-induced cell invasion. *Mol Cell Biol* 26(13):5155–5167.
29. Plaza-Menacho I, et al. (2011) Focal adhesion kinase (FAK) binds RET kinase via its FERM domain, priming a direct and reciprocal RET-FAK transactivation mechanism. *J Biol Chem* 286(19):17292–17302.
30. Sieg DJ, et al. (2000) FAK integrates growth-factor and integrin signals to promote cell migration. *Nat Cell Biol* 2(5):249–256.
31. Cai X, et al. (2008) Spatial and temporal regulation of focal adhesion kinase activity in living cells. *Mol Cell Biol* 28(1):201–214.
32. Goksoy E, et al. (2008) Structural basis for the autoinhibition of talin in regulating integrin activation. *Mol Cell* 31(1):124–133.
33. Palmer SM, Playford MP, Craig SW, Schaller MD, Campbell SL (2009) Lipid binding to the tail domain of vinculin: Specificity and the role of the N and C termini. *J Biol Chem* 284(11):7223–7231.
34. Legate KR, et al. (2011) Integrin adhesion and force coupling are independently regulated by localized PtdIns(4,5)2 synthesis. *EMBO J* 30(22):4539–4553.
35. McNamee HP, Liley HG, Ingber DE (1996) Integrin-dependent control of inositol lipid synthesis in vascular endothelial cells and smooth muscle cells. *Exp Cell Res* 224(1):116–122.
36. Di Paolo G, et al. (2002) Recruitment and regulation of phosphatidylinositol phosphate kinase type 1 gamma by the FERM domain of talin. *Nature* 420(6911):85–89.
37. Ling K, Doughman RL, Firestone AJ, Bunce MW, Anderson RA (2002) Type I gamma phosphatidylinositol phosphate kinase targets and regulates focal adhesions. *Nature* 420(6911):89–93.
38. Balabin IA, Yang W, Beratan DN (2009) Coarse-grained modeling of allosteric regulation in protein receptors. *Proc Natl Acad Sci USA* 106(34):14253–14258.
39. Wang Y, et al. (2013) Platelets lacking PIP5KIγ have normal integrin activation but impaired cytoskeletal-membrane integrity and adhesion. *Blood* 121(14):2743–2752.
40. Kanchanawong P, et al. (2010) Nanoscale architecture of integrin-based cell adhesions. *Nature* 468(7323):580–584.
41. Schlessinger J (2000) Cell signaling by receptor tyrosine kinases. *Cell* 103(2):211–225.
42. Brami-Cherrier K, et al. (2014) FAK dimerization controls its kinase-dependent functions at focal adhesions. *EMBO J* 33(4):356–370.
43. Fabry B, Klemm AH, Kienle S, Schäffer TE, Goldmann WH (2011) Focal adhesion kinase stabilizes the cytoskeleton. *Biophys J* 101(9):2131–2138.
44. Ritt M, Guan JL, Sivaramakrishnan S (2013) Visualizing and manipulating focal adhesion kinase regulation in live cells. *J Biol Chem* 288(13):8875–8886.
45. Grant BD, Adams JA (1996) Pre-steady-state kinetic analysis of cAMP-dependent protein kinase using rapid quench flow techniques. *Biochemistry* 35(6):2022–2029.
46. Taylor SS, Keshwani MM, Steichen JM, Kornev AP (2012) Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos Trans R Soc Lond B Biol Sci* 367(1602):2517–2528.
47. Chen TH, Chan PC, Chen CL, Chen HC (2011) Phosphorylation of focal adhesion kinase on tyrosine 194 by Met leads to its activation through relief of autoinhibition. *Oncogene* 30(2):153–166.
48. Sun Y, Ling K, Wagoner MP, Anderson RA (2007) Type I gamma phosphatidylinositol phosphate kinase is required for EGF-stimulated directional cell migration. *J Cell Biol* 178(2):297–308.

Supporting Information

Goñi et al. 10.1073/pnas.1317022111

SI Methods

Protein Expression and Purification. All expressed focal adhesion kinase (FAK) proteins are avian FAK. The WT and mutant FAK31-405 were expressed in *Escherichia coli* BL21 (DE3) cells as in the study by Ceccarelli et al. (1). WT and mutant FAK31-686 (without or with CFP and citrine labels) were expressed in Sf21 insect cells (Invitrogen) using the baculovirus expression system as in the study by Lietha et al. (2). Full-length FAK was expressed by transient transfection of HEK293GnT1 cells using polyethyleneimine as a transfection agent (3). The hSrc84-536 was coexpressed with YopH phosphatase in *E. coli* BL21 (DE3) cells as described by Seeliger et al. (4). All proteins were expressed as GST or 6× His tag fusions. Initial purification was performed by affinity purification, followed by protease cleavage to remove tags. Proteins were further purified by anion exchange (GE Healthcare) and size exclusion (GE Healthcare) chromatography.

Kinase Activity Assays. FAK autophosphorylation was monitored by Western blotting. Autophosphorylation reaction mixtures containing 0.4 μM FAK were split into three equal volumes after taking the $t = 0$ sample, and buffer (Hepes-buffered saline, pH 7.5), phosphatidyl choline (PC), or 6% (mol/mol) phosphatidylinositol-4,5-bisphosphate [PI(4,5)P2] vesicles (20 μM on the outer leaflet) were added and incubated for 20 min. Reactions were started with 1 mM ATP and stopped with SDS/PAGE loading buffer. Western blots were performed with an anti-pY397 antibody (Invitrogen), and colorimetric staining was performed with 4CN (Biorad).

Src phosphorylation of Y576 was monitored by Western blotting. To prevent confounding results due to low levels of Y576 autophosphorylation, autophosphorylation and Src phosphorylation reactions were performed in two separate steps. Initially 0.4, μM FAK was allowed to autophosphorylate for 2 min in the absence or presence of PC or $PI(4,5)P_2$ vesicles (in the same conditions as above). Autophosphorylation was stopped by adding 0.45 μM TAE226 (Chemexpress), a specific FAK inhibitor, and Src phosphorylation was started by adding 100 nM Src84-536. Reactions were stopped with SDS/PAGE loading buffer and blotted using an anti-pY576 antibody (Invitrogen) and colorimetric staining with 4CN (Biorad).

An antiphosphotyrosine-based ELISA method was used to quantify kinase autophosphorylation activity. Kinase reactions were performed with 200 nM WT 4.1, ezrin, radixin, moesin homology (FERM) + kinase fragment of WT FAK (FK-FAKwt) or FK-FAK180/183A in presence of 0–100 μM C8-PC, C8-$PI(4,5)P_2$, or $Ins(1,4,5)P_3$ (Echelon Biosciences). Reactions were started with 0.5 mM ATP and stopped after 1 min at room temperature (RT) with 300 mM EDTA. Phosphorylation levels were quantified by ELISAs using a PY20 antibody conjugated with HRP and colorimetric development with TMB (Calbiochem).

An enzyme-coupled spectrophotometric assay was used to determine ATP turnover of FAK proteins as described by Lietha et al. (2). In brief, reactions were performed with 1 μM kinase, 10 mM $MgCl_2$, 1 mM phosphoenolpyruvate, 0.25 mM NADH, 0.08 units/μL pyruvate kinase, 0.1 units/μL lactate dehydrogenase, and 100 μM E4Y (as polyGlu-Tyr, 4:1 Glu/Tyr; Sigma) in the absence or presence of soluble lipids or lipid vesicles. In Fig. 2, we used 100 μM soluble lipids [C8-PC or C8-$PI(4,5)P_2$] or 1.5 mM lipid vesicles [PC or 6% (mol/mol) $PI(4,5)P_2$, equaling 45 μM $PI(4,5)P_2$ per leaflet]. In Fig. S3*D*, lipid concentrations are indicated. Reactions were initiated with 0.5 mM ATP, and NADH depletion was monitored by A at 340 nm.

An ADP Hunter assay (DiscoveRx Corp) was used to determine FAK activity with the FAK FERM+linker fragment (F-FAK = FAK31-405 containing Y397) as a substrate. Kinase reactions were performed with 50 nM FK-FAK in 20 mM Hepes (pH 7.4), 150 mM NaCl, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), 0.2 mM $NaVO_3$, 200 μM ATP, 2 mM $MgCl_2$, and 40 μM FERM-linker substrate for 20 min at RT. The measurements were made in the absence or presence of 100 μM C8-PC or C8-$PI(4,5)P_2$ (Echelon Biosciences). Reactions were quenched by the addition of the ADP Hunter reagents and allowed to develop for 30 min at RT. Fluorescence was measured at 530-nm excitation and 590-nm emission wavelengths using an EnVision Multilabel Plate Reader (PerkinElmer).

EM. For negative staining, proteins were diluted to 0.1 mg/mL in 20 mM Hepes (pH 7.5) and 150 mM NaCl in the absence or presence of 200 μM C8-PC or C8-$PI(4,5)P_2$ or in the presence of polymerized PC liposomes containing 0% or 5% $PI(4,5)P_2$ (PolyPIPosomes; Echelon Biosciences). Samples were applied to glow-discharged, carbon-coated grids; negatively stained with 2% uranyl acetate; and imaged using a Tecnai G2 Spirit electron microscope (FEI) operated at 120 kV. Images were recorded using a Slow Scan CCD Camera Model 694 (Gatan) at a calibrated magnification of 21,000. For full-length FAK clusters with C8-$PI(4,5)P_2$, images were recorded using a TemCam-F416 4,096 × 4,096-pixel camera (TVIPS GmbH) at a calibrated magnification of 30,000 with 5.2 Å per pixel at the specimen level. A total of 574 individual particles were extracted using the *boxer* program implemented in EMAN (5). Raw data were analyzed using reference-free alignment and classification methods. The 3D model was build using the *makeinitialmodel.py* and *startAny* programs implemented in EMAN.

FRET Measurements. FRET measurements were performed with a 250 nM FRET sensor using a CFP excitation wavelength of 410 nm and recording emissions at 475 nm (CFP) and 525 nm (citrine) on a QuantaMaster QM2000-7 instrument (Photon Technology International). Relative FRET levels were obtained by calculating the emission ratio Em(525nm)/Em(475nm). In 2-min intervals, the following sequential additions were made: (*i*) 100 μM lipid [C8-PC, C8-$PI(4,5)P_2$, or no lipid], (*ii*) 100 nM Src84-536 (WT or K298M) or no Src, and (*iii*) 1 mM ATP and 2 mM $MgCl_2$. In Fig. 4*B* and Fig. S5 *B–D*, mean FRET values are plotted along with the SD.

Crystallization and Structure Determination of the FERM Domain of the FAK KAKTLR and Low-Resolution FERM Domain of WT FAK. The chFAK31-405wt [FERM domain of WT FAK (F-FAKwt)] or chFAK31-405 K216A, K218A, R221A, and K222A basic patch mutant FERM domain (F-FAK-KAKTLRK) was concentrated to ∼5 mg/mL in 20 mM Tris (pH 8.0), 150 mM NaCl, 5% glycerol, and 5 mM TCEP. Crystals were obtained with 18% PEG 4000, 325 mM $MgCl_2$, 100 mM Tris (pH 8.5), and 10 mM TCEP (F-FAKwt) or 14% PEG 4000, 200 mM $MgCl_2$, 100 mM Tris (pH 8.5), and 10 mM TCEP (F-FAK-KAKTLRK) as precipitant solutions and were frozen in the precipitant solution containing 15% ethylene glycol (F-FAK-KAKTLRK) or 15% ethylene glycol, 5% DMSO, and 4 mM *p*-(trifluoromethyl)phenol (F-FAKwt).

Data were collected at the Swiss Light Source (beamline X06SA) to 3.2 Å for F-FAKwt or at the European Synchrotron Radiation Facilities (beamline ID14-4) to a resolution of 3.15 Å for F-FAK-KAKTLRK. Data were processed with XDS (6), and

the Protein Data Bank (PDB) ID code 2AEH coordinates were used to obtain initial phases (1). Both structures contain two molecules in the asymmetrical unit [F-FAKwt: 51.9% solvent, Matthews Coefficient (V_m) = 2.56 Å^3/Da; F-FAK-KAKTLRK: 51.6% solvent, V_m = 2.54 Å^3/Da (7)]. Refinement was performed with Refmac (8), and manual rebuilding was performed with Coot (9). Final *R* factors calculated from reflections used in the refinement (R_{work}) or from 5% of reflections not used in the refinement (R_{free}) are 17.3/24.0 (R_{work}/R_{free}) for F-FAKwt and 22.3/27.8 (R_{work}/R_{free}) for F-FAK-KAKTLRK. The full crystallographic table is shown (Table S3). The structures have been deposited in the PDB under ID codes 4CYE (F-FAKwt) and 3ZDT (F-FAK-KAKTLRK).

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations of the catalytic domain of FAK were performed with the ACEMD program (10) and the AMBER99SB*-ILDN force field (11, 12). Simulations are based on the crystal structure (PDB ID code 2J0L) with unphosphorylated Y576 and Y577. For ATP-bound states, the crystallographic 5′-adenylylimidodiphosphate (AMP-PNP) was substituted with ATP, allowing ATP and Mg^{2+} to move freely. All of the resulting structures were energy-minimized and solvated in a cubic box of TIP3P water molecules (13). The edge of the box was chosen to have a minimum distance of 9 Å from the protein, resulting in a length of 87 Å after the equilibration at constant pressure in the isothermal-isobaric ensemble (NPT). The system charges were neutralized with positive Na^+ ions. The systems were gradually heated to 300 K and then equilibrated at constant pressure (NPT) for 5 ns and at constant volume (constant number of atoms, volume, and temperature) for 200 ns. The production runs were performed in the canonical ensemble, keeping the temperature at 300 K with a Langevin thermostat and a time step of 2 fs. The cutoff for the van der Waals interactions and short-range electrostatic interactions was set at 1.0 nm. Long-range electrostatic interactions were calculated by the particle mesh Ewald algorithm, with a mesh space of 0.1 nm. The simulations of the apo- and ATP-bound states were run for a total of 1,500 ns. The rms fluctuations were calculated alternatively on the whole trajectory discarding the first and last 200 ns or in 200-ns long windows with similar results.

All simulations with FK-FAK were based on the crystal structure with PDB ID code 2J0J (2). Initial geometries of the PI(4,5)P_2–FK-FAK complex were obtained using a molecular docking procedure. The putative PI(4,5)P_2 binding site has been assigned to the FERM basic patch (residues 216–222), based on the results reported in previous studies (14). Partial atomic charges on the PI(4,5)P_2 molecule were assigned using restrained electrostatic potential (RESP) methodology (15), utilizing the HF/6-31G* level of theory. We used PI(4,5)P_2 in its fully deprotonated state (net charge of −5), because this state has been reported to be dominant at an experimental pH of 7.5–8 (16), and because the basic patch of FAK is likely to replace protons of PI(4,5)P_2 upon binding (17, 18). Molecular docking was performed using the University of California, San Francisco DOCK 6.5 suite (19) with grid scoring in an implicit solvent. The grid spacing was 0.25 Å, and the grid included 12 Å beyond the FAK basic patch. The energy score was the sum of electrostatic and van der Waals contributions. In the course of the docking procedure, the PI(4,5)P_2 molecule was subjected to 2,500 cycles of molecular-mechanical energy minimization. The number of maximum ligand orientations was 5,000. The 10 best-scoring PI(4,5)P_2–FK-FAK complexes were further analyzed by means of MD simulations. The best structure was chosen according to the Coulomb interaction between the PI(4,5)P_2 head group and the basic patch in the FERM domain. We embedded the FAK-bound PI(4,5)P_2 into a bilayer of palmitoyl-oleoyl-glycero-3-phosphoethanolamine for a subsequent 150 ns of MD simulations and observed a highly stable FAK–PI(4,5)P_2 interaction that was structurally highly similar to the same complex in water only, validating our model.

All simulations containing FK-FAK were carried out using the MD software package GROMACS 4.0.5 (20). The optimized potentials for liquid simulations all-atom force field (21) for the protein and the TIP4P water model (13) were used. The temperature was kept constant at $T = 300$ K using velocity rescaling with a coupling time of 0.1 ps. In all simulations, the long-range electrostatic interactions were treated with the particle mesh Ewald (22) method, using a grid spacing of 0.12 nm with cubic interpolation. All bonds between hydrogens and heavy atoms were constrained using the LINCS algorithm (23). An integration time step of 2 fs was used. During all simulations, the neighbor list was updated every 10 steps.

Each of the three structures, the apo FK-FAK, the complex with PI(4,5)P_2, and the basic patch mutant, was solvated in a dodecahedral box containing about 200,000 atoms. Sodium and chloride counter-ions were added to neutralize the system with a concentration of 0.1 M. Before MD simulations, all systems were minimized using the steepest descent method for 10,000 steps, followed by 2-ns MD simulations during which position restraints were used on all bonds. Finally, the whole system was equilibrated during MD simulations of 120 ns. Two independent equilibrium MD simulations were performed for the apo FK-FAK.

Subsequently, we analyzed the equilibrium trajectories by force distribution analysis (24) as implemented in the GROMACS package to investigate the change in internal forces of the molecules under external perturbations. Atomic pairwise forces were calculated during the analysis, and residue forces were obtained by vector summation of atomic forces. We calculated the change in residue pairwise forces of PI(4,5)P_2–FK-FAK and FK-FAK-KAKTLRK, respectively, with respect to FK-FAK. Fig. S7*A* shows the largest continuous cluster of force changes upon PI(4,5)P_2 binding (over 140 pN) or KAKTLRK mutation (over 200 pN) (25).

Allosteric Coupling. The allosteric coupling maps have been obtained with an in-house reimplementation of the perturbation theory combined with normal mode analysis as proposed by Balabin et al. (26). The normal mode analysis was performed with the ProDy modeling package (27) on the Cα atoms of the PDB ID code 2J0L crystal structure.

Small-Angle X-Ray Scattering. FK-FAKwt and FK-FAK180/183A were prepared at three different concentrations (in the range of 1–4 mg/mL) in 20 mM Tris (pH 8.0), 200 mM NaCl, 5% glycerol, and 2 mM TCEP. Small-angle X-ray scattering data were collected at the Advanced Photon Source (Argonne National Laboratory) beamline 18ID (Biophysics Collaborative Access Team), acquiring 15 exposures of 1 s each. The scattering intensity I(Q) plots for each sample, were obtained by averaging measurements of all exposures and subtracting matched buffer plots. Distance distribution plots were obtained with the program GNOM (28) and were used for shape reconstructions with GASBOR (29). For both FK-FAKwt and FK-FAK180/180A, results from 20 calculations with GASBOR were averaged with DAMAVER (30).

The ensemble optimization method (31) was used to assess flexibility of domains relative to each other. The crystal structures of the FAK FERM (PDB ID code 2AEH) and kinase (PDB ID code 2J0L) domains were used as input to generate a pool of 10,000 structures with random linker conformations, and the ensemble of conformations that best fits the experimental scattering data was selected using a genetic algorithm. Theoretical scattering curves based on structural models and fits to experimental data (χ values) were calculated with CRYSOL (32), applying a constant subtraction.

Surface Plasmon Resonance. Vesicles for surface plasmon resonance (SPR) were prepared similar as described for vesicle pull-downs; however, instead of sonication, vesicles were subjected to five cycles of freezing (in liquid nitrogen), thawing (40 °C), and vortexing after every cycle. To obtain homogeneously sized vesicles, the lipid suspension was extruded for 15 passages through a 100-nm pore membrane using a miniextruder (Avanti Polar Lipids).

SPR was measured using a Biacore 3000 instrument (GE Healthcare). The L1 chip was coated with lipid vesicles at a total lipid concentration of 1.5 mM, containing 3% (mol/mol) $PI(4,5)P_2$ and 97% (mol/mol) PC, for 10 min at a flow rate of 1 μL/min. FK-FAKwt or FK-FAK180/183A was prepared in running buffer [20 mM Hepes (pH 7.5), 150 mM NaCl, 1 mM TCEP] at 1.6 μM, 3.1 μM, and 6.3 μM, and the samples were injected at a flow rate of 30 μL/min. Proteins were allowed to associate for 1 min and then to dissociate for at least 2 min. After every protein injection, the lipid layer was removed from the sensor chip by short pulses (30 s) of 1% octylglycoside and short pulses (20 s) of 2:3 isopropanol and 50 mM NaOH.

Dynamic Light Scattering. Dynamic light scattering was performed at 25 °C using a DynaPro instrument (Protein Solutions). The proteins (F-FAKwt, F-FAK-KAKTLRK, or FK-FAKwt) were prepared at 2 mg/mL in filtered 20 mM Hepes (pH 7.5), 150 mM NaCl, and 1 mM TCEP buffer in the presence of 100 μM C8-PC or C8-$PI(4,5)P_2$ (F-FAK and F-FAK-KAKTLRK) or 200 μM C8-$PI(4,5)P_2$ (FK-FAKwt), so as to match the conditions in EM studies. The samples were centrifuged at 16,100 × *g* for 2 h before dynamic light scattering measurements. The laser power was adjusted to 100%, and 30 measurements were acquired using a 10-s acquisition time. Before regularization, the data filter was set up as follows: amplitude, 0–1; baseline limit, 1 ± 1; and maximum sum of squares error statistic <100 (33). The results were processed with Dynamics V6 software (DynaPro), and the molecular weight values were calculated with the regularization algorithm. In Table S2, mean values of three independent experiments, along with the SE, are given for the hydrodynamic radius and derived molecular weight (Mw-R).

Cell Culture and Phosphatidylinositol 4-Phosphate 5-Kinase Type Iγ Knockdown. HeLa cells were grown in DMEM containing 10% bovine calf serum. A lentivirus-based shRNA system (Sigma–Aldrich) was used to reduce the phosphatidylinositol 4-phosphate 5-kinase type Iγ (PIP5KIγ) isoform 2 (CCGGCGATGAGAGGAGCTGGGTGTACTCGAGTACACCCAGCTCCTCTCATCGTTTTTG), along with a scramble shRNA control. HeLa cells were infected with virus-containing supernatants and polybrene (2 μg/mL) for 24 h at 37 °C. After the generation of a stable cell line with puromycin (1 μg/mL), two clones were selected and validated by Western blot analysis.

Cell Adhesion Assay. Coverslips were coated with fibronectin (10 μg/mL), fixed overnight in PBS, and washed three times with PBS. Cells (50×10^3) were plated in triplicate and fixed after 30 min of plating in 4% paraformaldehyde (PFA; Electron Microscopy Sciences). Cells were stained with 594-phalloidin and DAPI. Image acquisition was done using an inverted motorized microscope stage (Leica DMI 6000), and cell counting was performed using ImageJ software (National Institutes of Health). Assays were performed in triplicate ($n = 4$).

Immunofluorescence. Cells were seeded on plates with coverslips. After overnight serum starvation, cells were stimulated for 15 min and 2 h with 10% bovine calf serum and fibronectin (5 μg). Cells were fixed with 4% PFA in PBS for 10 min. Samples were blocked for 30 min with 1% BSA, 0.3% Triton X-100, 5% normal goat serum, 5% normal donkey serum, and 1% gelatin in PBS. The primary antibodies were used at the following ratios: paxillin, 1:100 (clone 5H11; Millipore); FAK, 1:100 (Millipore); and pFAKY397, 1:100 (Invitrogen). Secondary antibodies used were conjugated to FITC or Texas Red (Jackson Labs Technologies). The nucleus was stained with DAPI. Image acquisition was done using a confocal Leica Ultra-Espectral TCS-SP5 microscope. Quantifications were performed with Definiens Developer XD V2.0 software.

Western Blotting. Overnight serum-starved cells were trypsinized and plated on 10-μg fibronectin-coated plates for 2 h. Cells were lysed in radioimmunoprecipitation assay buffer. Cell lysates were cleared by centrifugation at 16,100 × *g*. Primary antibodies used were as follows: PIP5KIγ isoform 2 (1:500, recognizes residues 650–668; catalog no. K1894, Sigma–Aldrich), actin (1:5,000; Sigma–Aldrich), FAK (1:1,000; Millipore), pFAK(Y397) (1:1,000; Invitrogen), pFAK(Y576) (1:1,000; Invitrogen), and pFAK(Y577) (1:1,000; Invitrogen). Blots were developed with HRP-labeled secondary antibody (DAKO) and ECL detection (GE Healthcare).

1. Ceccarelli DF, Song HK, Poy F, Schaller MD, Eck MJ (2006) Crystal structure of the FERM domain of focal adhesion kinase. *J Biol Chem* 281(1):252–259.
2. Lietha D, et al. (2007) Structural basis for the autoinhibition of focal adhesion kinase. *Cell* 129(6):1177–1187.
3. Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res* 30(2):E9.
4. Seeliger MA, et al. (2005) High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci* 14(12):3135–3139.
5. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128(1):82–97.
6. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
7. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33(2):491–497.
8. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255.
9. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
10. Harvey MJ, Giupponi G, De Fabritiis G (2009) ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* 5:1632–1639.
11. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78(8):1950–1958.
12. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113(26):9004–9015.
13. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.
14. Cai X, et al. (2008) Spatial and temporal regulation of focal adhesion kinase activity in living cells. *Mol Cell Biol* 28(1):201–214.
15. Dupradeau FY, et al. (2010) The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* 12(28):7821–7839.
16. Graber ZT, Jiang Z, Gericke A, Kooijman EE (2012) Phosphatidylinositol-4,5-bisphosphate ionization and domain formation in the presence of lipids with hydrogen bond donor capabilities. *Chem Phys Lipids* 165(6):696–704.
17. Ellenbroek WG, et al. (2011) Divalent cation-dependent formation of electrostatic PIP2 clusters in lipid monolayers. *Biophys J* 101(9):2178–2184.
18. McLaughlin S, Wang J, Gambhir A, Murray D (2002) PIP(2) and proteins: Interactions, organization, and information flow. *Annu Rev Biophys Biomol Struct* 31:151–175.
19. Lang PT, et al. (2009) DOCK 6: Combining techniques to model RNA-small molecule complexes. *RNA* 6:1219–1230.
20. Van Der Spoel D, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26(16):1701–1718.
21. Jorgensen WL, Tirado-Rives J (1988) The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 110(6):1657–1666.
22. Darden T, York D, Pedersen L (1993) Particle mesh Ewald-an Nlog(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
23. Hess B, Bekker H, Berendsen JC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472.
24. Stacklies W, Seifert C, Graeter F (2011) Implementation of force distribution analysis for molecular dynamics simulations. *BMC Bioinformatics* 12:101.
25. Seifert C, Gräter F (2012) Force distribution reveals signal transduction in E. coli Hsp90. *Biophys J* 103(10):2195–2202.
26. Balabin IA, Yang W, Beratan DN (2009) Coarse-grained modeling of allosteric regulation in protein receptors. *Proc Natl Acad Sci USA* 106(34):14253–14258.

27. Bakan A, Meireles LM, Bahar I (2011) ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* 27(11):1575–1577.
28. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25:495–503.
29. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946–2953.
30. Volkov VV, Svergun DI (2003) Uniqueness of ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860–864.
31. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129(17):5656–5664.
32. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
33. Borgstahl GE, Gloria EO (2007) How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. *Methods in Molecular Biology, Macromolecular Crystallography Protocols*, ed Walker JM (Humana Press, Totowa, NJ), Vol 363, pp 109–130.

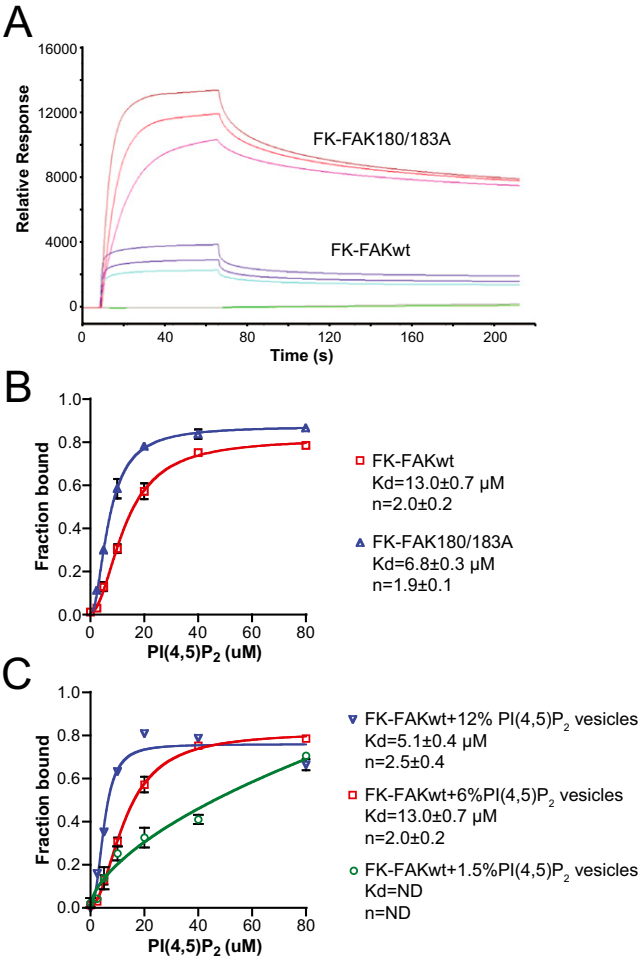


Fig. S1. FK-FAKwt and FK-FAK180/183A binding to PI(4,5)P_2 vesicles. (*A*) SPR binding measurements of FK-FAKwt and FK-FAK180/183A. Shown are SPR sensograms resulting from injections of FK-FAKwt (blue plots) and FK-FAK180/183A mutant (red plots) at 1.6 μM, 3.1 μM, or 6.3 μM (lower to higher response) into a flow cell, where vesicles containing 3% (mol/mol) PI(4,5)P_2 were immobilized onto an L1 sensor chip. Although quantification of binding parameters was not possible due to PI(4,5)P_2-induced clustering (Fig. 3), the higher response of FK-FAK180/183A suggests a higher affinity of this mutant toward PI(4,5)P_2 membranes compared with FK-FAKwt. This is in agreement with data obtained by vesicle pull-down (Fig. 1*D*). Buffer injections are plotted in green. (*B*) Binding data from vesicle pull-down experiments for FK-FAKwt and FK-FAK180/183A shown in Fig.1*D* are fitted with a cooperative model [y = B_{max} * x^n/(K_{dapp}^n + x^n), where B_{max} is y at saturation, n is the Hill coefficient, K_{dapp} is the apparent K_d, and x is the PI(4,5)P_2 concentration]. The fitting suggests positive co-operativity. (*C*) Vesicle pull-down binding data shown in Fig. 1*E* are fitted with a cooperative model as in *B*. ND, not determined.

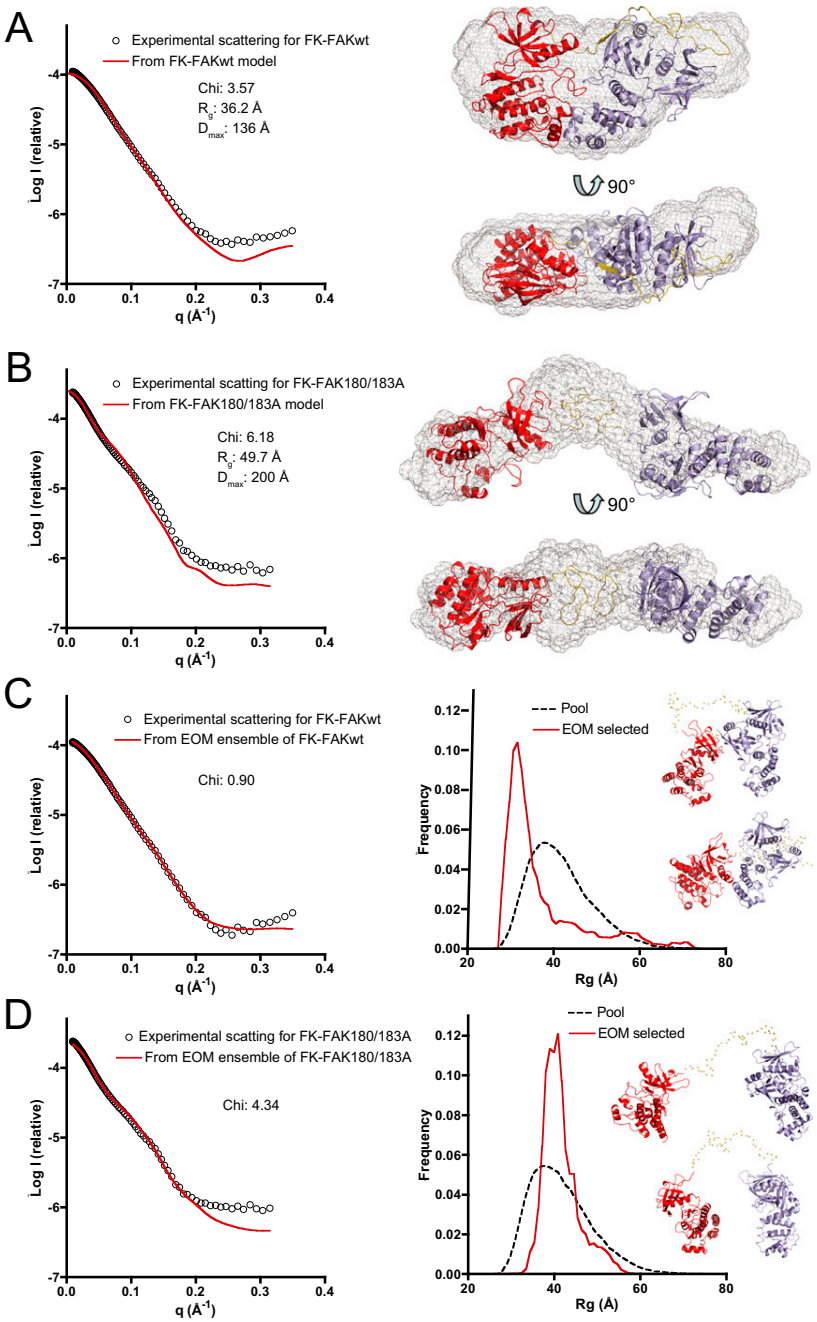


Fig. S2. Small-angle X-ray scattering analysis (SAXS) of FK-FAKwt and FK-FAK180/183A. Experimental SAXS curves for FK-FAKwt (*A*) and FK-FAK180/183A (*B*) are shown (*Left*, black circles), along with shape reconstructions calculated with GASBOR (1) (*Right*, gray mesh). (*Right*) Crystal structures of FERM (blue ribbon) and kinase (red ribbon) domains, as well as linker regions (yellow), are fitted into the envelopes. (*Left*) Theoretical scattering plots for the fitted models (red curve) and χ values were calculated with CRYSOL (2). Radius of gyration R_g and maximal dimension (D_{max}) values are derived with GNOM (3). SAXS envelopes of FK-FAKwt reveal a compact shape in agreement with the FK-FAK crystal structure (4), whereas SAXS reconstructions of FK-FAK180/183A suggest an open conformation where the FERM and kinase domains are dissociated. Flexibility of the domains was modeled for FK-FAKwt (*C*) and FK-FAK180/183A (*D*) using an ensemble optimized method (EOM) (5). (*Right*) R_g size distributions for EOM-selected ensembles (red lines) are compared with the random pool of 10,000 models (black dotted lines). EOM-generated models representing the two most populated clusters of selected conformations are shown with FERM and kinase domains in a ribbon representation and with the linker as a dotted line. (*Left*) Predicted scattering curves from the EOM-selected ensembles (red curves) are compared with experimental scattering (black circles). Discrepancies at a high q value are likely due to imperfect modeling of domain flexibility (6).

1. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946–2953.
2. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
3. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25:495–503.
4. Lietha D, et al. (2007) Structural basis for the autoinhibition of focal adhesion kinase. *Cell* 129(6):1177–1187.
5. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129(17):5656–5664.
6. Jacques DA, Guss JM, Svergun DI, Trewhella J (2012) Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallogr D Biol Crystallogr* 68(Pt 6):620–626.

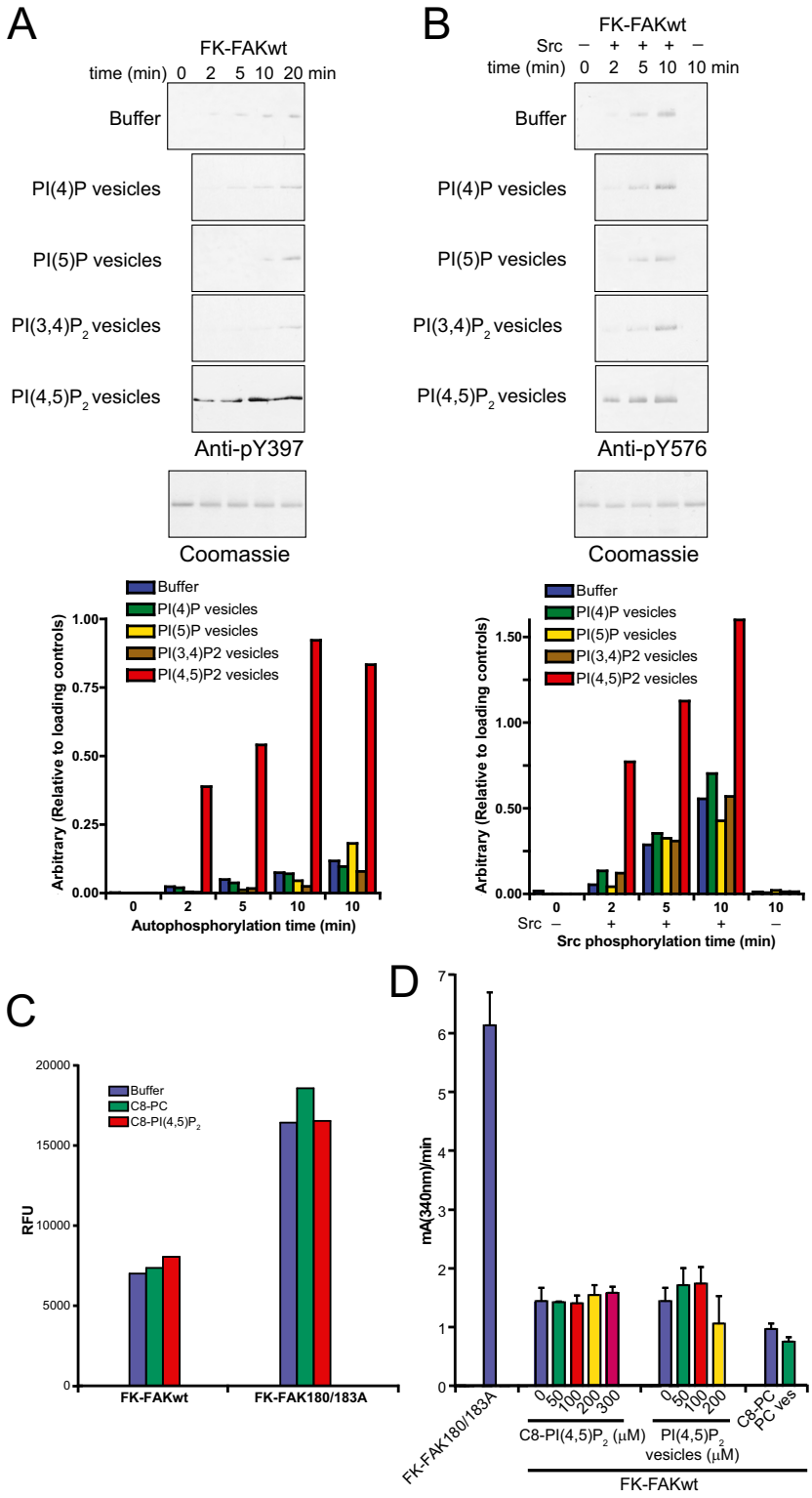


Fig. S3. Effect of phosphoinositides on autophosphorylation and Src phosphorylation and catalytic turnover. Autophosphorylations (*A*) and Src phosphorylations (*B*), as shown in Figs. 2 and 6, are performed in the presence of vesicles containing PI(4)P, PI(5)P, PI(3,4)P_2, and PI(4,5)P_2. The increase of autophosphorylation and Src phosphorylation rates is specific to PI(4,5)P_2. (*Lower*) Quantifications from the blots are shown relative to the Coomassie-stained loading controls. (*C*) Catalytic turnover was measured for FK-FAKwt and FK-FAK180/183A with FAK FERM + linker (FAK31-405, containing the Y397 autophosphorylation site) as a substrate using the ADP Hunter method for increased sensitivity. We confirm, using a more native substrate, the results shown in Fig. 2*D* (where polyE4Y was used as a substrate) that PI(4,5)P_2 does not increase catalytic turnover, whereas the FK-FAK180/183A mutations do. RFU, relative florescence units. (*D*) Catalytic turnover measured by the coupled kinase assay as in Fig. 2*D*. For FK-FAKwt, activity was measured in the presence of increasing PI(4,5)P_2 concentrations as indicated (soluble or in vesicles). Measurement of the PC controls (soluble or in vesicles) is performed at the lipid concentration corresponding to the highest PI(4,5)P_2 concentration. None of the PI(4,5)P_2 concentrations significantly affects the activity of FK-FAKwt. Error bars represent SD from three independent experiments.

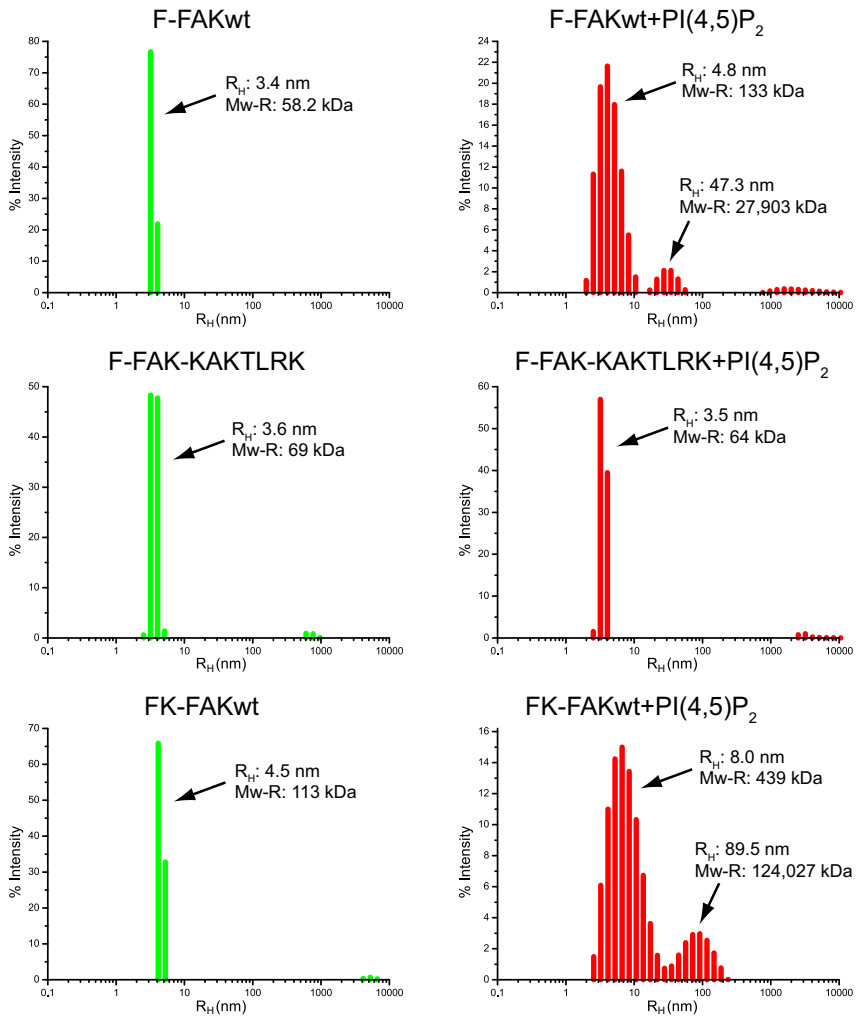


Fig. S4. PI(4,5)P_2 induces oligomers and polydispersity of F-FAK and FK-FAK. F-FAKwt, F-FAK-KAKTLRK, and FK-FAKwt proteins were analyzed by dynamic light scattering in the presence or absence of C8-PI(4,5)P_2, and the obtained data were processed by a regularization algorithm (Dynamics V6 software; Dynapro). Shown are the resulting binned histogram plots reporting the distribution of the particle size of the samples. R_H, hydrodynamic radius. The indicated molecular weight (Mw-R) values are derived from R_H, and in the case of polydisperse samples, they are the weighted average of more than one species (more details are provided in Table S2 and its legend). PI(4,5)P_2 induces oligomerization and polydispersity of F-FAKwt and FK-FAKwt but not F-FAK-KAKTLRK.

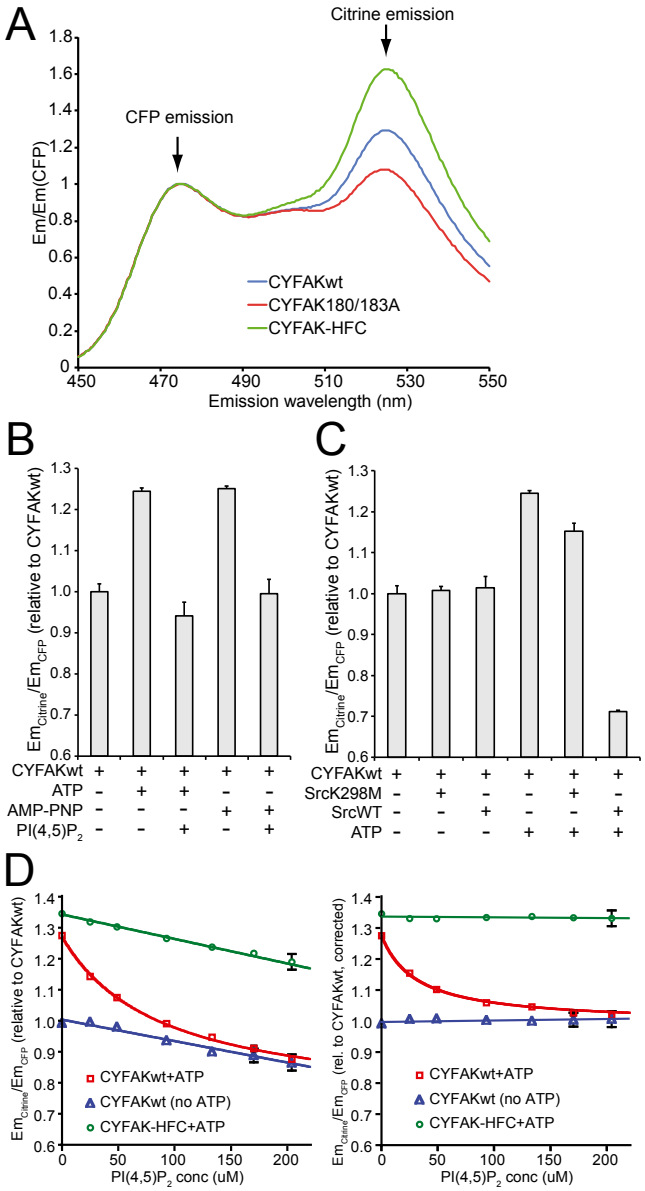


Fig. S5. FRET measurements with sensors containing CFP and citrine florescent proteins fused to FAK (CYFAK sensors). (*A*) Emission scans of FAK FRET sensors. The emission scans for the three FAK FRET sensors are shown, using an excitation wavelength of 410 nm (excitation of CFP). Curves are normalized with the peak CFP emission at 475 nm. The sensor with a WT FAK sequence (CYFAKwt) displays higher citrine emission signals (at 525 nm) than the open CYFAK180/180A mutant, and the high-FRET control sensor (CYFAK-HFC) shows the highest citrine emission. (*B*) Emission ratios of citrine (525 nm; $Em_{Citrine}$) and CFP (475 nm; Em_{CFP}) are plotted relative to the CYFAKwt ratio as a relative measure of FRET. The presence of ATP and the non-hydrolyzable ATP analog 5′-adenylylimidodiphosphate (AMP-PNP) increases the FRET level of CYFAKwt. In both cases C8-PI(4,5)P_2 reduces FRET signals, indicating that this effect is not phosphorylation-dependent. (*C*) FRET levels of CYFAKwt are not affected by Src (WT or kinase-dead K298M) in the absence of ATP. In the presence of ATP, inactive SrcK298M induces a small FRET reduction (indicating an effect due to Src binding), whereas SrcWT induces a large reduction in FRET [measured at the end of the reaction (i.e., the last 2 min) in Fig. 4*C*, *Left*], indicating that this effect is mostly due to Src phosphorylation of CYFAKwt. (*D*) The conformational effect of PI(4,5)P_2 shown in Fig. 4*B* is further investigated by titrating PI(4,5)P_2 concentrations (conc). Whereas PI(4,5)P_2 has only a linear nonconformational effect on CYFAKwt in the absence of ATP (same effect as for CYFAK-HFC), in the presence of ATP, PI(4,5)P_2 induces a dose-dependent reduction in FRET levels that approaches the level of CYFAKwt in the absence of ATP. The linear non-conformational reduction in FRET seen for CYFAK-HFC and CYFAKwt without ATP (*Left*) is added to the data points (*Right*) to correct for the non-conformational effect.

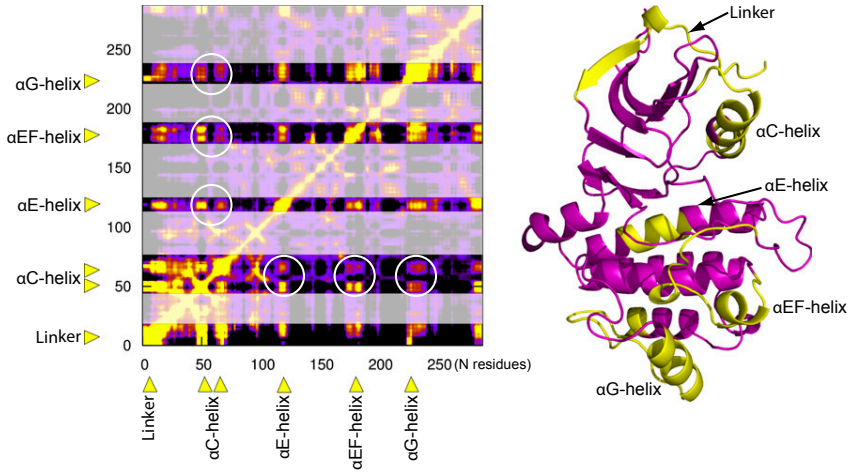


Fig. S6. Allosteric coupling analysis of the FAK. Allosteric coupling map for residues in the FAK domain (PDB ID code 2J0L) as described by Balabin et al. (1). The main allosteric sites are labeled, whereas other regions are shaded for clarity. Allosteric coupling ranges from yellow (high coupling) to black (no coupling). Whereas residues proximal in space are expected to couple, residues far in space are coupling allosterically. High coupling is observed for the αC-helix in the kinase N-lobe with the αE-, αEF-, and αG-helices in the kinase C-lobe (white circles). (*Right*) Structural elements of the FAK domain found to couple allosterically are mapped in yellow onto the kinase structure in a ribbon representation.

1. Balabin IA, Yang W, Beratan DN (2009) Coarse-grained modeling of allosteric regulation in protein receptors. *Proc Natl Acad Sci USA* 106(34):14253–14258.

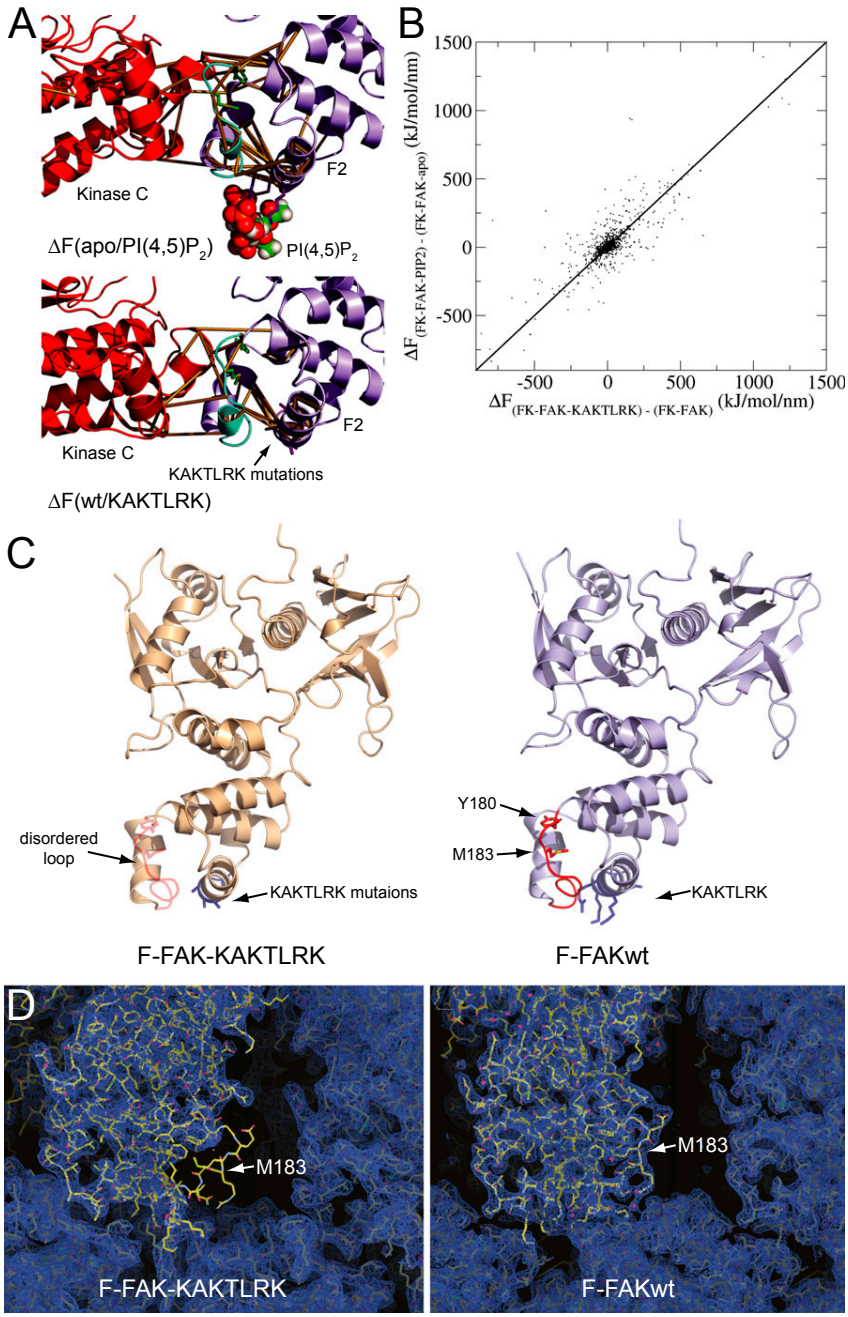


Fig. S7. Effect of neutralizing the basic KAKTLRK region. (*A*) MD analysis of FK-FAK indicates similar changes in the force distribution (ΔF) when the basic patch is neutralized by PI(4,5)P_2 binding (*Upper*) or KAKTLRK mutation (*Lower*). Resulting ΔFs propagate from the basic patch to the loop (cyan) containing Y180/M183 (green) at the FERM/kinase interface. Brown cylinders indicate differences in residue pairwise forces between PI(4,5)P_2-bound and apo FK-FAK above 140 pN (*Upper*) or between FK-FAKwt and FK-FAK-KAKTLRK (*Lower*, above 200 pN because the effect is more pronounced with the mutant) and measure minor conformational rearrangements as described by Stacklies et al. (1). (*B*) Charge neutralization at the basic patch by PI(4,5)P_2 binding or basic patch mutation (FK-FAK-KAKTLRK) results in a quantitatively reproducible change in the force distribution pattern. Differences in residue pairwise force (ΔF) between FK-FAK–PI(4,5)P_2 and apo FK-FAK are plotted against those between mutant FK-FAK-KAKTLRK and FK-FAKwt. Each data point corresponds to a force difference between two residues in the FERM F2 and/or kinase C-lobe. The plot shows a good correlation between ΔF values resulting from PI(4,5)P_2 binding and from KAKTLRK mutation. (*C*) Crystal structure of F-FAK-KAKTLRK. The crystal structure of F-FAK-KAKTLRK (*Left*, PDB ID code 3ZDT) is shown in a ribbon representation and is compared with an F-FAKwt structure solved at a similar resolution (*Right*, PDB ID code 4CYE). The basic patch residues (or alanine mutations) are shown as blue sticks. The asymmetric unit cell of F-FAK-KAKTLRK contains two molecules, which superimpose well (rmsd of Cα atoms = 0.638 Å). Shown is the molecule with chain identifier A. The overall structure is very similar to F-FAKwt (with rmsd values compared to 4CYE between 0.563 and 1.121 Å, depending on the chain); however, in F-FAK-KAKTLRK, the loop containing residues Y180 and M183 is disordered in both molecules in the asymmetric unit (shown in red transparent color). (*D*) Crystal packing environment of the FERM F2 lobe in F-FAK-KAKTLRK (chain B) is compared with the environment of the F2 lobe in F-FAKwt (PDB ID code 4CYE; chain A). The 2Fo-Fc electron density is shown in blue, countered at 1σ.

1. Stacklies W, Seifert C, Graeter F (2011) Implementation of force distribution analysis for molecular dynamics simulations. *BMC Bioinformatics* 12:101.

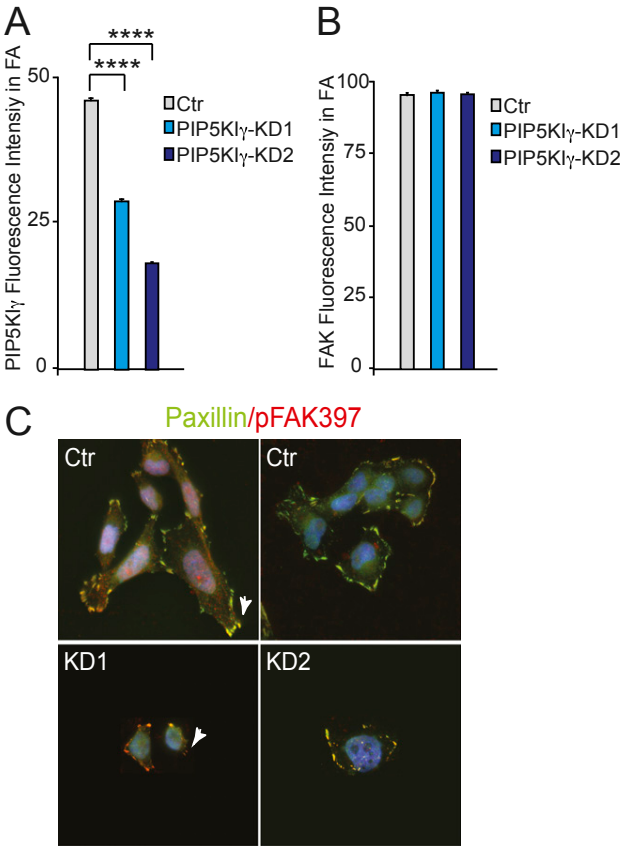


Fig. S8. PIP5KIγ knockdown affects focal adhesion size but not FAK recruitment. (*A*) Quantification of the PIP5KIγ immunofluorescence signal in focal adhesions upon knockdown of PIP5KIγ compared with scramble controls (Ctr). Data are mean values ± SEM. ****$P < 0.0001$, unpaired Student *t* test. (*B*) Quantification of the total FAK intensity levels at focal adhesion sites upon knockdown of PIP5KIγ compared with Ctr. Data are mean values ± SEM. Unpaired Student *t* test. (*C*) Immunofluorescence staining of paxillin (green) and pFAK397 (red) in PIP5KIγ knockdown 1 (KD1) and KD2 clones and Ctr. Arrows indicate focal adhesion sites.

Table S1. Molecular weight determination from SAXS data

Protein	Concentration, mg/mL	$I(0)$, 10^{-3}	*Mw* (*i*),* $M_r \times 10^{-3}$	*Mw* (*ii*),† $M_r \times 10^{-3}$	Theoretical *Mw*, $M_r \times 10^{-3}$
Cytochrome C	7.25	0.164 (1)			12.4
	3.6	0.077809 (2)			12.4
FK-FAKwt	1.0	0.11655	64.5	67.5	75.4
	1.6	0.20732	71.0	74.3	75.4
	4.6	0.70401	83.9	87.8	75.4
FK-FAK180/183A	1.1	0.13347	66.5	69.6	75.2
	1.6	0.22452	76.9	80.5	75.2
	3.3	0.5158	85.7	89.7	75.2

Scattering intensity at scattering angle = 0 [$I(0)$] is determined by extrapolation from Guinier plots (1). The molecular weight (*Mw*) of FK-FAK proteins (*X*) is derived using the relationship between $I(0)$ and *Mw* and comparison with a known protein [here Cytochrome C (*CytC*)] as reported by Krigbaum and Kügler (2): $\frac{I(0)_{CytC}}{I(0)_X} = \frac{Mw_{CytC} \cdot C_{CytC}}{Mw_X \cdot C_X}$, where *C* is the concentration in milligrams per milliliters.

**Mw* (*i*) is determined with $I(0)_{CytC}$, where C_{CytC} = 7.25 mg/mL.

†*Mw* (*ii*) is determined with $I(0)_{CytC}$, where C_{CytC} = 3.6 mg/mL. FK-FAKwt and FK-FAK180/183A appear monomeric at the measured concentration range.

1. Guinier A (1938) The diffusion of x-rays under the extremely weak angles applied to the study of fine particles and colloidal suspension. *C R Hebd Seances Acad Sci* 206:1374–1376.
2. Krigbaum WR, Kügler FR (1970) Molecular conformation of egg-white lysozyme and bovine alpha-lactalbumin in solution. *Biochemistry* 9(5):1216–1223.

Table S2. Summary of dynamic light scattering analysis

Sample	R_H, nm	*Mw*-R, $M_r \times 10^{-3}$	Modal distribution	Dispersity
F-FAKwt	3.46 ± 0.05	61 ± 2.4	Monomodal	Monodisperse
+ C8-PC	3.64 ± 0.07	69 ± 3.0	Monomodal	Monodisperse
+ C8-PIP2	4.4 ± 0.22	104 ± 14	Monomodal	Polydisperse
F-FAK-KAKTLRK	3.7 ± 0	72 ± 3.1	Monomodal	Moderate polydisperse
+ C8-PC	3.7 ± 0.06	75 ± 2.8	Monomodal	Monodisperse
+ C8-PIP2	3.6 ± 0.06	66 ± 2.1	Monomodal	Moderate polydisperse
FK-FAKwt	4.5 ± 0.11	111 ± 7.5	Monomodal	Monodisperse
+ C8-PIP2	7.7 ± 0.61	403 ± 71	Multimodal	Polydisperse

Mw-R, molecular weight estimated from R_H (based on an empirical curve of known globular proteins and their measured hydrodynamic radius); PIP2, phosphatidylinositol 4,5-bisphosphate 2; R_H, hydrodynamic radius [weighted average of bins comprising the highest intensity peak (histograms in Fig. S4)]. For polydisperse samples, the R_H and *Mw*-R are based on a weighted average of more than one species. The baseline parameter was used to judge the modal distribution (1), and the normalized polydispersity (%Pd = Pd/R_H) was used to classify the dispersity of the main peak according to Borgstahl and Gloria (1). The sum of squares error statistic was less than 50 for all measurements. For R_H and *Mw*-R, mean values of three independent experiments, along with the SE, are given.

1. Borgstahl GE, Gloria EO (2007) How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. *Methods in Molecular Biology, Macromolecular Crystallography Protocols*, ed Walker JM (Humana Press, Totowa, NJ), Vol 363, pp 109–130.

Table S3. X-ray diffraction data collection and refinement statistics

	Data collection (PDB ID code)	
	F-FAKwt (4CYE)	F-FAK-KAKTLRK (3ZDT)
Data Collection		
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c*; Å	43.9, 146.1, 69.1	41.6, 175.5, 60.0
α, β, γ;°	90.0, 97.0, 90.0	90.0, 93.9, 90.0
Resolution, Å	73–3.2 (3.37–3.2)*	50–3.15 (3.34–3.15)
R_{merge}	8.4 (33.4)*	9.7 (51.5)
I/σ (I)	7.9 (2.0)*	12.5 (2.6)
Completeness, %	99.9 (99.9)*	99.6 (99.6)
Redundancy	5.6 (5.7)*	3.77 (3.76)
Refinement		
Resolution, Å	73.06–3.20	43.88–3.15
No. of reflections	13,590	14,023
R_{work}/R_{free}	17.3/24.0	22.3/27.8
No. of atoms	5,439	5,342
Protein	5,393	5,335
Water	46	7
Mean B-factor (overall), Å^2	58.6	79.6
rmsd		
Bond lengths, Å	0.008	0.010
Bond angles, °	1.133	1.808

Diffraction data were collected on one single crystal.
*Highest resolution shell is shown in parentheses.

7.2 Annex 2

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Design, Synthesis, and Evaluation of Novel Imidazo[1,2-*a*][1,3,5]triazines and Their Derivatives as Focal Adhesion Kinase Inhibitors with Antitumor Activity

Pascal Dao,[†] Nikaia Smith,[†] Céline Tomkiewicz-Raulet,[‡] Expédite Yen-Pon,[†] Marta Camacho-Artacho,[§] Daniel Lietha,[§] Jean-Phillipe Herbeuval,[†] Xavier Coumoul,[‡] Christiane Garbay,[†] and Huixiong Chen*[,†,∥]

[†]Chemistry & Biology, Nucleo(s)tides & Immunology for Therapy (CBNIT), CNRS UMR8601, Université Paris Descartes, PRES Sorbonne Paris Cité, UFR Biomédicale, 45 rue des Saints-Pères, 75270 Cedex 06 Paris, France

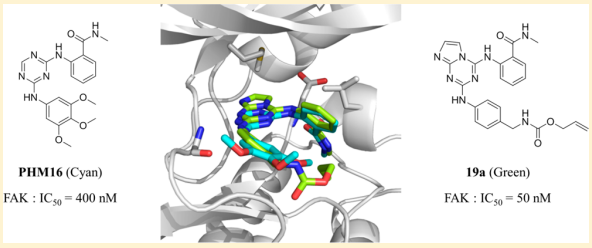
[‡]Toxicologie, Pharmacologie et Signalisation Cellulaire, INSERM, UMR S 1124, Université Paris Descartes, PRES Sorbonne Paris Cité, UFR Biomédicale, 45 rue des Saints-Pères, 75270 Cedex 06 Paris, France

[§]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Calle Melchor Fernández Almagro 3, Madrid 28029, Spain

[∥]School of Chemical Engineering and Light Industry, Guangdong University of Technology, no. 100 Waihuan Xi road, Education Mega Center, Guangzhou, 510006, China

Supporting Information

ABSTRACT: A series of triazinic inhibitors of focal adhesion kinase (FAK) have been recently shown to exert antiangiogenic activity against HUVEC cells and anticancer efficacy against several cancer cell lines. We report herein that we further explored the heterocyclic core of these inhibitors by a fused imidazole ring with the triazine to provide imidazo[1,2-*a*][1,3,5]triazines. Importantly, these new compounds displayed 10^{-7}–10^{-8} M IC_{50} values, and the best inhibitor showed IC_{50} value of 50 nM against FAK enzymatic activity. Several inhibitors potently inhibited the proliferation of a panel of cancer cell lines expressing high levels of FAK. Apoptosis analysis in U87-MG and HCT-116 cell lines suggested that these compounds delayed cell cycle progression by arresting cells in the G2/M phase of the cell cycle, retarding cell growth. Further investigation demonstrated that these compounds strongly inhibited cell-matrix adhesion, migration, and invasion of U87-MG cells.



INTRODUCTION

Focal adhesion kinase (FAK), a cytoplasmic tyrosine kinase and scaffold protein localized to focal adhesions, is uniquely positioned at the convergence point of integrins and receptor tyrosine kinase signal transduction pathways, which transmit signals from the extracellular matrix (ECM) to the cell cytoskeleton. FAK has been identified as a key pathogenic mediator of a variety of diseases. The most widely studied disease in this regard is cancer. This protein is emerging as a promising therapeutic target because it is highly expressed at both the transcriptional and translational level in various cancers. Ample evidence has indicated that FAK signaling pathways can stimulate tumor progression and metastasis through their regulation of cell migration, invasion, ECM, and angiogenesis (for a recent review of FAK, refer to refs 1–4). The initial attempts at inhibition focused upon down-regulation of FAK expression, demonstrating that silencing FAK by transfection of dominant-negative C-terminal FAK-CD decreased adhesion, colonization, and tumor growth in breast cancer cells.[5] Similarly, the inhibition of FAK expression with antisense oligonucleotides and siRNA to FAK led to cell rounding, detachment, reduction of invasion, increased apoptosis, and finally induced *in vivo* tumorigenesis.[6,7] It was also observed that in two transgenic mouse models there is a strong correlation between FAK expression and late-stage tumorigenesis, indicating that FAK is required for progression of early tumors to late-stage adenocarcinoma.[8,9] These studies confirmed scientific rationale for further development of FAK inhibitors as anticancer agents.

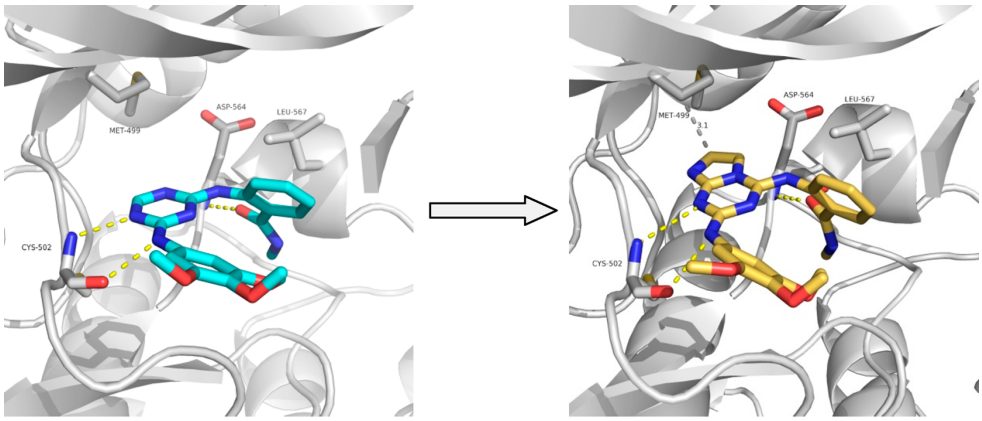
FAK has also been involved in angiogenesis as an important modulator during development, as evidenced by the early embryonic lethality of mice engineered to harbor an endothelial specific deletion of this protein.[10] FAK signaling pathway was connected to increased production of vascular endothelial growth factor (VEGF) and urokinase plasminogen activator, known factors involved in tumor progression.[11,12] It was

Special Issue: New Frontiers in Kinases

Received: May 20, 2014
Published: September 2, 2014

 © 2014 American Chemical Society

Chart 1



PHM16 **Imidazotriazinic compound**

reported that FAK expression in endothelial cells is necessary for the formation of new blood vessels, for the stability of the vascular network, and for the survival of endothelial cells.[13] Overexpression of FAK in vascular endothelial cells directly promotes angiogenesis in transgenic mice.[14] On the other hand, endothelial FAK deletion in adult mice inhibited tumor growth and reduced tumor angiogenesis.[15] A correlation between microvessel density on tumor specimen and FAK expression was found in a clinical study with breast cancer patients.[16] Therefore, several FAK inhibitors have been successfully developed,[17−20] and two of them were extensively examined in preclinical or clinical studies.[21,22] Their efficacy in tumor models may be a result of their ability to potently inhibit both tumor growth and tumor-associated angiogenesis.

Previously, we have described the generation of a series of triazinic inhibitors of FAK, which show antiangiogenic activity against HUVEC cells and anticancer efficacy against several cancers in cell culture studies (glioblastoma U-87MG, colon cancer HCT-116, breast cancer MDA-MB-231, and prostate cancer PC-3).[18,19] X-ray crystallographic analysis of the cocrystal structure of these compounds in the FAK kinase domain revealed that the mode of interaction is highly similar to that observed in the complex of TAE-226,[23] a potent ATP-competitive inhibitor of FAK designed by Novartis Pharma AG that stabilizes an unusual helical conformation of the DFG motif in which the φ torsion angle of Asp564 is rotated by 113° compared with the active kinase domain. This compound has shown potent antiproliferative and antitumor effects *in vitro* and *in vivo* in several types of malignancies including brain tumors,[24] esophageal cancer,[25] breast cancer,[26] ovarian cancer,[17] and gastrointestinal stroma tumor.[22] However, TAE-226 never entered clinical trials, because this compound also inhibited the insulin receptor with an IC_{50} of 40 nM and showed important side effects in animal studies including severely affected glucose metabolism and glucose blood levels.[27] Interestingly our triazinic compounds did not inhibit the insulin receptor (IR) kinase at 1 μM. Compared with TAE-226, our compounds showed similar antiangiogenic and anticancer activities but a poorer affinity for FAK with IC_{50} values in the range of 10^{-7} M, which could be worthy of further development.

In an attempt to design compounds with inhibitory potency enhancement toward FAK kinase activity, we decided to explore the heterocyclic core by the fused imidazole ring with the triazine to give an imidazo[1,2-*a*][1,3,5]triazine (Chart 1), which we hoped to have potentially an interaction with a gatekeeper residue Met499 in the kinase domain, and we also expected to retain the functional selectivity against the IR because the poorer affinity of 1,3,5-triazinic inhibitors for FAK might be attributed to the missing chlorine atom, which in TAE226 makes van der Waals interactions with this residue and significantly contributes to binding capacity.

Molecular docking is an application where the lowest energy binding mode of a small molecule bound to a protein target is predicted based on computationally calculated ligand−protein interactions. We first docked one imidazotriazinic compound into FAK based on the cocrystal structure of PHM16 with the FAK kinase domain (PDB ID 4c7t, see Chart 1). This compound could bind to the ATP-binding site in a very similar way to PHM16 and fits well in the nucleotide binding pocket with the imidazotriazine ring located in the adenine pocket and the imidazole ring oriented toward residue Met499 (3.1 Å). We hypothesized that the imidazotriazine cycle could be used as a good scaffold to develop novel FAK inhibitors. Recently, 1*H*-pyrrolo[2,3-*b*]- and 3*H*-imidazolo[4,5-*b*]pyridine scaffolds were also used to prepare FAK inhibitors with good affinity.[28]

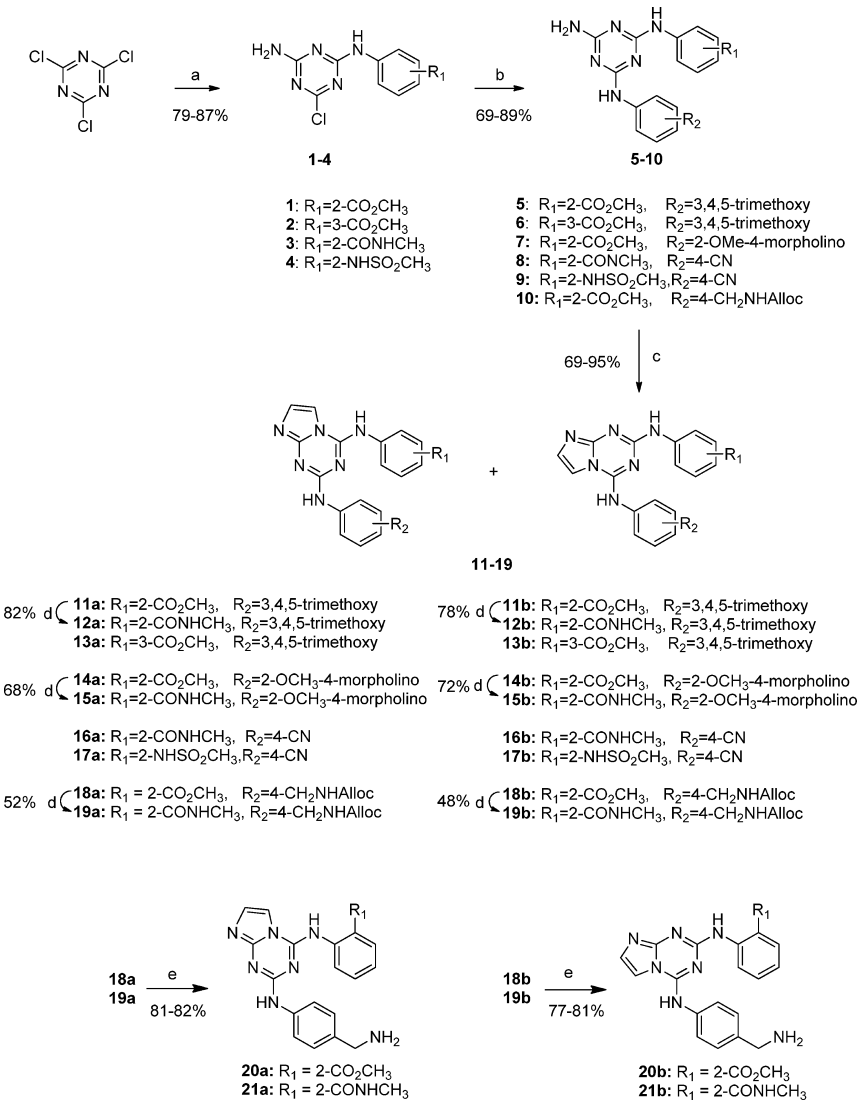
This paper describes the design, the synthesis, the biological evaluation, and the docking study of imidazotriazinic compounds as novel FAK inhibitors. The compounds potently suppressed the enzymatic activities of FAK with IC_{50} values in the 10^{-7}−10^{-8} M range. The best inhibitor of this series of compounds showed an IC_{50} value of 50 nM. These compounds also strongly suppressed the proliferation of human cancer cells in which FAK expression was at high levels. Furthermore, the representative compounds potently suppressed cancer cell adhesion, migration, and invasion, indicating their potential to serve as new lead compounds for further anticancer drug discovery.

CHEMISTRY

Synthesis of imidazo[1,2-*a*][1,3,5]triazine scaffold is described in Schemes 1 and 2. Starting from cyanuric chloride (Scheme 1), the first chlorine was displaced by nucleophilic substitution

dx.doi.org/10.1021/jm500784e | *J. Med. Chem.* 2015, 58, 237−251

Scheme 1[a]



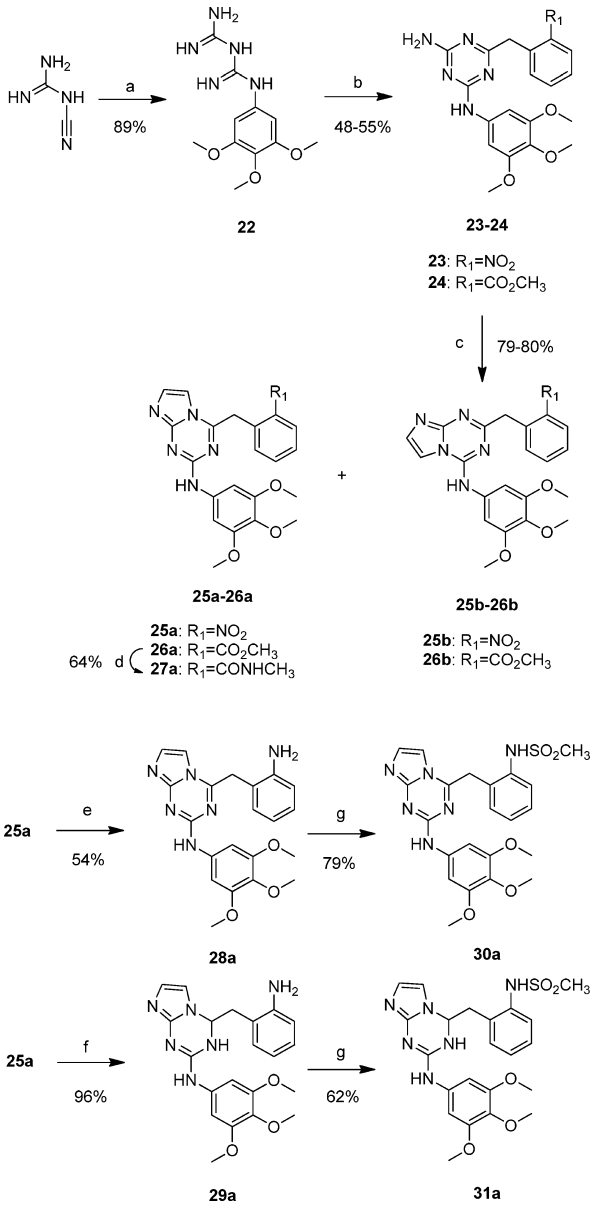
[a](a) Arylamine/THF/DIEA, −20°C, and NH_3/THF, TA; (b) arylamine/dioxane/$Pd(OAc)_2$/rac-BINAP/Cs_2CO_3/MW, 100 °C, 10 min; (c) 2-chloroacetaldehyde/EtOH/reflux; (d) $CH_3NH_2/CaCl_2$/MW, 100 °C, 20 min; (e) $Pd(PPh_3)_4$, $NaBH_4$, THF, TA.

with arylamines at −10 °C to produce the mono-substituted intermediates. Intermediates **1**−**4** were synthesized from cyanuric chloride (Scheme 1) by nucleophilic substitution with arylamines at −10 °C followed by reaction with ammonia in a one-pot procedure. The displacement of the last chlorine by the corresponding arylamines was accomplished by a Pd-catalyzed arylamination procedure, affording compounds **5**−**10** in good yields. Next, compounds **11**, **13**, **14**, and **16**−**18** were obtained in the form of two regioisomers by the condensation of compounds **5**−**10** with 4.2-fold excess of 2-chloroacetaldehyde. These two regioisomers were isolated by column chromatography, and their respective structures were confirmed by NOE experiments. In theory, it was possible from this cyclocondensation to obtain a pair of isomeric products that differ in the position of the imidazole ring, because the starting compounds **5**−**10** are differently substituted in the 4 and 6 positions. In general, the minor product is identified as the regioisomer B. Aminolysis of esters **11**, **14**, and **18** was accomplished using methylamine and $CaCl_2$ under MW irradiation to afford compounds **12**, **15**, and **19**. Finally, the protecting N-Alloc groups of compounds **18** and **19** were removed using a palladium catalyzed reductive deprotection procedure to produce the desired products **20** and **21**, respectively, in good yields.

The second series of imidazo[1,2-*a*][1,3,5]triazine derivatives (Scheme 2) were synthesized from compound **22**, which was prepared according to a previously published method.[29] Subsequent cyclization of compound **22** with methyl 2-nitrophenylacetate or methyl 2-(2-methoxy-2-oxoethyl)-benzoate was performed in methanolic sodium methoxide to afford trisubstituted triazines **23** and **24**. Next, compounds **25** and **26** were obtained in the form of two regioisomers by the condensation of compounds **23** and **24** with 2-chloroacetaldehyde as described in the first series. Aminolysis of ester **26a** led

dx.doi.org/10.1021/jm500784e | J. Med. Chem. 2015, 58, 237−251

Scheme 2[a]



[a](a) 3,4,5-Trimethoxyaniline/dioxane/MW, 90 °C, 15 min; (b) phenylacetate derivatives/MeONa/THF/reflux; (c) 2-chloroacetaldehyde/EtOH, 120 °C; (d) $Al(CH_3)_3/NH_4OH/CH_2Cl_2$, RT, 12 h; (e) $Na_2S_2O_4/NaHCO_3/THF/EtOH/H_2O$, RT; (f) $H_2/PtO_2/MeOH/THF$; (g) CH_3SO_2Cl/pyridine, 50 °C.

to the corresponding amide derivative **27a**. Finally, the synthesis of compounds **30a** and **31a** was accomplished by either chemical reduction using $Na_2S_2O_4$ or catalytic hydrogenation using PtO_2 and further substitution by methanesulfonyl chloride as described in Scheme 2.

RESULTS AND DISCUSSION

In Vitro Activity against FAK Kinase. The *in vitro* inhibition of FAK kinase activity of the newly synthesized compounds was evaluated using a TR-FRET based kinase assay as described previously.[18,19] One reported inhibitor of FAK, TAE-226, was included to validate the screening conditions. Under the experimental conditions, TAE-226 inhibited the activity of FAK with IC_{50} value of 7 nM (Table 1), which was similar to previously reported data.[27]

As shown in Table 1, comparison of regioisomers A and B of imidazo[1,2-*a*][1,3,5]triazines revealed that, in general, the first one is slightly more active as FAK inhibitor, especially for **19a**, which showed higher potency affinity to FAK (17.6-fold) than **19b**. It is interesting to note that the introduction of the 3,4,5-trimethoxyphenylamino group (ring B) at the position 2 of the imidazo[1,2-*a*][1,3,5]triazine ring and the phenylamino moiety (ring A) bearing an ester function at the position 4 of the same ring (compounds **11a**) resulted in a large increase in inhibitory potency on FAK kinase activity (IC_{50} 280 nM), compared with the corresponding 1,3,5-triazinic FAK inhibitor (IC_{50} 2.3 μM),[18] showing specific contributions of the heterocyclic core of imidazo[1,2-*a*][1,3,5]triazine. The displacement of the ester group from ortho (compound **11a**) to meta position in the arylamine on the imidazo[1,2-*a*][1,3,5]triazine ring A (compound **13a**) was not well tolerated and led to a decreased affinity toward FAK. This is consistent with what was observed for related 1,3,5-triazinic inhibitors.[18] However, replacing the ester group on the arylamine (ring A for compound **11a**) with an amide moiety (compound **12a**) displayed only a marginal improvement in inhibitory activity.

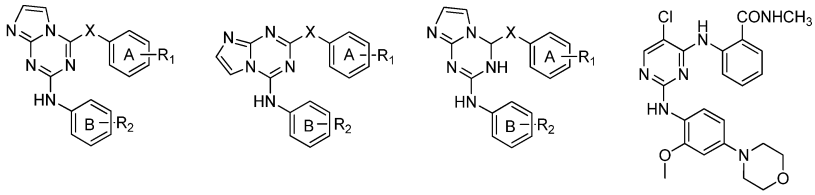
Then, we investigated the replacement of 3,4,5-trimethoxy groups on the phenyl ring (ring B) by different groups. When the 3,4,5-trimethoxyphenyl group (compound **12a**) is replaced by 2-methoxy-4-morpholinophenyl moiety, the corresponding compounds **15a** and **16a** did not display any relevant inhibitory activity changes. In contrast, increased inhibitory potencies are observed when the phenyl ring (ring B) is substituted by a CH_2NH_2 group or $CH_2NHAlloc$ group in the para position (**19a** and **21a** compared with **12a**), which led to a potent inhibitor of FAK (IC_{50} 50 nM for compound **19a**). This may be attributed to an additional interaction with the Glu506 residue of the enzymatic site, which could further enhance the binding capacity. The replacement of NH moiety between the phenyl group and the triazine moiety in the first series of imidazo[1,2-*a*][1,3,5]triazines by a CH_2 group leading to **26**, **27**, and **30** did not demonstrate an improved affinity for FAK (**26a** compared with **11a** and **27a** compared with **12a**), suggesting that the NH group did not play an important role for the inhibitory activity of FAK. Finally, a dihydroimidazo[1,2-*a*][1,3,5]triazine scaffold, a novel heterocyclic core, was not well tolerated and led to a decreased binding affinity toward FAK (**31a** compared with **30a**), showing that this scaffold is not favorable for the enzymatic site probably due to the dihydrotriazine ring, which is not planar.

Molecular Docking Study of Compound 19a. In an attempt to gain insight into the putative binding mode of the potent inhibitor **19a** with the FAK kinase, compared with its regioisomer **19b**, molecular docking of these compounds into the ATP binding site of the apo-FAK kinase (PDB ID 4c7t) was performed. The resulting 3D binding modes of compounds **19a**,**b** to FAK are depicted in Figure S1, Supporting Information.

For compound **19a**, the triazine ring, the two aniline rings (A and B), and the carboxamide moiety take very similar positions as observed for PHM16[19] bound to FAK. The two hydrogen bonds to the kinase hinge region made by PHM16 with its [1,3,5]triazine ring and the aniline moiety (ring B) are

dx.doi.org/10.1021/jm500784e | J. Med. Chem. 2015, 58, 237−251

Table 1. In Vitro Enzymatic Activities of Novel Imidazo[1,2-*a*][1,3,5]triazines Compared with TAE-226



no.	R^2	R^1	X	regioisomer	FRET IC_{50} (μM)
11a	3,4,5-OCH_3	2-CO_2CH_3	NH	A	0.28 ± 0.06
11b		2-CO_2CH_3	NH	B	0.57 ± 0.1
13a		3-CO_2CH_3	NH	A	0.87 ± 0.05
13b		3-CO_2CH_3	NH	B	1.20 ± 0.4
12a		2-$CONHCH_3$	NH	A	0.23 ± 0.08
12b		2-$CONHCH_3$	NH	B	0.31 ± 0.03
14a	2-OCH_3-4-morpholino	2-CO_2CH_3	NH	A	0.37 ± 0.03
14b		2-CO_2CH_3	NH	B	0.57 ± 0.07
15a		2-$CONHCH_3$	NH	A	0.33 ± 0.2
15b		2-$CONHCH_3$	NH	B	0.51 ± 0.04
16a	4-CN	2-$CONHCH_3$	NH	A	0.71 ± 0.06
16b		2-$CONHCH_3$	NH	B	1.12 ± 0.3
17a		2-$NHSO_2CH_3$	NH	A	0.92 ± 0.02
17b		2-$NHSO_2CH_3$	NH	B	1.73 ± 0.1
18a	4-CH_2NHAlloc	2-CO_2CH_3	NH	A	0.50 ± 0.02
18b		2-CO_2CH_3	NH	B	1.05 ± 0.1
19a		2-$CONHCH_3$	NH	A	0.05 ± 0.001
19b		2-$CONHCH_3$	NH	B	0.88 ± 0.02
20a	4-CH_2NH_2	2-CO_2CH_3	NH	A	0.24 ± 0.02
20b		2-CO_2CH_3	NH	B	0.28 ± 0.05
21a		2-$CONHCH_3$	NH	A	0.12 ± 0.02
21b		2-$CONHCH_3$	NH	B	0.22 ± 0.08
26a	3,4,5-OCH_3	2-CO_2CH_3	CH_2	A	0.37 ± 0.02
27a		2-$CONHCH_3$	CH_2	A	0.18 ± 0.01
30a		2-$NHSO_2CH_3$	CH_2	A	0.84 ± 0.02
31a		2-$NHSO_2CH_3$	CH_2	A	8.91 ± 0.9
TAE226					0.007 ± 0.002

conserved in **19a**; however, additional interactions are made by the imidazo[1,2-*a*] ring in **19a**, most notably with Met499 (Figure S1A, Supporting Information). The CO of the carboxamide group of **19a** is located near the DFG (D564-F565-G566) motif of the activation loop of the kinase domain and forms a hydrogen bond with the backbone nitrogen of Asp564 of the DFG motif. This hydrogen bond together with hydrophobic contacts between ring A and Leu567 stabilize the short helical conformation of residues 565−568 and the typical DFG conformation also seen in PHM16 and TAE226 bound structures, with the side chain of Asp564 pointing up toward the kinase N-lobe. The CH_2NH on the phenyl ring B forms a hydrogen bond with Ile-428, and the CO of the alloc group in **19a** has an additional interaction with Glu506. These interactions together with the interactions made by the imidazo[1,2-*a*] ring may be responsible for the increased binding affinity to FAK for **19a** compared with PHM16.

In contrast, **19b** cannot bind in the same mode as **19a**, because the imidazole ring is too close to Glu-500 in the hinge region, leading to steric clashes with this residue (Figure S1F, Supporting Information). As a result, the imidazo and triazine rings are pushed further out of the pocket, preventing hydrogen bonds between **19b** and the kinase hinge and the hydrogen bonds that are formed are much more solvent exposed and provide lower binding energy, likely explaining the lower affinity of **19b** compared with **19a**. As shown in Figure S1C,F, Supporting Information, **19a** but not **19b** fits well into the pocket as observed from a slice through the kinase just above the binding site.

Kinase Selectivity Profile of Compound 19a. We further profiled the best inhibitor, **19a**, *in vitro* against a panel of more than 30 kinases (ABL1, AKT1, ALK, CDK1, CDK2, CDK3, CDK4, CDK5, CHK1, EGFR, ERBB2, ERK1, ERK2, FGFR-1, FGFR-2, FGFR-3, FYN, GSK3β, IGF1R, IR, JAK3, c-Kit, LCK, c-Met, PDGFRA, Pyk2, RAF1, c-Src, mTOR/FRAP1, FLT1/VEGFR1, KDR/VEGFR2, and FLT4/VEGFR3) determined by Reaction Biology Corporation, Malvern, PA, USA to investigate its selectivity (Table S1, Supporting Information). A radioactive in vitro kinase assay was performed using [γ^{33}P]ATP and human recombinant protein in the presence of 1 μM of compound **19a**. The results revealed that compound **19a** displays an excellent selectivity profile against the tested kinases. As expected, this compound did not inhibit the IR (insulin receptor) kinase at 1 μM in contrast to TAE-226.[27] It is also interesting to note that at 1 μM **19a** did not inhibit Pyk2, which shares a highly similar structural organization and a high sequence homology with approximately 60% identity in the central catalytic domain and 40% identity in

dx.doi.org/10.1021/jm500784e | *J. Med. Chem.* 2015, 58, 237−251

both the N- and C-terminal domains with FAK.[30] This high affinity binding to FAK but not Pyk2 was also observed for the compound PF-573,228 reported by Pfizer.[31,32]

Inhibition of FAK Autophosphorylation and Growth Inhibitory Activity on Cancer Cell Lines. The antiproliferative effects of several compounds, compared with TAE-226, were investigated on human glioblastoma (U-87MG), human colon carcinoma (HCT-116), human metastatic breast cancer (MDA-MB-231), and human prostate cancer (PC-3), which express high levels of FAK.[33]

The inhibition of FAK autophosphorylation was first tested in these cancer cell lines, using FACEFAK ELISA kit (Active Motif Europe, Belgium). As shown in Figure 1, FAK

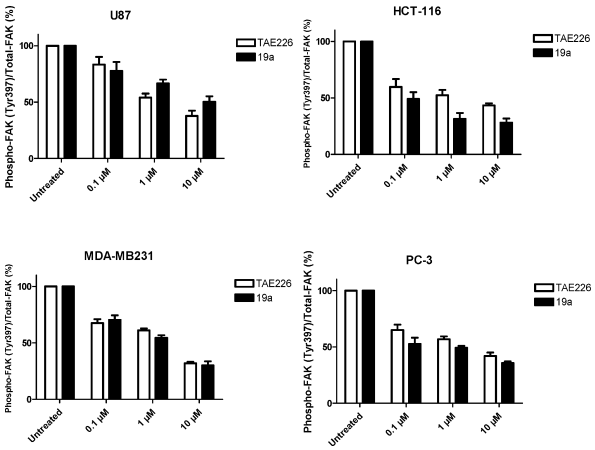


Figure 1. Cells (U-87MG, HCT-116, MDA-MB-231, and PC-3) were treated for 48 h with 0.1, 1, and 10 μM of compound **19a** or TAE-226 as control. Data obtained by the Face method are expressed as the fold decrease relative to the control and represent mean of triplicate readings of three independent experiments. A decrease of FAK tyrosine 397 phosphorylation in a dose-dependent manner is shown for **19a** and for TAE-226.

autophosphorylation was significantly inhibited by treatment with the best inhibitor, **19a**, compared with TAE-226. Consistent with its inhibitory activity shown against the FAK kinase, compound **19a** blocked tyrosine 397 phosphorylation in a dose-dependent manner in these cancer cell lines, suggesting that these inhibitors are able to effectively inhibit cellular FAK autophosphorylation and phosphorylation of kinase targets at low concentrations.

Then, we investigated the inhibition of cancer cell growth with several imidazo[1,2-*a*][1,3,5]triazine compounds using the WST-1 assay. As shown in Table 2, the IC_{50} values of these compounds against different tumor cells ranged between 0.27 and 4.2 μM except for compounds **12a** and **15a** to which MDA-MB-231 cells are not sensitive in terms of decrease in cell survival. Among them, U-87MG cells were the most sensitive to

Table 2. Antiproliferation Activity, IC_{50} (μM)

no.	U-87MG	HCT-116	MDA-MB-231	PC-3
12a	4.2 ± 0.3	0.95 ± 0.2	>10	0.35 ± 0.19
15a	2.31 ± 0.2	0.34 ± 0.08	>10	0.27 ± 0.08
19a	2.89 ± 0.3	0.52 ± 0.21	0.80 ± 0.27	0.29 ± 0.17
21a	0.37 ± 0.04	0.31 ± 0.18	1.08 ± 0.45	1.27 ± 0.28
TAE-226	1.2 ± 0.3	0.4 ± 0.08	2.8 ± 0.2	1.6 ± 0.2

21a with an IC_{50} value of 0.37 μM, significantly lower than that of TAE-226. HCT-116 cells were as sensitive to these compounds as to TAE-226, except for **12a**, which exhibited less cytotoxicity. PC-3 cells are more sensitive to these compounds than to TAE-226. MDA-MB-231 cells showed more sensitivity to compounds **19a** and **21a** than to TAE-226.

The effects of these compounds and TAE-226 on the tumorigenicity of these cancer cell lines were also examined using an *in vitro* colony formation assay. These compounds inhibited colony formation of the four cancer cell lines in a dose-dependent manner. In general, these compounds, like TAE-226, showed less cytotoxic effects but had stronger antitumorigenic effects on the cancer cell lines. As shown in Table 3, these compounds showed a similar effect on U-87MG

Table 3. Inhibition of Colony Formation, IC_{50} (μM)

no.	U-87MG	HCT-116	MDA-MB-231	PC-3
12a	0.22 ± 0.04	0.76 ± 0.06	3.9 ± 0.4	0.99 ± 0.2
15a	1.4 ± 0.1	0.02 ± 0.01	>10	0.63 ± 0.04
19a	0.42 ± 0.04	0.13 ± 0.02	0.14 ± 0.01	0.63 ± 0.04
21a	0.22 ± 0.03	0.12 ± 0.01	0.52 ± 0.03	1.4 ± 0.2
TAE-226	0.19 ± 0.01	0.23 ± 0.02	1.9 ± 0.2	0.26 ± 0.1

cells as TAE-226, except for **15a**. Similar results were also observed for HCT-116 cells, except for **15a**, which demonstrated a stronger antitumorigenic effect with an IC_{50} value of 20 nM, compared with TAE-226. The different cellular properties (cell permeability, intracellular stability, and distribution) or off-target effects might account for this strong inhibition of **15a** on HCT-116 cancer cells. Concerning MDA-MB-231 cancer cells, compounds **19a** and **21a** exhibited more potent anticancer activity than TAE-226. However, these compounds showed slightly weaker effects on PC-3 cells than TAE-226.

Effects of Compounds 19a and 21a on Apoptosis and on Cell Cycle in U-87MG and HCT-116 Cell Lines. To further investigate the mechanism underlying the antiproliferative effect of these compounds, we first examined whether the inhibitory effects of two good FAK inhibitors, **19a** and **21a**, on U-87MG and HCT-116 cell lines were due to apoptosis by examining the expression of annexin V using FACS analysis.

As shown in Figure 2, we tested apoptosis caused by these two compounds and compared them with TAE-226. In U-87MG cells, 3 μM **19a** and 0.3 μM **21a** did not induce significant apoptosis. This finding is similar to that reported in the literature[19,27] and suggests that these two compounds may disrupt cell cycle progression. Interestingly, when the HCT-116 cell line was exposed to **19a** at 3 μM, it resulted in a significant increase in apoptosis. Indeed, cell apoptosis was increased by 25.9% for **19a** compared with control. This effect is less potent than that observed with TAE-226 (41.8%). Low concentrations of **19a** and TAE-226 were also able to induce significant apoptosis in HCT-116 cell lines (Figure S2, Supporting Information). In contrast, **21a** did not induce significant apoptosis.

Next, we wished to determine whether **19a**- or **21a**-induced decrease of FAK phosphorylation would result in cell cycle arrest in U-87MG and HCT-116 cell lines, compared with TAE-226. Cell cycle analysis was performed using flow cytometry (Figure 3). Cell lines were treated with **19a** at 3 μM, **21a** at 0.4 μM, or TAE-226 at 1 μM. When the U-87MG cell line was exposed to **19a** or **21a** for 24 h, the DNA contents

dx.doi.org/10.1021/jm500784e | *J. Med. Chem.* 2015, 58, 237−251

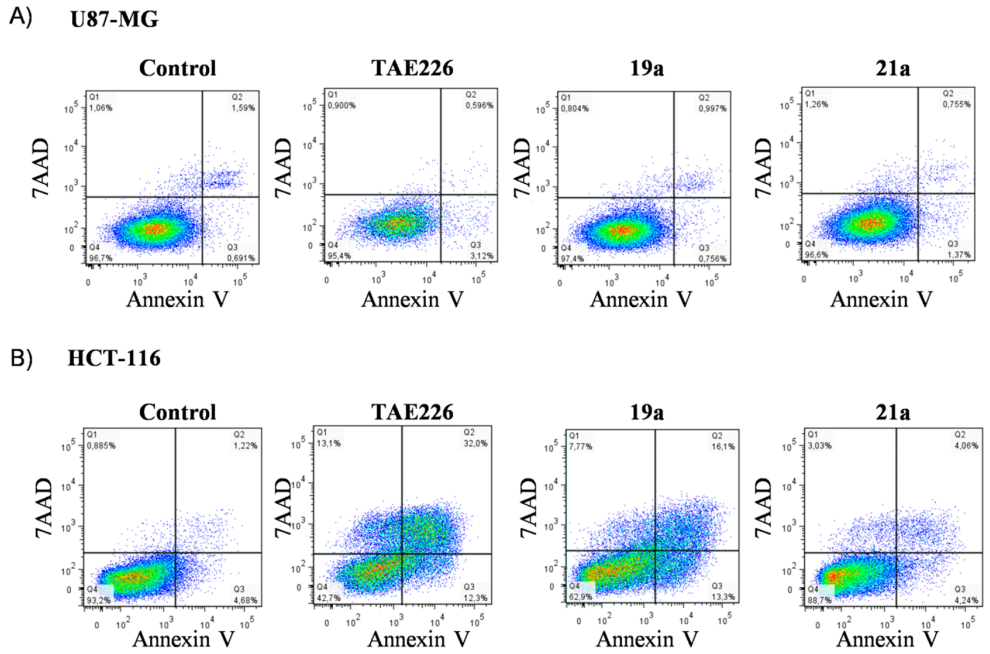


Figure 2. Two cancer cell lines (U87-MG and HCT-116) were plated at 10^5 cells per well in a 6-well culture dish, incubated overnight, and then cultured with **19a**, **21a**, or TAE-226 for 24 h. Cells were stained with annexin V and 7AAD and analyzed by FACS calibur. Cells were then classified as alive with no apoptosis (i.e., 7AAD negative, annexin V negative), dead (i.e., 7AAD positive), or apoptotic (i.e., annexin V positive). (A) U87-MG cells were treated with 3 μM **19a**, 0.3 μM **21a**, or 1 μM TAE-226; (B) HCT-116 cells were treated with 0.5 μM **19a**, 0.4 μM **21a**, or 0.3 μM TAE-226. One representative experiment is shown.

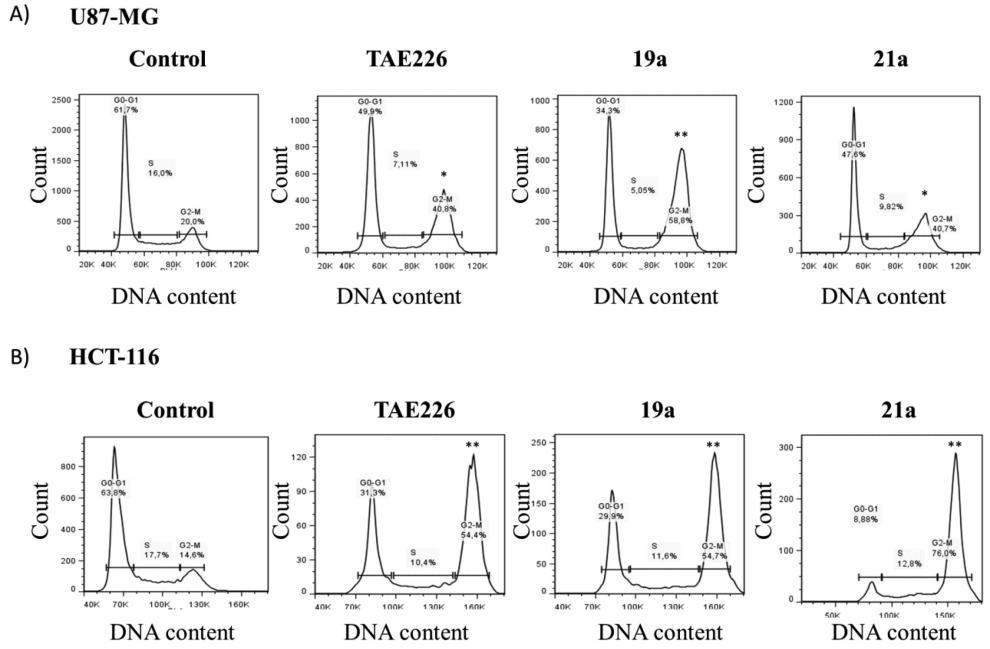


Figure 3. Cell cycle distribution was studied by flow cytometry: cells (U87-MG and HCT-116) were plated at 10^5 per well in a 6-well culture dish, incubated overnight, and then exposed to **19a**, **21a**, or TAE-226 along with a vehicle treatment control for 24 h. Cell cycle distribution was represented by histograms. Asterisks denote statistically significant differences ($*p < 0.01$; $**p < 0.001$) as determined by unpaired Student's t test. (A) U87-MG cells were treated with 3 μM **19a**, 0.3 μM **21a**, or 1 μM TAE-226; (B) HCT-116 cells were treated with 0.5 μM **19a**, 0.4 μM **21a**, or 0.3 μM TAE-226. One representative experiment is shown.

of the live population were 61.7%, 16.0%, and 20.0% for untreated cells, 34.3%, 5.1%, and 58.8% for **19a**-treated cells, and 47.6%, 9.8%, and 40.7% for **21a**-treated cells, respectively, for G0/G1, S, and G2/M phase. This finding is similar to that

dx.doi.org/10.1021/jm500784e | J. Med. Chem. 2015, 58, 237−251

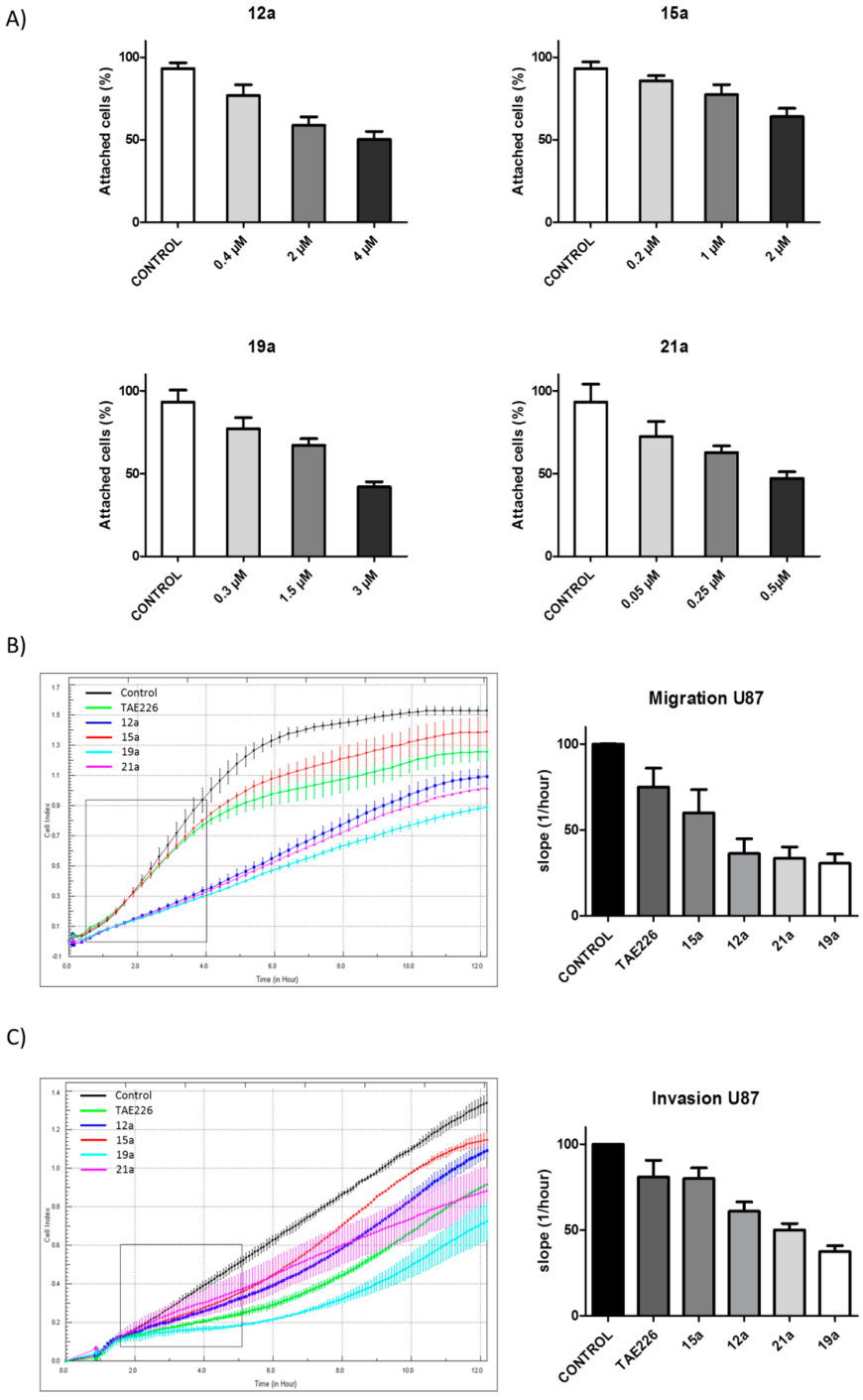


Figure 4. Effects of compounds **12a**, **15a**, **19a**, and **21a** on the cell attachment, migration, and invasion in U-87MG cell lines. (A) Cells were treated for 30 min in the presence of **12a**, **15a**, **19a**, **21a**, or TAE226. The attached cells were removed and counted. (B) Comparative xCELLigence analysis of the effect of different treatments on the migration of U87-MG human cancer cell lines. CI values are plotted on the graph. The slopes $(1/h)$, representing the cell migrating ability, were calculated based on CI values in the time interval indicated on the plots by the boxes. The mean slopes calculated with DMSO (control) were set to 100. The experiment was repeated two times and was consistent. (C) Comparative xCELLigence analysis of the effect of different treatment on the invasion of U87-MG human cancer cell lines. CI values are plotted on the graph. The slopes $(1/h)$, representing the cell invasion ability, were calculated based on CI values in the time interval indicated on the plots by the boxes. The mean slopes calculated with DMSO (control) were set to 100. The experiment was repeated two times and was consistent.

dx.doi.org/10.1021/jm500784e | *J. Med. Chem.* 2015, 58, 237−251

observed with TAE-226. Indeed, treatment with 3 μM **19a** or 0.4 μM **21a** resulted in a 2.9-fold and 2-fold, respectively increase in the fraction of cells arrested at G2. Similar results were also observed for the HCT-116 cell line. When these cells were treated with **19a** at 3 μM or **21a** at 0.4 μM, the number of cells in G2−M phase increased from 14.6% to 54.7% and 61.0%, whereas cells in G0/G1 phase decreased from 63.8% to 29.9% and 24.4%, respectively, compared with the controls. Low concentrations of **19a**, **21a**, and TAE-226 were also able to induce significant cell cycle arrest in U-87MG and HCT-116 cell lines (Figure S3, Supporting Information). Taken together, these results suggest that these compounds delayed cell cycle progression by arresting cells in the G2/M phase of the cell cycle, retarding U-87MG and HCT-116 cell growth.

Effects of Compounds 12a, 15a, 19a, and 21a on the Adherence, Migration, and Invasion of U-87MG and HCT-116 Cell Lines. Metastasis is multistep process that results from the convergence of a variety of cellular processes. Some of these processes include cell migration, invasion, adhesion into the surrounding stroma and through vessel walls, and survival in a foreign environment. Furthermore, the ability of tumor cells to migrate, invade, and metastasize is associated with increased FAK expression.[33] Indeed, a role for FAK in cell migration was first suggested by observations of its increased expression or activation in the migrating keratinocytes in epidermal wound healing or ECs migrating into the wounded monolayer in vitro, respectively.[34,35] Interference with FAK function via antisense oligonucleotides or by FRNK expression inhibits the motility of various cell types.[7,36] FAK−/− fibroblasts derived from FAK KO mouse embryo showed a significant decrease in cell migration compared with the cells from wild-type mice.[37] Because FAK is critical for cell attachment, migration, and invasion, we fist treated the U-87MG cells with different doses of our best inhibitors of FAK (**12a**, **15a**, **19a**, and **21a**) for the adhesion assay. We observed that these compounds caused dose-dependent decrease of attachment (Figure 4A). Among them, **21a** and **19a** showed strong potencies for adhesion inhibition. Indeed, at low dose of 0.5 μM **21a** or 3 μM **19a**, the attachment decreased about 50%, compared with that treated with TAE-226, which only decreased 36% of attachment at 1 μM (Figure S4, Supporting Information).

Next, we performed measurements of U-87MG cell migration and invasion using modified 16-well plates on the xCELLigence DP device form Roche Diagnostics. This assay is based on real time monitoring of cell migration or invasion, which captures cell responses during the entire course of an experiment. The cell index (CI) reflects the electrical impedance across the interdigitated microelectrodes integrated on the lower side of the 8 μm pore membrane, caused by the cells invading through the Matrigel or collagen, across the membrane and spreading on the lower side of the membrane. As shown in Figure 5B, treatment with 4 μM **12a**, 2 μM **15a**, 3 μM **19a**, and 0.4 μM **21a** significantly decreased the number of cells migrating through the collagen-coated Transwell membrane chambers after 12 h. Interestingly, these compounds were even more potent for the cell mobility reduction than TAE-226. Among them, **19a** and **21a** showed strong inhibition of U87-MG cell migration. Indeed, as shown by the slopes ($1/h$), representing the cell migration ability, the migration of U87-MG cells across collagen-coated Transwell membrane was reduced by 66.4% in the presence of 0.4 μM **21a** and 69.5% in the presence of 3 μM **19a** compared with the reduction of only 25.5% in the presence of 1 μM TAE-226. This result is consistent with that observed in the adhesion inhibition and also confirmed by Boyden chamber assay, which showed that our compounds inhibited cell migration in a dose-dependent manner (Figure S5A, Supporting Information).

Cell invasion is a prerequisite for establishment of metastases at secondary sites requiring movement through basement membranes.[38] To allow for measurement of cellular invasion, the membrane separating the upper and lower compartment of the CIM-plate 16 was coated with a layer of Matrigel, forcing the cell to digest the Matrigel prior to moving the bottom side of the membrane. As shown in Figure 4C, these compounds significantly reduced the invasive propensity of U87-MG cells compared with the vehicle control. Quantification of these results showed that compounds **19a** and **21a** could inhibit U87-MG cell invasion by at least 50% of untreated control, supporting their potent anti-invasion activity. As observed in the inhibition of cell migration, these compounds were more potent than TAE-226 in inhibiting cell invasion, except for compound **15a**, which showed the similar result as TAE-226. This result is consistent with that observed in the cellular migration and also confirmed by Boyden chamber assay, which showed that these compounds inhibited cell invasion in a dose-dependent manner (Figure S5B, Supporting Information). Taken together, these compounds can prevent cell adhesion, migration, and invasion of human U87-MG cells in vitro, which suggest that inhibition of FAK may uncouple adhesions from the substrate while still allowing them to be connected to the mobile actin network in the cell and this uncoupling would lead to defects in the generation or maintenance of tension at sites of adhesion, affecting thus cell migration, invasion, and tumor metastasis.[39]

CONCLUSION

In summary, a series of novel imidazo[1,2-*a*][1,3,5]triazines and their derivatives were found to be selective FAK inhibitors against the tested kinases. These compounds displayed IC_{50} values in the range of 10^{-7}−10^{-8} M and the best inhibitor showed IC_{50} values of 50 nM against FAK. Several inhibitors potently inhibited the proliferation of a panel of cancer cell lines expressing high levels of FAK, including U87-MG, HCT-116, MDA-MB-231, and PC-3. Apoptosis analysis in U87-MG and HCT-116 cell lines suggested that these compounds (**19a** and **21a**) may disrupt cell cycle progression but do not induce apoptosis as measured by 7AAD and annexin V, except for **19a** in the HCT-116 cell line. Indeed, using flow cytometry analysis, when the U-87MG and HCT-116 cell lines were exposed to these two compounds, the number of cells in G2−M phase increased and those in G0/G1 phase decreased compared with the controls. These observations suggest that they inhibited the cell growth through a mechanism associated with preventing progression through the G2−M phase of the cell cycle. Further investigation demonstrated that these compounds (**12a**, **15a**, **19a**, and **21a**) strongly inhibited cell-matrix adhesion, migration, and invasion in U87-MG cells in a dose-dependent manner. The careful evaluation of pharmacokinetics and pharmacodynamics *in vivo* will be required to fully understand the role of FAK inhibition in the regulation of tumorigenesis and metastatic progression. These compounds may provide useful tools to assess the role of adhesion signaling in the regulation of cellular proliferation and migration and may have important therapeutic applications in the prevention or treatment of cancer. More importantly, they will offer possible

dx.doi.org/10.1021/jm500784e | *J. Med. Chem.* 2015, 58, 237−251

alternative therapeutic approaches for the currently difficult treatment of distant metastases.

EXPERIMENTAL SECTION

All commercial materials were used without further purification. Microwave irradiation was carried out with a microwave monomode reactor (infrared detector for temperature). Melting points were determined on a Kofler apparatus as uncorrected values. Analytical thin-layer chromatography was performed on precoated 250 μm layer thickness silica gel 60 F254 plates and visualized with UV light. Column chromatography was performed using silica gel 60 (40−63 μm). ^{1}H NMR and ^{13}C NMR spectra were measured on 250 or 500 MHz spectrometer in DMSO-d_6 or $CDCl_3$ with chemical shift (δ) given in parts per million (ppm) relative to TMS as internal standard and recorded at 23 °C. MS (ESI) was determined by using a Q-Tof1 spectrometer with Z-spray source. High-resolution mass spectra (HRMS) were performed on Q-TOF Micro micromass positive ESI (CV = 30 V).

General Method for the Synthesis of Compounds 6−10. A solution of the corresponding compound (**2**−**4**,[40] 0.18 mmol, 1 equiv) with arylamine (0.18 mmol, 1 equiv) and Cs_2CO_3 (0.27 mmol, 1.5 equiv) was introduced into a 10 mL microwave vial. The mixture was purged with argon before the addition of $Pd(OAc)_2$ (7.2 nmol, 0.04 equiv) and BINAP (7.2 nmol, 0.04 equiv). The mixture was subjected to microwave irradiation for 10 min at 100 °C using irradiation power of 100 W. The solvent was evaporated under vacuum, and the crude residue was purified by silica gel column chromatography eluted with mixture of dichloromethane and methanol.

*Methyl 3-[4-Amino-6-(3,4,5-trimethoxyphenylamino)-1,3,5-triazin-2-ylamino]benzoate (**6**).* White solid, mp 150 °C, 124 mg (87%). ^{1}H NMR (250 MHz, DMSO): δ 9.25 (s, 1H), 8.97 (s, 1H), 8.33 (dd, J = 7.5 Hz, 1.5 Hz, 1H), 8.16 (s, 1H), 7.57 (dd, J = 7.5, 1.5 Hz, 1H), 7.42 (td, J = 7.5, 1.5 Hz, 1H), 7.16 (s, 2H), 6.68 (s, 2H), 3.85 (s, 3H), 3.74 (s, 6H), 3.64 (s, 3H).

*Methyl 2-[4-Amino-6-(2-methoxy-4-morpholinophenylamino)-1,3,5-triazin-2-ylamino]benzoate (**7**).* Violet solid, mp 174 °C, 132.5 mg (88%). ^{1}H NMR (250 MHz, DMSO): δ 10.27 (s, 1H), 7.95 (dd, J = 8.0, 1.5 Hz, 1H), 7.85 (s, 1H), 7.61 (d, J = 8.5 Hz, 1H), 7.5 (td, J = 7.5, 1.5 Hz, 1H), 7.03 (td, J = 7.5, 1.5 Hz, 1H), 6.72 (bs, 2H), 6.65 (d, J = 2.0 Hz, 1H), 6.50 (dd, J = 8.5, 2.0 Hz, 1H), 3.88 (s, 3H), 3.81 (s, 3H), 3.77 (t, J = 4.5 Hz, 4H), 3.13 (t, J = 4.5 Hz, 4H).

*2-[4-Amino-6-(4-cyanophenylamino)-1,3,5-triazin-2-ylamino]-N-methylbenzamide (**8**).* Pale yellow solid, mp 163 °C, 107 mg (89%).^{1}H NMR (250 MHz, DMSO): δ 9.52 (s, 1H), 8.41 (s, 1H), 8.04 (d, J = 4.5 Hz, 1H), 7.96 (d, J = 9.0, 2H), 7.94 (dd, J = 7.5, 1.5 Hz, 1H), 7.65 (d, J = 9 Hz, 2H), 7.41 (td, J = 7.5, 1.5 Hz, 1H), 7.35 (dd, J = 7.5, 1.5 Hz, 1H), 7.12 (td, J = 7.5, 1.5 Hz, 1H), 6.87 (bs, 2H), 2.65 (d, J = 4.5 Hz, 3H).

*N-{2-[4-Amino-6-(4-cyanophenylamino)-1,3,5-triazin-2-ylamino]phenyl}methanesulfonamide (**9**).* White solid, mp 184 °C, 95 mg (72%). ^{1}H NMR (250 MHz, DMSO): δ 9.69 (s, 1H), 8.43 (s, 1H), 8.15 (bs, 1H), 8.0 (d, J = 9.0 Hz, 2H), 7.95 (dd, J = 7.5, 1.5 Hz, 1H), 7.66 (d, J = 9 Hz, 2H), 7.35 (td, J = 7.5, 1.5 Hz, 1H), 7.30 (dd, J = 7.5, 1.5 Hz, 1H), 7.17 (td, J = 7.5, 1.5 Hz, 1H), 6.93 (bs, 2H), 2.94 (s, 3H).

*Methyl 2-{4-Amino-6-[4-((benzyloxycarbonylamino)methyl)-phenylamino]-1,3,5-triazin-2-ylamino}benzoate (**10**).* White solid, mp 192 °C, 104 mg (69%). ^{1}H NMR (250 MHz, DMSO): δ 10.38 (s, 1H), 9.30 (s, 1H), 8.95 (d, J = 7.5 Hz, 1H), 7.98 (dd, J = 7.5, 1.5 Hz, 1H), 7.73 (d, J = 7.5 Hz, 1H), 7.68 (d, J = 8.0 Hz, 2H), 7.57 (td, J = 7.5, 1.5 Hz, 1H), 7.18 (d, J = 8.0 Hz, 2H), 7.06 (td, J = 7.5, 1.5 Hz, 1H), 6.89 (br, 2H), 6.03−5.83 (m, 1H), 5.38−5.13 (m, 2H), 4.51 (dt, J = 5.3, 1.5 Hz, 2H), 4.16 (d, J = 6.0 Hz, 2H), 3.89 (s, 3H).

General Method for the Synthesis of Compounds 11, 13, 14, 16, 17, 18, 25, and 26. A solution of the corresponding compound (**5**[40]−**10** and **23**, 0.25 mmol, 1 equiv) and 2-chloroacetaldehyde (50% in H_2O, 0.07 mL, 1.05 mmol, 4.2 equiv) in DMSO was stirred at 120 °C for 5−7 h. The solvent was evaporated under vacuum, and the crude residue was treated with $NaHCO_3$ saturated aqueous solution. This alkaline solution was extracted with dichloromethane. The organic phase was dried over Na_2SO_4, reduced to a small volume, and purified by silica gel column chromatography eluted with a mixture of dichloromethane and methanol.

*Methyl 2-[2-(3,4,5-Trimethoxyphenylamino)imidazo[1,2-a][1,3,5]triazin-4-ylamino]benzoate (**11a**) and Methyl 2-[4-(3,4,5-Trimethoxyphenylamino)imidazo[1,2-a][1,3,5]triazin-2-ylamino]-benzoate (**11b**).* White solid, mp 196 °C, 70 mg (62%). ^{1}H NMR (250 MHz, DMSO): δ 10.47 (s, 1H), 9.11 (dd, J = 8.0, 1.5 Hz, 1H), 7.94 (dd, J = 8.0, 1.5 Hz, 1H), 7.70 (d, J = 1.5 Hz, 1H), 7.5 (td, J = 8.0, 1.5 Hz, 1H), 7.1 (d, J = 1.5 Hz, 1H), 7.01 (br, 2H), 6.95 (td, J = 8.0, 1.5 Hz, 1H), 3.87 (s, 3H), 3.80 (s, 6H), 3.67 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.2, 156.5, 152.3, 151.4, 147.5, 143.6, 133.9, 132.3, 130.5, 129.2, 119.3, 118.9, 112.7, 106.9, 101.0, 60.1, 55.5, 52.0. MS (ESI) m/z 451.2 $[M + 1]^+$. White solid, mp 198 °C, 37 mg (33%). ^{1}H NMR (250 MHz, DMSO): δ 10.82 (s, 1H), 10.05 (br, 1H), 8.95 (dd, J = 8.0, 1.5 Hz, 1H), 8.01 (d, J = 1.0 Hz, 1H), 8.0 (dd, J = 8.0, 1.5 Hz, 1H), 7.62 (td, J = 8.0, 1.5 Hz, 1H), 7.38 (d, J = 1.0 Hz, 1H), 7.23 (br, 2H), 7.08 (td, J = 8.0, 1.5 Hz, 1H), 3.87 (br, 9H), 3.71 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.4, 157.5, 153.6, 148.9, 141.3, 133.5, 130.2, 128.4, 124.3, 119.7, 107.5, 101.1, 60.2, 55.7, 52.0. MS (ESI) m/z 451.2 $[M + 1]^+$.

*Methyl 3-[2-(3,4,5-Trimethoxyphenylamino)imidazo[1,2-a][1,3,5]triazin-4-ylamino]benzoate (**13a**) and Methyl 3-[4-(3,4,5-Trimethoxyphenylamino)imidazo[1,2-a][1,3,5]triazin-2-ylamino]-benzoate (**13b**).* White solid, mp 154 °C, 65 mg (58%). ^{1}H NMR (250 MHz, DMSO): δ 9.9 (br, 1H), 9.64 (s, 1H), 8.38 (d, J = 2.0 Hz, 1H), 8.17 (dd, J = 7.5, 1.5 Hz, 1H), 7.93 (s, 1H), 7.56 (dd, J = 7.5, 1.5 Hz, 1H), 7.4 (td, J = 7.5, 1.5 Hz, 1H), 7.33 (d, J = 2.0 Hz, 1H), 7.02 (br, 2H), 3.84 (s, 3H), 3.80 (s, 6H), 3.70 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 52.0, 55.5, 55.5, 60.1, 101.0, 101.0, 106.9, 112.7, 118.9, 119.3, 129.2, 129.2, 130.5, 132.3, 133.9, 143.6, 147.5, 151.4, 152.3, 152.3, 156.5, 168.2. MS (ESI) m/z 451.2 $[M + 1]^+$. White solid, mp 164 °C, 30 mg (26%). ^{1}H NMR (250 MHz, DMSO): δ 10.11 (br, 1H), 9.32 (s, 1H), 8.39 (d, J = 2.0 Hz, 1H), 8.15 (td, J = 7.5, 1.5 Hz, 1H), 7.93 (dd, J = 7.5, 1.5 Hz, 1H), 7.75 (td, J = 7.5, 1.5 Hz, 1H), 7.57 (dd, J = 7.5, 1.5 Hz, 1H), 7.30 (d, J = 2.0 Hz, 1H), 7.20 (br, 2H), 3.87 (s, 3H), 3.68 (s, 6H), 3.62 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 52.0, 55.7, 55.7, 60.2, 101.1, 101.1, 107.5, 119.7, 119.7, 124.3, 128.4, 130.2, 130.2, 130.2, 133.5, 133.5, 141.3, 148.9, 148.9, 153.6, 157.5, 168.4. MS (ESI) m/z 451.2 $[M + 1]^+$.

*Methyl 2-[2-(2-Methoxy-4-morpholinophenylamino)imidazo-[1,2-a][1,3,5]triazin-4-ylamino]benzoate (**14a**) and Methyl 2-[4-(2-Methoxy-4-morpholinophenylamino)-imidazo[1,2-a][1,3,5]triazin-2-ylamino]benzoate (**14b**).* Violet solid, mp 168 °C, 67 mg (56%).^{1}H NMR (250 MHz, DMSO): δ 10.40 (s, 1H), 9.76 (s, 1H), 8.76 (d, J = 8.6 Hz, 1H), 7.97 (d, J = 1.6 Hz, 1H), 7.93 (d, J = 1.6 Hz, 1H), 7.5 (ddd, J = 8.5, 7.0, 1.6 Hz, 1H), 7.35 (s, 1H), 7.3 (d, J = 8.5 Hz, 1H), 7.03 (t, J = 8 Hz, 1H), 6.75 (d, J = 2.5 Hz, 1H), 6.63 (dd, J = 8.5, 2.5 Hz, 1H), 3.86 (s, 3H), 3.84−3.75 (m, 7H), 3.27−3.20 (m, 4H). ^{13}C NMR (500 MHz, DMSO): δ 168.7, 157.3, 155.0, 148.8, 148.2, 147.1, 141.5, 131.6, 131.5, 131.3, 124.3, 123.5, 123.2, 122.8, 122.4, 115.4, 107.0, 102.3, 65.5, 56.7, 52.0, 47.3. MS (ESI) m/z 476.2 $[M + 1]^+$. Violet solid, mp 152 °C, 33 mg (28%).^{1}H NMR (250 MHz, DMSO): δ 10.40 (s, 1H), 9.76 (s, 1H), 8.76 (d, J = 8.6 Hz, 1H), 7.97 (d, J = 1.6 Hz, 1H), 7.93 (d, J = 1.6 Hz, 1H), 7.5 (ddd, J = 8.5, 7.0, 1.6 Hz, 1H), 7.35 (s, 1H), 7.3 (d, J = 8.5 Hz, 1H), 7.03 (t, J = 8 Hz, 1H), 6.75 (d, J = 2.5 Hz, 1H), 6.63 (dd, J = 8.5, 2.5 Hz, 1H), 3.86 (s, 3H), 3.84−3.75 (m, 7H), 3.27−3.20 (m, 4H). ^{13}C NMR (500 MHz, DMSO): δ 168.8, 156.8, 155.0, 148.9, 148.9, 146.9, 141.8, 131.6, 131.5, 131.3, 124.1, 123.9, 123.6, 122.6, 115.1, 107.1, 101.8, 65.6, 56.8, 52.1, 47.3. MS (ESI) m/z 476.2 $[M + 1]^+$.

*2-[2-(4-Cyanophenylamino)imidazo[1,2-a][1,3,5]triazin-4-ylamino]-N-methylbenzamide (**16a**) and 2-[4-(4-Cyanophenylamino)-imidazo[1,2-a][1,3,5]triazin-2-ylamino]-N-methylbenzamide (**16b**).* Gray solid, mp 164 °C, 40 mg (42%). ^{1}H NMR (250 MHz, DMSO): δ 11.49 (br, 1H), 10.89 (s, 1H), 10.11 (s, 1H), 8.64 (d, J = 2.0 Hz, 1H), 7.93 (d, J = 9.0 Hz, 2H), 7.85−7.63 (m, 4H), 7.51 (d, J = 2.0 Hz, 1H), 7.48 (td, J = 7.5, 1.5 Hz, 1H), 7.19 (dd, J = 7.5, 1.5 Hz, 1H), 3.06 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 154.6, 148.9, 148.3, 141.1, 140.6, 132.2, 131.6, 130.6, 127.6, 126.9, 124.6, 123.8, 122.6, 119.1, 107.1, 103.3, 26.2. MS (ESI) m/z 385.1 $[M + 1]^+$. Gray

solid, mp 158 °C, 26 mg (27%). ^{1}H NMR (250 MHz, DMSO): δ 9.72 (br, 1H), 9.57 (s, 1H), 8.04 (dd, J = 8.0, 1.5 Hz, 1H), 7.95−7.75 (m, 4H), 7.68 (d, J = 8.0 Hz, 1H), 7.54 (td, J = 8.0, 1.5 Hz, 1H), 7.34 (d, J = 1.5 Hz, 1H), 7.25 (td, J = 8.0, 1.5 Hz, 1H), 6.84 (dd, J = 8.0, 1.5 Hz, 1H), 3.01 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 155.0, 148.9, 148.8, 141.1, 132.6, 131.6, 129.7, 129.5, 126.3, 124.8, 124.2, 122.2, 119.1, 107.1, 105.6, 26.2. MS (ESI) m/z 385.1 $[M + 1]^+$.

*N-{2-[2-(4-Cyanophenylamino)imidazo[1,2-a][1,3,5]triazin-4-ylamino]phenyl}methanesulfonamide (**17a**) and N-{2-[4-(4-Cyanophenylamino)imidazo[1,2-a][1,3,5]triazin-2-ylamino]-phenyl}methanesulfonamide (**17b**).* Gray solid, mp 166 °C, 46 mg (44%). ^{1}H NMR (250 MHz, DMSO): δ 9.77 (s, 1H), 7.92 (d, J = 9.0 Hz, 2H), 7.83 (d, J = 2.0 Hz, 1H), 7.62 (d, J = 9.0 Hz, 2H), 7.56−7.50 (m, 2H), 7.38 (d, J = 2.0 Hz, 1H), 7.30−7.19 (m, 2H), 2.97 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 154.6, 148.9, 148.3, 141.1, 131.6, 131.4, 130.6, 128.9, 127.5, 126.7, 123.8, 122.2, 121.6 119.1, 107.1, 103.3, 42.2. MS (ESI) m/z 421.1 $[M + 1]^+$. Gray solid, mp 162 °C, 32 mg (31%). ^{1}H NMR (250 MHz, DMSO): δ 8.74 (s, 1H), 7.87−7.65 (m, 5H), 7.37 (td, J = 7.5, 1.5 Hz, 1H), 7.34 (d, J = 2.0 Hz, 1H), 7.27 (dd, J = 7.5, 1.5 Hz, 1H), 7.19 (td, J = 7.5, 1.5 Hz, 1H), 2.92 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 155.0, 148.9, 148.8, 141.1, 132.5, 131.6, 129.5, 129.0, 128.2, 126.3, 124.8, 122.7, 121.6, 119.3, 107.2, 105.6, 42.3. MS (ESI) m/z 421.1 $[M + 1]^+$.

*Methyl 2-((2-((4-((((Vinyloxy)carbonyl)amino)methyl)phenyl)-amino)imidazo[1,2-a][1,3,5]-triazin-4-yl)amino)benzoate (**18a**) and Methyl 2-((4-((4-((((Vinyloxy)-carbonyl)amino)methyl)phenyl)-amino)imidazo[1,2-a][1,3,5]triazin-2-yl)amino)benzoate (**18b**).* White solid, mp 174 °C, 61 mg (52%). ^{1}H NMR (250 MHz, DMSO): δ 10.45 (s, 1H), 9.30 (s, 1H), 8.92 (dd, J = 7.5, 1.5 Hz, 1H), 8.05−7.95 (m, 4H), 7.55 (d, J = 2.0 Hz, 1H), 7.3 (d, J = 8.0 Hz, 2H), 7.03 (td, J = 7.5, 1.5 Hz, 1H), 7.18 (d, J = 8.0 Hz, 2H), 7.06 (td, J = 7.5, 1.5 Hz, 1H), 6.89 (br, 2H), 6.06−5.82 (m, 1H), 5.39−5.12 (m, 2H), 4.53 (dt, J = 5.3, 1.5 Hz, 2H), 4.23 (d, J = 6.0 Hz, 2H), 3.87 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.8, 158.6, 154.6, 148.9, 148.3, 141.6, 138.2, 134.9, 134.1, 131.6, 131.5, 131.3, 128.4, 123.2, 122.9, 122.4, 120.2, 117.4, 107.1, 66.0, 52.1, 43.5. MS (ESI) m/z 474.2 $[M + 1]^+$. White solid, mp 166 °C, 37 mg (31%). ^{1}H NMR (250 MHz, DMSO): δ 11.37 (s, 1H), 9.26 (s, 1H), 8.14 (dd, J = 7.5, 1.5 Hz, 1H), 7.81 (d, J = 2.0 Hz, 1H), 7.67−7.54 (m, 2H), 7.42 (td, J = 7.5, 1.5 Hz, 1H), 7.23 (d, J = 8.0, 1.5 Hz, 2H), 7.20 (dd, J = 7.5, 1.5 Hz, 1H), 7.03 (d, J = 8.0, 1.5 Hz, 2H), 6.52 (s, 1H), 5.96−5.77 (m, 1H), 5.29−5.10 (m, 2H), 4.63 (dt, J = 5.3, 1.5 Hz, 2H), 4.32 (d, J = 6.0 Hz, 2H), 3.90 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.8, 158.6, 155.0, 148.9, 148.8, 141.8, 137.2, 135.4, 134.1, 131.6, 131.5, 131.3, 128.0, 124.1, 122.6, 121.7, 117.4, 107.1, 66.0, 52.1, 43.5. MS (ESI) m/z 474.2 $[M + 1]^+$.

*4-(2-Nitrobenzyl)-N-(3,4,5-trimethoxyphenyl)imidazo[1,2-a]-[1,3,5]triazin-2-amine (**25a**) and 2-(2-Nitrobenzyl)-N-(3,4,5-trimethoxyphenyl)imidazo[1,2-a][1,3,5]triazin-4-amine (**25b**).* Yellow solid, mp 129 °C, 62 mg (57%). ^{1}H NMR (250 MHz, DMSO) δ 9.48 (s, 1H), 8.23 (dd, J = 7.5 Hz, 1.5 Hz, 1H), 7.90 (d, J = 2.0 Hz, 1H), 7.85 (td, J = 7.5 Hz, 1.5 Hz, 1H), 7.74 (dd, J = 7.5 Hz, 1.5 Hz, 1H), 7.69 (td, J = 7.5 Hz, 1.5 Hz, 1H), 7.43 (d, J = 2.0 Hz, 1H), 7.21 (br, 2H), 4.89 (s, 2H), 3.70 (s, 6H), 3.62 (s, 3H). Gray solid, mp126 °C, 25 mg (22%).^{1}H NMR (250 MHz, DMSO) δ 10.11 (s, 1H), 8.11 (d, J = 2.0 Hz, 1H), 8.05 (dd, J = 7.5 Hz, 1.5 Hz, 1H), 7.72 (td, J = 7.5 Hz, 1.5 Hz, 1H), 7.62 (dd, J = 7.5 Hz, 1.5 Hz, 1H), 7.57 (td, J = 7.5 Hz, 1.5 Hz, 1H), 7.54 (d, J = 2.0 Hz, 1H), 7.07 (br, 2H), 4.89 (s, 2H), 3.74 (s, 6H), 3.68 (s, 3H).

*Methyl 2-((2-(3,4,5-Trimethoxyphenylamino)imidazo[1,2-a]-[1,3,5]triazin-4-yl)methyl)-benzoate (**26a**) and Methyl 2-((4-((3,4,5-Trimethoxyphenyl)amino)imidazo[1,2-a][1,3,5]triazin-2-yl)-methyl)benzoate (**26b**).* White solid, mp 162 °C, 62 mg (55%). ^{1}H NMR (250 MHz, CDCl3): δ 8.15 (dd, J = 7.5, 1.5 Hz, 1H), 7.61 (td, J = 7.5, 1.5 Hz, 1H), 7.51 (td, J = 7.5, 1.5 Hz, 1H), 7.50 (d, J = 2.0 Hz, 1H), 7.43 (d, J = 2.0 Hz, 1H), 7.39 (dd, J = 7.5, 1.5 Hz, 1H), 7.06 (s, 2H), 4.72 (s, 2H), 3.88 (s, 6H), 3.83 (s, 3H), 3.78 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 167.2, 158.2, 155.0, 153.3, 150.3, 134.8, 134.6, 134.1, 134.0, 132.8, 132.3, 131.4, 129.6, 128.1, 107.0, 97.0, 61.0, 56.2, 52.3, 38.5. MS (ESI) m/z 450.2 $[M + 1]^+$. White solid, mp 154 °C, 27 mg (24%). ^{1}H NMR (250 MHz, CDCl3): δ 10.86 (s, 1H), 7.89 (dd, J = 7.5, 1.5 Hz, 1H), 7.63 (dd, J = 7.5, 1.5 Hz, 1H), 7.62 (d, J = 2 Hz, 1H), 7.54 (td, J = 7.5, 1.5 Hz, 1H), 7.37 (td, J = 7.5, 1.5 Hz, 1H), 7.34 (d, J = 2.0 Hz, 1H), 6.14 (s, 2H), 4.11 (s, 2H), 3.87 (s, 3H), 3.80 (s, 6H). ^{13}C NMR (500 MHz, DMSO): δ 175.0, 167.3, 165.1, 152.6, 147.5, 138.4, 134.6, 132.9, 132.8, 132.0, 131.9, 130.1, 130.0, 126.7, 106.9, 100.3, 60.0, 55.8, 51.8, 42.9. MS (ESI) m/z 450.2 $[M + 1]^+$.

General Method for the Synthesis of Compounds 12, 15, and 19. A solution of the corresponding compound (**11**, **14**, **18**, and **26a**, 0.05 mmol, 1 equiv), methylamine (2 M in MeOH, 10 equiv, 0.5 mmol), and $CaCl_2$ (0.15 mmol, 3equiv) in MeOH was introduced into a 10 mL microwave vial. The mixture was subjected to microwave irradiation for 20 min at 80 °C using irradiation power of 100 W. The solvent was evaporated under vacuum and the crude residue was purified by silica gel column chromatography eluted with mixture of dichloromethane and methanol.

*N-Methyl-2-[2-(3,4,5-trimethoxyphenylamino)imidazo[1,2-a]-[1,3,5]triazin-4-ylamino]benzamide (**12a**).* White solid, mp 186 °C, 18 mg (82%). ^{1}H NMR (250 MHz, DMSO): δ 11.37 (s, 1H), 9.98 (br, 1H), 8.86 (dd, J = 8.5, 1.5 Hz, 1H), 8.71 (q, J = 4.5 Hz, 1H), 8.03 (d, J = 2.0 Hz, 1H), 7.72 (dd, J = 8.0, 1.5 Hz, 1H), 7.5 (td, J = 8.0, 1.5 Hz, 1H), 7.35 (d, J = 2.0 Hz, 1H), 7.27 (br, 2H), 7.05 (td, J = 8.0, 1.5 Hz, 1H), 3.88 (s, 6H), 3.71 (s, 3H), 2.81 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.2, 156.5, 152.3, 151.4, 147.5, 143.6, 133.9, 132.3, 130.5, 129.2, 119.3, 118.9, 112.7, 106.9, 101.0, 60.1, 55.5, 26.2. MS (ESI) m/z 450.2 $[M + 1]^+$.

*N-Methyl-2-[4-(3,4,5-trimethoxyphenylamino)imidazo[1,2-a]-[1,3,5]triazin-2-ylamino]benzamide (**12b**).* White solid, mp 172 °C, 17 mg (78%). ^{1}H NMR (250 MHz, DMSO): δ 11.32 (s, 1H), 8.87 (dd, J = 8.0, 1.5 Hz, 1H), 8.71 (q, J = 4.5 Hz, 1H), 7.96 (d, J = 1.5 Hz, 1H), 7.71 (dd, J = 8.0, 1.5 Hz, 1H), 7.49 (td, J = 8.0, 1.5 Hz, 1H), 7.33 (d, J = 1.5 Hz, 1H), 7.24 (br, 2H), 7.04 (t, J = 7.5, 1.5 Hz, 1H), 3.88 (s, 6H), 3.71 (s, 3H), 2.81 (d, J = 4.7 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.4, 156.6, 156.7, 155.0, 148.9, 148.8, 141.1, 134.5, 134.4, 132.6, 131.6, 129.7, 126.3, 124.2, 122.2, 107.1, 102.8, 102.9, 60.6, 56.8, 56.9, 26.2. MS (ESI) m/z 450.2 $[M + 1]^+$.

*2-[2-(2-Methoxy-4-morpholinophenylamino)imidazo[1,2-a]-[1,3,5]triazin-4-ylamino]-N-methylbenzamide (**15a**).* White solid, mp 192 °C, 16 mg (68%). ^{1}H NMR (250 MHz, DMSO): δ 10.72 (s, 1H), 8.66 (d, J = 8.0 Hz, 1H), 8.61 (q, J = 4.5 Hz, 1H), 7.89 (s, 1H), 7.86 (d, J = 2.0 Hz, 1H), 7.62 (dd, J = 8.0, 1.5 Hz, 1H), 7.37 (td, J = 8.0, 1.5 Hz, 1H), 7.29 (s, 1H), 7.27 (d, J = 8.0 Hz, 1H), 6.99 (td, J = 8.0, 1.5 Hz, 1H), 6.61 (dd, J = 8.0, 1.5 Hz, 1H), 3.84−3.73 (m, 7H), 3.24 (m, 4H), 2.76 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 157.3, 155.0, 148.9, 148.3, 147.1, 140.6, 132.2, 131.6, 127.6, 126.9, 124.6, 124.3, 123.6, 122.6, 115.5, 107.1, 102.3, 65.6, 56.8, 47.3, 26.2. MS (ESI) m/z 475.2 $[M + 1]^+$.

*2-[4-(2-Methoxy-4-morpholinophenylamino)imidazo[1,2-a]-[1,3,5]triazin-2-ylamino]-N-methylbenzamide (**15b**).* White solid, mp 187 °C, 17 mg (72%). ^{1}H NMR (250 MHz, DMSO): δ 11.06 (s, 1H), 8.76 (q, J = 4.5 Hz, 1H), 8.40 (dd, J = 7.5, 1.5 Hz, 1H), 7.72 (td, J = 7.5, 1.5 Hz, 1H), 7.14 (d, J = 2.0 Hz, 1H), 7.10 (dd, J = 7.5, 1.5 Hz, 1H), 7.07 (td, J = 7.5, 1.5 Hz, 1H), 7.02 (d, J = 2.0 Hz, 1H), 6.89 (dd, J = 8.0, 1.5 Hz, 1H), 6.60 (dd, J = 8.0, 1.5 Hz, 1H), 6.05 (s, 1H), 3.86−3.70 (m, 7H), 3.42−3.20 (m, 4H), 2.86 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 156.8, 155.0, 148.9, 148.3, 146.9, 141.1, 132.6, 131.6, 129.7, 126.3, 124.2, 123.9, 123.6, 122.2, 115.1, 107.1, 101.8, 65.6, 56.8. MS (ESI) m/z 475.2 $[M + 1]^+$.

*Vinyl 4-((4-((2-(Methylcarbamoyl)phenyl)amino)imidazo[1,2-a]-[1,3,5]triazin-2-yl)amino)benzylcarbamate (**19a**).* White solid, mp 187 °C, 12 mg (52%). ^{1}H NMR (250 MHz, DMSO): δ 11.24 (s, 1H), 9.85 (q, J = 4.5 Hz, 1H), 7.82 (dd, J = 7.5, 1.5 Hz, 1H), 7.79 (d, J = 2.0 Hz, 2H), 7.67 (d, J = 2.0 Hz, 2H), 7.47 (td, J = 7.5, 1.5 Hz, 1H), 7.26 (dd, J = 7.5, 1.5 Hz, 1H), 7.23 (d, J = 8.0 Hz, 1H), 7.15 (td, J = 7.5, 1.5 Hz, 1H), 7.03 (d, J = 8.0 Hz, 1H), 6,97 (s, 1H), 5.99−5.82 (m, 1H), 5.46−5.16 (m, 2H), 4.73 (d, J = 6.0 Hz, 2H), 4.50 (dt, J = 5.3, 1.5 Hz, 2H), 2.82 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 158.6, 154.6, 148.9, 148.3, 140.6, 138.2, 134.9, 134.1, 132.2, 131.6, 128.4, 127.6, 126.9, 124.6, 122.6, 120.2, 117.4, 107.1, 66.0, 43.5, 26.2. MS (ESI) m/z 473.5 $[M + 1]^+$.

dx.doi.org/10.1021/jm500784e | J. Med. Chem. 2015, 58, 237−251

*Vinyl 4-((2-((2-(Methylcarbamoyl)phenyl)amino)-imidazo[1,2-a][1,3,5]triazin-4-yl)amino)benzylcarbamate (**19b**).* White solid, mp 179 °C, 11 mg (48%). ^{1}H NMR (250 MHz, DMSO): δ 9.76 (q, J = 4.5 Hz, 1H), 8.58 (s, 1H), 7.76 (dd, J = 7.7, 1.4 Hz, 2H), 7.65 (d, J = 7.5 Hz, 2H), 7.34 (td, J = 7.4, 1.4 Hz, 1H), 7.28 (d, J = 7.5 Hz, 2H), 7.13 (d, J = 2.2 Hz, 1H), 7.08 (dd, J = 7.5, 1.5 Hz, 1H), 7.07 (td, J = 7.5, 1.5 Hz, 1H), 7.01 (d, J = 7.5 Hz, 2H), 6.71 (s, 1H), 6.04 (s, 1H), 5.96−579 (m, 1H), 5.23−5.12 (m, 2H), 4.75 (d, J = 6.0 Hz, 2H), 4.33 (s, 2H), 2.79 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 158.6, 155.0, 148.9, 148.8, 141.1, 137.2, 135.4, 134.1, 132.6, 131.6, 129.7, 128.0, 126.3, 124.2, 122.2, 121.7, 117.4, 107.1, 66.0, 43.5, 26.2. MS (ESI) m/z 473.2 $[M + 1]^+$.

General Method for the Synthesis of Compounds 20 and 21. A solution of the corresponding compound (**18** or **19**, 0.2 mmol, 1 equiv) with morpholine (1 mmol, 5 equiv), triphenylphosphine (0.04 mmol, 0.2 equiv), and tetrakis(triphenylphosphine)palladium(0) (0.01 mmol, 0.05 equiv) in dichloromethane was stirred at room temperature for 4 h. The solvent was evaporated under vacuum, and the crude residue was purified by silica gel column chromatography eluted with a mixture of dichloromethane and methanol.

*Methyl 2-{2-[4-(Aminomethyl)phenylamino]imidazo[1,2-a][1,3,5]triazin-4-ylamino}-benzoate (**20a**).* White solid, mp 186 °C, 63 mg (82%). ^{1}H NMR (250 MHz, DMSO): δ 11.25 (s, 1H), 10.12 (s, 1H), 7.90 (dd, J = 7.5, 1.5 Hz, 1H), 7.65 (d, J = 2.0 Hz, 1H), 7.45 (td, J = 7.5, 1.5 Hz, 1H), 7.24 (d, J = 8.0 Hz, 2H), 7.22 (dd, J = 7.5, 1.5 Hz, 1H), 7.16 (d, J = 2.0 Hz, 1H), 7.10 (td, J = 7.5, 1.5 Hz, 1H), 7.00 (d, J = 8.0 Hz, 2H), 5.92 (br, 2H), 4.02 (s, 2H), 3.80 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.8, 154.6, 148.9, 148.3, 141.6, 137.0, 136.2, 131.6, 131.5, 131.3, 127.4, 123.2, 122.9, 122.4, 120.2, 107.1, 52.1, 46.2. MS (ESI) m/z 390.2 $[M + 1]^+$.

*Methyl 2-{4-[4-(Aminomethyl)phenylamino]imidazo[1,2-a][1,3,5]triazin-2-ylamino}benzoate (**20b**).* White solid, mp 164 °C, 60 mg (77%). ^{1}H NMR (250 MHz, DMSO): δ 11.06 (s, 1H), 10.65 (s, 1H), 7.82 (dd, J = 7.5, 1.5 Hz, 1H), 7.70 (d, J = 2.0 Hz, 1H), 7.52 (d, J = 2.0 Hz, 1H), 7.42 (td, J = 7.5, 1.5 Hz, 1H), 7.24 (d, J = 8.0 Hz, 2H), 7.08 (dd, J = 7.5, 1.5 Hz, 1H), 7.05 (td, J = 7.5, 1.5 Hz, 1H), 6.92 (d, J = 8.0 Hz, 2H), 5.75 (br, 2H), 4.00 (s, 2H), 3.88 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 168.8, 155.0, 148.9, 148.8, 141.8, 136.5, 136.2, 131.6, 131.5, 131.3, 127.2, 124.1, 122.6, 121.7, 107.1, 52.1, 46.2. MS (ESI) m/z 390.2 $[M + 1]^+$.

*2-{2-[4-(Aminomethyl)phenylamino]imidazo[1,2-a][1,3,5]triazin-4-ylamino}-N-methylbenzamide (**21a**).* White solid, mp 171 °C, 63 mg (81%). ^{1}H NMR (250 MHz, DMSO): δ 11.60 (s, 1H), 10.05 (s, 1H), 8.82 (q, J = 4.5 Hz, 1H), 7.87 (dd, J = 7.5, 1.5 Hz, 1H), 7.60 (d, J = 2.0 Hz, 1H), 7.42 (td, J = 7.5, 1.5 Hz, 1H), 7.32 (dd, J = 7.5, 1.5 Hz, 1H), 7.20 (d, J = 8.0 Hz, 2H), 7.18 (td, J = 7.5, 1.5 Hz, 1H), 7.12 (d, J = 2.0 Hz, 1H), 6.94 (d, J = 8.0 Hz, 2H), 5.70 (br, 2H), 3.95 (s, 2H), 2.82 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 154.6, 148.9, 148.3, 140.6, 137.0, 136.2, 132.2, 131.6, 127.6, 127.4, 126.9, 124.6, 122.6, 120.2, 107.1, 46.2, 26.2. MS (ESI) m/z 389.2 $[M + 1]^+$.

*2-{4-[4-(Aminomethyl)phenylamino]imidazo[1,2-a][1,3,5]triazin-2-ylamino}-N-methylbenzamide (**21b**).* White solid, mp 154 °C, 261.1 mg (82.6%). ^{1}H NMR (250 MHz, DMSO): δ 12.06 (s, 1H), 11.22 (s, 1H), 8.50 (q, J = 4.5 Hz, 1H), 7.78 (dd, J = 7.5, 1.5 Hz, 1H), 7.77 (d, J = 2.0 Hz, 1H), 7.52 (d, J = 2.0 Hz, 1H), 7.41(td, J = 7.5, 1.5 Hz, 1H), 7.22 (d, J = 8.0 Hz, 2H), 7.14 (td, J = 7.5, 1.5 Hz, 1H), 7.09 (dd, J = 7.5, 1.5 Hz, 1H), 6.88 (d, J = 8.0 Hz, 2H), 5.87 (br, 2H), 3.90 (s, 2H), 2.77 (d, J = 4.5 Hz, 3H). ^{13}C NMR (500 MHz, DMSO): δ 171.5, 155.0, 148.9, 148.8, 141.1, 136.5, 136.2, 132.6, 131.6, 129.7, 127.2, 126.3, 124.2, 122.2, 121.7, 107.1, 46.2. MS (ESI) m/z 389.2 $[M + 1]^+$.

General Method for the Synthesis of Compounds 23 and 24. A mixture of sodium methoxide (1.5 mmol, 1.5 equiv) prepared from Na and methanol, 3,4,5-trimethoxyphenylbiguanide chlorohydrate (1 mmol, 1 equiv), and methyl 2-(2-nitrophenyl)acetate or methyl 2-(2-methoxy-2-oxoethyl)benzoate (3 mmol, 3 equiv) in dry THF (3 mL) was introduced into a 50 mL round-bottomed flask equipped with a condenser and a magnetic stirring bar. The flask was placed in the microwave cavity and exposed to microwave irradiation for 20 min at 70 °C using irradiation power of 100 W. On cooling to room temperature, the mixture was evaporated under vacuum, and the residue was subjected to flash chromatography (silica gel, 5% methanol/CH_2Cl_2) to afford the desired product as a white solid.

*6-(2-Nitrobenzyl)-N2-(3,4,5-trimethoxyphenyl)-1,3,5-triazine-2,4-diamine (**23**).* White solid, mp 157 °C, 197 mg (48%). ^{1}H NMR (250 MHz, DMSO): δ 9.20 (s, 1H), 8.04 (dd, J = 7.5, 1.5 Hz, 1H), 7.71 (td, J = 7.5, 1.5 Hz, 1H), 7.59 (dd, J = 7.5, 1.5 Hz, 1H), 7.55 (td, J = 7.5, 1.5 Hz, 1H), 7.12 (br, 2H), 6.99 (br, 2H), 4.22 (s, 2H), 3.70 (s, 6H), 3.59 (s, 3H).

*Methyl 2-((4-Amino-6-(3,4,5-trimethoxyphenylamino)-1,3,5-triazin-2-yl)methyl)-benzoate (**24**).* White solid, mp 168 °C, 234 mg (55%). ^{1}H NMR (250 MHz, CDCl3): δ 7.97 (dd, J = 7.5, 1.5 Hz, 1H), 7.50 (td, J = 7.5, 1.5 Hz, 1H), 7.36 (td, J = 7.5, 1.5 Hz, 1H), 7.35 (dd, J = 7.5, 1.5 Hz, 1H), 6.79 (s, 2H), 4.35 (s, 2H), 3.82 (s, 6H), 3.79 (s, 3H).

***N*-Methyl-2-((2-(3,4,5-trimethoxyphenylamino)imidazo[1,2-*a*][1,3,5]triazin-4-yl)methyl)benzamide (27a).** A solution of methylamine (2.5 M in hexane, 0.25 mmol, 1 equiv) and NH_4OH (28% in H_2O, 0.25 mmol, 1 equiv) in CH_2Cl_2 was stirred for 15 min under argon at room temperature, and compound **26a** (0.25 mmol, 1 equiv) was added. The solvent was evapored under vacuum, the crude residue was extracted with dichloromethane, and the organic phase was purified by silica gel column chromatography eluted with a mixture of dichloromethane and methanol. White solid, mp 152 °C, 72 mg (64%). ^{1}H NMR (250 MHz, DMSO): δ 9.52 (s, 1H), 8.29 (q, J = 4.5 Hz, 1H), 7.79 (d, J = 1.5 Hz, 1H), 7.56 (d, J = 7.0 Hz, 1H), 7.50−7.39 (m, 3H), 7.34 (d, J = 1.5 Hz, 1H), 7.26 (s, 2H), 4.60 (s, 2H), 3.74 (s, 6H), 3.63 (s, 3H), 2.63 (d, J = 4.5 Hz, 3H). MS (ESI) m/z 449.2 $[M + 1]^+$.

4-(2-Aminobenzyl)-*N*-(3,4,5-trimethoxyphenyl)imidazo[1,2-*a*][1,3,5]triazin-2-amine (28a). A solution of compound **25a**[40] (0.2 mmol, 1 equiv) in THF (5 mL), water (2.5 mL), and EtOH (5 mL) with sodium dithionite (0.4 mmol, 2 equiv) and sodium bicarbonate (0.4 mmol, 2 equiv) was stirred at room temperature for 24 h. The mixture was evapored under vacuum, and the crude residue was purified by silica gel column chromatography and eluted with a mixture of dichloromethane and methanol to give 44 mg (54%), light-yellow solid, mp 234 °C. ^{1}H NMR (250 MHz, DMSO): δ 12.80 (s, 1H), 9.59 (s, 1H), 7.18 (d, J = 2.0 Hz, 1H), 7.01 (td, J = 7.5, 1.5 Hz, 1H), 6.96 (d, J = 1.5 Hz, 1H), 6.78−6.66 (m, 4H), 6.52 (td, J = 7.5, 1.5 Hz, 1H), 5.98 (t, J = 5 Hz, 2H), 3.76 (s, 6H), 3.63 (s, 3H), 3.22−2.94 (m, 2H).

4-(2-Aminobenzyl)-*N*-(3,4,5-trimethoxyphenyl)-3,4-dihydroimidazo[1,2-*a*][1,3,5]triazin-2-amine (29a). To a solution of compound **25a** (0.2 mmol, 1 equiv) in methanol (10 mL), 10% platinum oxide (8 mg) is added under a nitrogen atmosphere. The reaction vessel is charged with hydrogen and evacuated three times until the reaction is under a hydrogen atmosphere. The reaction is stirred overnight. The reaction mixture is filtered through a pad of Celite and washed with methanol. The filtrate is concentrated to a small volume and purified by flash chromatography to give 78 mg (96%), light-yellow solid, mp 260 °C. ^{1}H NMR (250 MHz, DMSO): δ 9.75 (s, 1H), 7.67 (d, J = 2.0 Hz, 1H), 7.51 (dd, J = 7.5, 1.5 Hz, 1H), 7.32 (s, 2H), 7.30 (s, 2H), 7.27 (d, J = 2.0 Hz, 1H), 7.13 (td, J = 7.5, 1.5 Hz, 1H), 7.04 (dd, J = 7.5, 1.5 Hz, 1H), 6.82 (td, J = 7.5, 1.5 Hz, 1H), 4.41 (s, 2H), 3.76 (s, 6H), 3.64 (s, 3H).

General Method for the Synthesis of Compounds 30a and 31a. A solution of the corresponding compound (**28a** or **29a**, 0.2 mmol, 1 equiv) in pyridine (2 mL) with acetic anhydride (0.4 mmol, 2 equiv) was stirred at 50 °C for 5 h. The solvent was evapored under vacuum, and the crude residue was purified by silica gel column chromatography eluted with a mixture of dichloromethane and methanol.

*N-{2-[4-(3,4,5-Trimethoxyphenylamino)imidazo[1,2-a][1,3,5]-triazin-2-ylmethyl]phenyl}methanesulfonamide (**30a**).* Light yellow solid, mp 204 °C, 76 mg (79%). ^{1}H NMR (250 MHz, DMSO): δ 11.54 (br, 1H), 9.01 (s, 1H), 8.87 (dd, J = 7.5, 1.5 Hz, 1H), 8.03 (d, J = 2.0 Hz, 1H), 7.37 (td, J = 7.5, 1.5 Hz, 1H), 7.36 (d, J = 2.0 Hz, 1H), 7.28 (td, J = 7.5, 1.5 Hz, 1H), 7.26 (dd, J = 7.5, 1.5 Hz, 1H), 7.16 (s,

2H), 4.30 (s, 2H), 3.68 (s, 9H), 2.97 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 41.0, 56.7, 56.7, 60.5, 101.5, 101.5, 127.8, 130.1, 130.7, 134.3, 134.8, 135.2, 136.3, 148.9, 157.5, 157.5, 161.3, 165.9, 169.6. MS (ESI) m/z 485.2 $[M + 1]^+$.

*N-{2-[2-(3,4,5-Trimethoxyphenylamino)-3,4-dihydro-imidazo-[1,2-a][1,3,5]triazin-4-ylmethyl]phenyl}acetamide (**31a**).* Yellow solid, mp 187 °C, 60 mg (62%). ^{1}H NMR (250 MHz, DMSO): δ 10.27 (s, 1H), 9.70 (s, 1H), 7.72 (d, J = 2.0 Hz, 1H), 7.54 (dd, J = 7.5, 1.5 Hz, 1H), 7.34 (s, 2H), 7.32 (d, J = 2.0 Hz, 1H), 7.10 (td, J = 7.5, 1.5 Hz, 1H), 7.04 (dd, J = 7.5, 1.5 Hz, 1H), 6.80 (td, J = 7.5, 1.5 Hz, 1H), 4.38 (s, 2H), 3.76 (s, 6H), 3.61 (s, 3H), 3.08 (s, 3H). ^{13}C NMR (500 MHz, DMSO): δ 161.0, 156.7, 152.5, 139.7, 134.5, 134.2, 132.0, 131.7, 131.2, 126.8, 125.2, 121.3, 110.9, 101.5, 101.0, 61.3, 56.9, 56.6. 42.3, 36.3. MS (ESI) m/z 487.2 $[M + 1]^+$.

Ligand Docking. Compounds **19a** and **19b** were docked into the apo-FAK inactive conformation (PDB ID 4c7t), which is similar to the TAE-226 bound conformation with the "DFG motif" adopting a helical conformation. The kinase receptor was kept rigid, while the flexible ligands were positioned in the binding pocket according to the PHM16 location. A standard protocol using Autodock 4.2[41] was carried out. In brief, the protein structure was screened within a 50 × 50 × 50 Å^3 cube centered at the "DFG motif" applying a space grid of 0.375 Å. The lowest energy and most populated clusters were selected after 100 cycles of running a Lamarckian genetic algorithm with a population size of 150. A maximum of 2.5 million energy evaluations was applied, and results were clustered using a tolerance of 2.0 Å.

In Vitro Kinase Assay. Ten microliters of assay mixture containing 100 nM FAK substrate ULight-poly GT, 0.1 nM FAK, 25 μM ATP, and 1 μL of compounds at desired concentrations in kinase buffer (50 mM Tris HCl, 1 mM EGTA, 10 mM $MgCl_2$, 2 mM DTT, 0.01% Tween 20, pH 7.4) was added into a 384-well plate. After incubation at 30 °C for 1.6 h, the kinase reaction was stopped by the addition of 5 μL of 40 mM EDTA in LANCE detection buffer 10× (Perkin Elmer) for 5 min and then 5 μL of the Eu-labeled antibody, 8 nM in detection buffer. The plate was incubated at 30 °C for 1 h, and the TR-FRET signal was detected with an Envision plate reader. For each compound, the IC_{50} value was determined from a sigmoid dose−response curve using Graph-Pad Prism (GraphPad Software, San Diego, CA, USA).

FACE FAK ELISA Assay. Cells (U-87MG, HCT-116, MDA-MB-231, and PC-3) were cultured in 96-well plates at 3 × 10^3 cells per well and treated with the compounds at different concentrations. Following treatment, the cells were fixed using 4% formaldehyde in PBS. Each well was then incubated with a primary antibody that recognizes either phosphorylated FAK or total FAK as suggested by the manufacturer. The phospho-FAK antibody recognizes FAK only when it is phosphorylated at Tyr397. The total-FAK antibody recognizes FAK regardless of its phosphorylation state. Subsequent incubation with secondary HRP-conjugated antibody and developing solution provided an easily quantified colorimetric response at 450 nm. The relative cell number in each well is then determined using the provided Crystal Violet solution and quantified at 595 nm. The levels of FAK phosphorylation were normalized by both the levels of total FAK protein and total cell number in each well.

Cell Culture and Reagents. Human glioma U-87MG (ATCC HTB-14), colorectal carcinoma HCT-116 (ATCC CCL-247), breast adenocarcinoma MDA-MB-231 (ATCC CRM-HTB-26), and prostate adenocarcinoma PC-3 (ATCC CRL-1435) were purchased from American Type Culture Collection (ATCC) and cultured at 37 °C in a 5% CO_2 humidified incubator and maintained in DMEM supplemented with 10% fetal bovine serum and 5% penicillin−streptomycin.

Cell Viability Assay. Cell viability was determined in 96-well plates by using the 4-[3-(4-iodophenyl)-2-(4-nitrophenyl)-2*H*-5-tetrazolio]-1,3-benzene disulfonate (WST-1) (Roche Diagnostics, Penzberg, Germany). The exponentially growing cells were seeded at 5 × 10^3 cells per well in 96-well plates and incubated overnight. The medium was then removed, and the cells were treated for 48 h with 200 μL of medium containing various concentrations of different compounds prepared in DMSO. Then, WST-1 was added to each well, and the cells were further incubated for 2 h. The optical density was analyzed with a microplate reader (Bio-Rad) at 450 nm to determine the cell viability; the reference wavelength was 620 nm. Data were calculated as the ratio of the values obtained for the treated cells to those for the controls (DMSO). The results are expressed as the mean of three independent experiments with three determinations per tested concentration per experiment. For each compound, the IC_{50} values were calculated using nonlinear regression model (logarithmic inhibitor vs normalized response-variable slope) in Graph-Pad Prism (GraphPad Software, San Diego, CA, USA).

Colony Assays. Cancer cell lines were seeded in 6-well plates at a density of 250 cells per well for HCT-116 and 300 cells per well for U-87MG, MDA-MB-231, and PC-3 and then incubated overnight. The medium was then removed before treatment at various concentrations of different compound for 2 weeks. The colonies were washed with PBS and fixed with 4% paraformaldehyde solution at room temperature for 30 min, stained with 0.1% crystal violet for 10 min, and finally, washed with water. Positive colony formation (more than 50 cells/colony) was counted, and colony formation rate was calculated. For each compound, the IC_{50} values were calculated using nonlinear regression model (logarithmic inhibitor vs normalized response-variable slope) in Graph-Pad Prism (GraphPad Software, San Diego, CA, USA).

Apoptosis Determination. Cell apoptosis was evaluated on U-87MG and HCT-116 cells using annexin V-PE (BD Pharmingen) in conjunction with the vital dye 7-amino-actinomycin D (7-AAD) to differentiate early apoptosis (annexin V$^+$ 7-AAD$^-$) from late apoptosis/necrosis (annexin V$^+$ 7-AAD$^+$). The cells were seeded in 24-well plates at 5 × 10^4 cells per well and incubated overnight before treatment for 48 h. The treated cells were collected and incubated in annexin V binding buffer with annexin V-PE and 7-AAD for 30 min at room temperature. The cells were then washed, resuspended in annexin V binding buffer, and analyzed by flow cytometry immediately on a BD LSR II (Becton Dickinson, San Jose, CA). Results were analyzed using FlowJo software (Treestar, Ashland, OR).

Flow Cytometer Analysis of Cell Cycle. Cell cycle was analysized on U-87MG or HCT-116 cells using propidium iodide. The cells were seeded in 24-well plates at 5 × 10^4 cells per well and incubated overnight before treatment for 48 h. The treated cells were collected, resuspended, and incubated for 15 min at 37 °C with PI/triton buffer. Cells were finally acquired by flow cytometer, and cell cycle was analyzed using Dean−Jett−Fox methods on a FACS calibur (Becton Dickinson, San Jose, CA).

Cell Adhesion Assay. Cells cultured in media containing 10% serum were collected by trypsinization and diluted to a concentration of 100 000 cells/mL in media containing 10% serum and then treated for 30 min with DMSO and each of compounds **12a**, **15a**, **19a**, and **21a** at various concentrations. Next, 200 μL of the treated cells with inhibitor (20 000 cells) was added to a 96-well plate and incubated for 1 h to allow attachment. Cells were fixed with paraformaldehyde 4% for 30 min and then stained by Toluidine Blue 0.1% in PFA 4% for 30 min. After washing with PBS, the optical density was analyzed with a microplate reader (Bio-Rad) at 620 nm to determine the attached cells.

xCELLigence Real-Time Cell Analysis (RTCA) of Migration and Invasion. Cell migration and invasion experiments were carried out using the xCELLigence RTCA DP instrument (Roche Diagnostics GmbH, Mannheim, Germany), which was placed in a humidified incubator at 37 °C and 5% CO_2. The experiments were performed using modified 16-well plates (CIM-16, Roche Diagnostics GmbH, Mannheim, Germany) with each well consisting of an upper and a lower chamber separated by a microporous membrane. Microelectrodes were attached to the underside of the membrane for impedance-based detection of migrated cells.

Prior to each experiment, U-87MG cells were cultured in 24-well plates at 5 × 10^4 cells per well, incubated overnight, and then exposed to different compounds or DMSO in 3% FBS for 24 h. Initially, 165 and 15 μL of media were added to the lower and upper chambers, respectively, and the CIM-16 plate was locked in the RTCA DP device at 37 °C and 5% CO_2 during 60 min to obtain equilibrium according to the manufacturer's guidelines. After this incubation period, a measurement step was performed as a background signal, generated by

cell-free media. To initiate an experiment, cells were harvested, resuspended in 3% FBS-medium, counted, and seeded in the upper chamber applying 4×10^4 cells in 100 μL. After cell addition, CIM-16 plates were incubated during 30 min at room temperature to allow the cells to settle onto the membrane according to the manufacturer's guidelines, they were locked in the RTCA DP device in the incubator, and the impedance value of each well was automatically monitored by the xCELLigence system and expressed as a cell index value (CI). Lower chambers contained media with 10% FBS in order to assess chemotactic migration when exposed to FBS. Each condition was performed in duplicate with a programmed signal detection schedule of every 5 min during 10 h of incubation.

An identical protocol was performed to follow the invasion, with the application of a layer of collagen on the upper side of the membranes and dynamic process follow-up during 12 hours. All data have been recorded by the supplied RTCA software (vs 1.2.1).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

In vitro profile of compound **19a** against a panel of ten kinases, purity data of tested compounds (HPLC), apoptosis in U-87MG and HCT-116 cells assessed by annexin V/7-AAD double staining, cell cycle distribution studied by flow-cytometry, effects of TAE-226 on the cell attachment in U-87MG cell lines, effects of compounds **TAE-226**, **12a**, **15a**, **19a**, and **21a** on the migration/invasion in U-87MG cell lines, using Boyden chamber assays, and experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +331-428-640-85. Fax: +331-428-640-82. E-mail: huixiong.chen@parisdescartes.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Agence Nationale Recherche (Grant RNA-NTO5-2_42590) and Visiotac Pharma.

■ ABBREVIATIONS

FAK, focal adhesion kinase; HUVEC, human umbilical vein endothelial cell; ECM, extracellular matrix; siRNA, small interfering RNA; VEGF, vascular endothelial growth factor; ATP, adenosine triphosphate; DFG, aspartate-phenylalanine-glycine; IR, insulin receptor; NOE, nuclear Overhauser effect; MW, microwave; RT, room temperature; TR-FRET, fluorescence resonance energy transfer; EGFR, epidermal growth factor receptor; Pyk2, proline-rich tyrosine kinase 2; FGFR2, fibroblast growth factor receptor 2; FLT1/VEGFR1, Fms-related tyrosine kinase 1/vascular endothelial growth factor receptor 1; IGF1R, insulin-like growth factor 1 receptor; KDR/VEGFR2, kinase insert domain receptor/vascular endothelial growth factor receptor 2; PDGFRA, α-type platelet-derived growth factor receptor; EGTA, ethylene glycol tetraacetic acid; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; PBS, phosphate buffered saline; DMEM, Dulbecco's modified Eagle medium

■ REFERENCES

(1) Zhang, J.; Hochwald, S. N. The Role of FAK in Tumor Metabolism and Therapy. *Pharmacol. Ther.* **2014**, *142*, 154−163.

(2) Cance, W. G.; Kurenova, E.; Marlowe, T.; Golubovskaya, V. Disrupting the Scaffold to Improve Focal Adhesion Kinase-Targeted Cancer Therapeutics. *Sci. Signaling* **2013**, *6*, No. pe10.

(3) Lechertier, T.; Hodivala-Dilke, K. Focal Adhesion Kinase and Tumour Angiogenesis. *J. Pathol.* **2012**, *226*, 404−412.

(4) Ma, W. W. Development of Focal Adhesion Kinase Inhibitors in Cancer Therapy. *Anticancer Agents Med. Chem.* **2011**, *11*, 638−642.

(5) Golubovskaya, V. M.; Zheng, M.; Zhang, L.; Li, J.-L.; Cance, W. G. The Direct Effect of Focal Adhesion Kinase (FAK), Dominant-Negative FAK, FAK-CD and FAK siRNA on Gene Expression and Human MCF-7 Breast Cancer Cell Tumorigenesis. *BMC Cancer* **2009**, *9*, No. 280.

(6) Xu, L. H.; Owens, L. V.; Sturge, G. C.; Yang, X.; Liu, E. T.; Craven, R. J.; Cance, W. G. Attenuation of the Expression of the Focal Adhesion Kinase Induces Apoptosis in Tumor Cells. *Cell Growth Differ.* **1996**, *7*, 413−418.

(7) Hauck, C. R.; Sieg, D. J.; Hsia, D. A.; Loftus, J. C.; Gaarde, W. A.; Monia, B. P.; Schlaepfer, D. D. Inhibition of Focal Adhesion Kinase Expression or Activity Disrupts Epidermal Growth Factor-Stimulated Signaling Promoting the Migration of Invasive Human Carcinoma Cells. *Cancer Res.* **2001**, *61*, 7079−7090.

(8) McLean, G. W.; Brown, K.; Arbuckle, M. I.; Wyke, A. W.; Pikkarainen, T.; Ruoslahti, E.; Frame, M. C. Decreased Focal Adhesion Kinase Suppresses Papilloma Formation during Experimental Mouse Skin Carcinogenesis. *Cancer Res.* **2001**, *61*, 8385−8389.

(9) Parsons, J. T.; Slack-Davis, J.; Tilghman, R.; Roberts, W. G. Focal Adhesion Kinase: Targeting Adhesion Signaling Pathways for Therapeutic Intervention. *Clin. Cancer Res.* **2008**, *14*, 627−632.

(10) Shen, T.-L.; Park, A. Y.-J.; Alcaraz, A.; Peng, X.; Jang, I.; Koni, P.; Flavell, R. A.; Gu, H.; Guan, J.-L. Conditional Knockout of Focal Adhesion Kinase in Endothelial Cells Reveals Its Role in Angiogenesis and Vascular Development in Late Embryogenesis. *J. Cell Biol.* **2005**, *169*, 941−952.

(11) Weis, S. M.; Cheresh, D. A. Pathophysiological Consequences of VEGF-Induced Vascular Permeability. *Nature* **2005**, *437*, 497−504.

(12) Duffy, M. J. The Urokinase Plasminogen Activator System: Role in Malignancy. *Curr. Pharm. Des.* **2004**, *10*, 39−49.

(13) Braren, R.; Hu, H.; Kim, Y. H.; Beggs, H. E.; Reichardt, L. F.; Wang, R. Endothelial FAK Is Essential for Vascular Network Stability, Cell Survival, and Lamellipodial Formation. *J. Cell Biol.* **2006**, *172*, 151−162.

(14) Peng, X.; Ueda, H.; Zhou, H.; Stokol, T.; Shen, T.-L.; Alcaraz, A.; Nagy, T.; Vassalli, J.-D.; Guan, J.-L. Overexpression of Focal Adhesion Kinase in Vascular Endothelial Cells Promotes Angiogenesis in Transgenic Mice. *Cardiovasc. Res.* **2004**, *64*, 421−430.

(15) Tavora, B.; Batista, S.; Reynolds, L. E.; Jadeja, S.; Robinson, S.; Kostourou, V.; Hart, I.; Fruttiger, M.; Parsons, M.; Hodivala-Dilke, K. M. Endothelial FAK Is Required for Tumour Angiogenesis. *EMBO Mol. Med.* **2010**, *2*, 516−528.

(16) Huang, Y.; Zhang, S. [Focal adhesion kinase expression and angiogenesis in breast carcinoma]. *Nan Fang Yi Ke Da Xue Xue Bao* **2007**, *27*, 1370−1373.

(17) Halder, J.; Lin, Y. G.; Merritt, W. M.; Spannuth, W. A.; Nick, A. M.; Honda, T.; Kamat, A. A.; Han, L. Y.; Kim, T. J.; Lu, C.; Tari, A. M.; Bornmann, W.; Fernandez, A.; Lopez-Berestein, G.; Sood, A. K. Therapeutic Efficacy of a Novel Focal Adhesion Kinase Inhibitor TAE226 in Ovarian Carcinoma. *Cancer Res.* **2007**, *67*, 10976−10983.

(18) Dao, P.; Jarray, R.; Le Coq, J.; Lietha, D.; Loukaci, A.; Lepelletier, Y.; Hadj-Slimane, R.; Garbay, C.; Raynaud, F.; Chen, H. Synthesis of Novel Diarylamino-1,3,5-Triazine Derivatives as FAK Inhibitors with Anti-Angiogenic Activity. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4552−4556.

(19) Dao, P.; Jarray, R.; Smith, N.; Lepelletier, Y.; Coq, J. L.; Lietha, D.; Hadj-Slimane, R.; Herbeuval, J.-P.; Garbay, C.; Raynaud, F.; Chen, H. Inhibition of Both Focal Adhesion Kinase and Fibroblast Growth Factor Receptor 2 Pathways Induces Anti-Tumor and Anti-Angiogenic Activities. *Cancer Lett.* **2014**, *348*, 88−99.

(20) Golubovskaya, V. M. A Small Molecule Inhibitor, 1,2,4,5-Benzenetetraamine Tetrahydrochloride, Targeting the Y397 Site of

Focal Adhesion Kinase Decreases Tumor Growth. *J. Med. Chem.* **2008**, *51*, 7405−7416.
(21) Roberts, W. G. Antitumor Activity and Pharmacology of a Selective Focal Adhesion Kinase Inhibitor, PF-562,271. *Cancer Res.* **2008**, *68*, 1935−1944.
(22) Sakurama, K.; Noma, K.; Takaoka, M.; Tomono, Y.; Watanabe, N.; Hatakeyama, S.; Ohmori, O.; Hirota, S.; Motoki, T.; Shirakawa, Y.; Yamatsuji, T.; Haisa, M.; Matsuoka, J.; Tanaka, N.; Naomoto, Y. Inhibition of Focal Adhesion Kinase as a Potential Therapeutic Strategy for Imatinib-Resistant Gastrointestinal Stromal Tumor. *Mol. Cancer Ther.* **2009**, *8*, 127−134.
(23) Lietha, D.; Eck, M. J. Crystal Structures of the FAK Kinase in Complex with TAE226 and Related Bis-Anilino Pyrimidine Inhibitors Reveal a Helical DFG Conformation. *PLoS One* **2008**, *3*, No. e3800.
(24) Shi, Q.; Hjelmeland, A. B.; Keir, S. T.; Song, L.; Wickman, S.; Jackson, D.; Ohmori, O.; Bigner, D. D.; Friedman, H. S.; Rich, J. N. A Novel Low-Molecular Weight Inhibitor of Focal Adhesion Kinase, TAE226, Inhibits Glioma Growth. *Mol. Carcinog.* **2007**, *46*, 488−496.
(25) Watanabe, N.; Takaoka, M.; Sakurama, K.; Tomono, Y.; Hatakeyama, S.; Ohmori, O.; Motoki, T.; Shirakawa, Y.; Yamatsuji, T.; Haisa, M.; Matsuoka, J.; Beer, D. G.; Nagatsuka, H.; Tanaka, N.; Naomoto, Y. Dual Tyrosine Kinase Inhibitor for Focal Adhesion Kinase and Insulin-like Growth Factor-I Receptor Exhibits Anticancer Effect in Esophageal Adenocarcinoma in Vitro and in Vivo. *Clin. Cancer Res.* **2008**, *14*, 4631−4639.
(26) Kurio, N.; Shimo, T.; Fukazawa, T.; Takaoka, M.; Okui, T.; Hassan, N. M. M.; Honami, T.; Hatakeyama, S.; Ikeda, M.; Naomoto, Y.; Sasaki, A. Anti-Tumor Effect in Human Breast Cancer by TAE226, a Dual Inhibitor for FAK and IGF-IR in Vitro and in Vivo. *Exp. Cell Res.* **2011**, *317*, 1134−1146.
(27) Liu, T.-J.; LaFortune, T.; Honda, T.; Ohmori, O.; Hatakeyama, S.; Meyer, T.; Jackson, D.; de Groot, J.; Yung, W. K. A. Inhibition of Both Focal Adhesion Kinase and Insulin-like Growth Factor-I Receptor Kinase Suppresses Glioma Proliferation in Vitro and in Vivo. *Mol. Cancer Ther.* **2007**, *6*, 1357−1367.
(28) Heinrich, T.; Seenisamy, J.; Emmanuvel, L.; Kulkarni, S. S.; Bomke, J.; Rohdich, F.; Greiner, H.; Esdar, C.; Krier, M.; Grädler, U.; Musil, D. Fragment-Based Discovery of New Highly Substituted 1*H*-Pyrrolo[2,3-*b*]- and 3*H*-Imidazolo[4,5-*b*]-pyridines as Focal Adhesion Kinase Inhibitors. *J. Med. Chem.* **2013**, *56*, 1160−1170.
(29) Dao, P.; Garbay, C.; Chen, H. Regioselective Synthesis of imidazo[1,2-a][1,3,5]triazines and 3,4-dihydroimidazo[1,2-a][1,3,5]-triazines from [1,3,5]triazin-2,4-Diamines. *Tetrahedron* **2013**, *69*, 3867−3871.
(30) Lev, S.; Moreno, H.; Martinez, R.; Canoll, P.; Peles, E.; Musacchio, J. M.; Plowman, G. D.; Rudy, B.; Schlessinger, J. Protein Tyrosine Kinase PYK2 Involved in Ca(2+)-Induced Regulation of Ion Channel and MAP Kinase Functions. *Nature* **1995**, *376*, 737−745.
(31) Kath, J. C.; Luzzio, M. J. Composes Pour Traiter Le Developpement Anormal de Cellules. International Patent WO2004056786 A2, July 8, 2004.
(32) Slack-Davis, J. K.; Martin, K. H.; Tilghman, R. W.; Iwanicki, M.; Ung, E. J.; Autry, C.; Luzzio, M. J.; Cooper, B.; Kath, J. C.; Roberts, W. G.; Parsons, J. T. Cellular Characterization of a Novel Focal Adhesion Kinase Inhibitor. *J. Biol. Chem.* **2007**, *282*, 14845−14852.
(33) Gabarra-Niecko, V.; Schaller, M. D.; Dunty, J. M. FAK Regulates Biological Processes Important for the Pathogenesis of Cancer. *Cancer Metastasis Rev.* **2003**, *22*, 359−374.
(34) Romer, L. H.; McLean, N.; Turner, C. E.; Burridge, K. Tyrosine Kinase Activity, Cytoskeletal Organization, and Motility in Human Vascular Endothelial Cells. *Mol. Biol. Cell* **1994**, *5*, 349−361.
(35) Gates, R. E.; King, L. E., Jr.; Hanks, S. K.; Nanney, L. B. Potential Role for Focal Adhesion Kinase in Migrating and Proliferating Keratinocytes near Epidermal Wounds and in Culture. *Cell Growth Differ.* **1994**, *5*, 891−899.
(36) Taylor, J. M.; Mack, C. P.; Nolan, K.; Regan, C. P.; Owens, G. K.; Parsons, J. T. Selective Expression of an Endogenous Inhibitor of FAK Regulates Proliferation and Migration of Vascular Smooth Muscle Cells. *Mol. Cell. Biol.* **2001**, *21*, 1565−1572.
(37) Ilic, D. Reduced Cell Motility and Enhanced Focal Adhesion Contact Formation in Cells from FAK-Deficient Mice. *Nature* **1995**, *377*, 539−544.
(38) Brábek, J.; Mierke, C. T.; Rösel, D.; Veselý, P.; Fabry, B. The Role of the Tissue Microenvironment in the Regulation of Cancer Cell Motility and Invasion. *Cell Commun. Signaling* **2010**, *8*, No. 22.
(39) Beningo, K. A.; Dembo, M.; Kaverina, I.; Small, J. V.; Wang, Y. L. Nascent Focal Adhesions Are Responsible for the Generation of Strong Propulsive Forces in Migrating Fibroblasts. *J. Cell Biol.* **2001**, *153*, 881−888.
(40) Dao, P.; Garbay, C.; Chen, H. High Yielding Microwave-Assisted Synthesis of Tri-Substituted 1,3,5-Triazines Using Pd-Catalyzed Aryl and Heteroarylamination. *Tetrahedron* **2012**, *68*, 3856−3860.
(41) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.

Supporting information

Design, synthesis and evaluation of novel imidazo[1,2-a][1,3,5]triazines and their derivatives as Focal adhesion kinase inhibitors with antitumor acitivity

Pascal Dao[a], Nikaia Smith[a], Céline Tomkiewicz-Raulet[b], Expédite Yen-Pon[a], Marta Camacho-Artacho[c], Daniel Lietha[c], Jean-Philippe Herbeuval[a], Christiane Garbay[a], Huixiong Chen[a,d]*

[a]*Chemistry & Biology, Nucleo(s)tides & Immunology for Therapy (CBNIT), CNRS UMR8601, Université Paris Descartes, PRES Sorbonne Paris Cité, UFR Biomédicale, 45 rue des Saints-Pères, 75270 Paris Cedex 06, France*

[b]*Toxicologie, Pharmacologie et Signalisation Cellulaire, INSERM, UMR S 1124, Université Paris Descartes, PRES Sorbonne Paris Cité, UFR Biomédicale, 45 rue des Saints-Pères, 75270 Paris Cedex 06, France*

[c]*Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Calle Melchor Fernández Almagro 3, Madrid 28029, Spain*

[d]*School of Chemical Engineering and Light Industry,, Guangdong University of Technology, n° 100 Waihuan Xi road, Education Mega Center, Guangzhou, 510006, China*

Table S1. In vitro profile of compound **19a** against a panel of ten kinases

Kinase	Substrate	% inhibition at 1 μM	IC_{50} (μM)
ABL1[a]	EAIYAAPFAKKK	0 %	ND[c]
AKT1[a]	KGSGSGRPRTSSFAEG	3.1%	ND[c]
ALK[a]	poly[Glu:Tyr] (4:1)	5.2%	ND[c]
CDK1/cyclin B[a]	Histone H1	4.9%	ND[c]
CDK2/cyclin A[a]	Histone H1	0 %	ND[c]
CDK3/cyclin E[a]	Histone H1	0 %	ND[c]
CDK4/cyclin D1[a]	RB-CTF	0 %	ND[c]
CDK5/p35[a]	Histone H1	0 %	ND[c]
CHK1[a]	KKKVSRSGLYRSPSMPENLNRPR	0 %	ND[c]
EGFR[a]	poly[Glu:Tyr] (4:1)	0%	ND[c]
ERBB2/HER2[a]	poly[Glu:Tyr] (4:1)	0 %	ND[c]
ERK1[a]	MBP (Myelin Basic Protein)	0 %	ND[c]
ERK2/MAPK1[a]	MBP (Myelin Basic Protein)	0 %	ND[c]
FGFR1[a]	KKKSPGEYVNIEFG	1.9%	ND[c]
FGFR2[a]	poly[Glu:Tyr] (4:1)	0%	ND[c]
FGFR3[a]	poly[Glu:Tyr] (4:1)	1.2%	ND[c]
FLT1/VEGFR1[a]	poly[Glu:Tyr] (4:1)	0%	ND[c]
FLT4/VEGFR3[a]	poly[Glu:Tyr] (4:1)	2.6%	ND[c]
FAK[b]	Ulight-poly[Glu:Tyr] (4:1)	89.8%	0.05 ± 0.01
FYN[a]	poly[Glu:Tyr] (4:1)	4.0%	ND[c]
GSK3β[a]	YRRAAVPPSPSLSRHSSPHQ(pS)EDEEE	0 %	ND[c]
IGF1R[a]	KKKSPGEYVNIEFG	8.2%	ND[c]
IR[a]	poly[Glu:Tyr] (4:1)	4.6%	ND[c]
JAK3[a]	GEEEEYFELVKKKK	0 %	ND[c]

KDR/VEGFR2[a]	poly[Glu:Tyr] (4:1)	0%	ND[c]
c-Kit[a]	poly[Glu:Tyr] (4:1)	0 %	ND[c]
LCK[a]	poly[Glu:Tyr] (4:1)	0 %	ND[c]
c-MET	KKKSPGEYVNIEFG	10.2%	ND[c]
PDGFRA[a]	poly[Glu:Tyr] (4:1)	0%	ND[c]
Pyk2[a]	poly[Glu:Tyr] (4:1)	4.1%	ND[c]
RAF1[a]	MEK1 (K97R)	0%	ND[c]
c-Src[a]	poly[Glu:Tyr] (4:1)	0%	ND[c]
mTOR/FRAP1[a]	4EBP1	0%	ND[c]

[a]Inhibition of different kinases with **19a** was determined using radioactive in vitro kinase assay (conducted by Reaction Biology Corp. Malvern, USA), [b]**19a** was evaluated for its ability to inhibit FAK kinase activity through a TR-FRET based kinase assay. [c]ND: Not Determined. The results represent means SEM of 3 independent experiments performed in triplicate.

Table S-2: Purity data of tested compounds (HPLC)

Compound	Time retention (min)	Purity (HPLC)[a]
11a	16.4	98.8
11b	14.6	99.4
13a	11.2	95.1
13b	9.7	95,1
12a	14.8	98.3
12b	12.2	96.6
14a	18.6	95.0
14b	16.5	98.2
15a	17.5	99.6
15b	15.8	99.8
16a	13.5	96.4
16b	12.0	99.4
17a	11.5	95.6
17b	9.7	99.8
18a	14.2	99.6
18b	12.6	95.8
19a	11.7	96.9
19b	10.2	96.0
20a	8.7	95.2
20b	7.1	98.1
21a	7.6	97,3
21b	5.3	99.8
26a	12.5	99.6
27a	10.7	96.2
30a	16.4	98.4

31a	14.2	95.4

[a]HPLC analyses were carried out on a Shimadzu LC-10AD system (Shimadzu, Kyoto, Japan) with a fixed wavelength UV detector at 214 nm. An Agilent Zorbax Extend C18 Column (Agilent Technologies, Palo Alto, CA, USA) (250 mm x 4.6 mm, 5 µM) was used with a flow rate of 1.0 mL/min for the sample. The injection volume was 30 µL. Mobile phase was consisted of H20 + 0.1% TFA (A) and CH_3CN/H_2O: 70:30 (v/v) + 0.1% TFA (B) with gradient elution: 0~45 min (95:5 – 10:90).

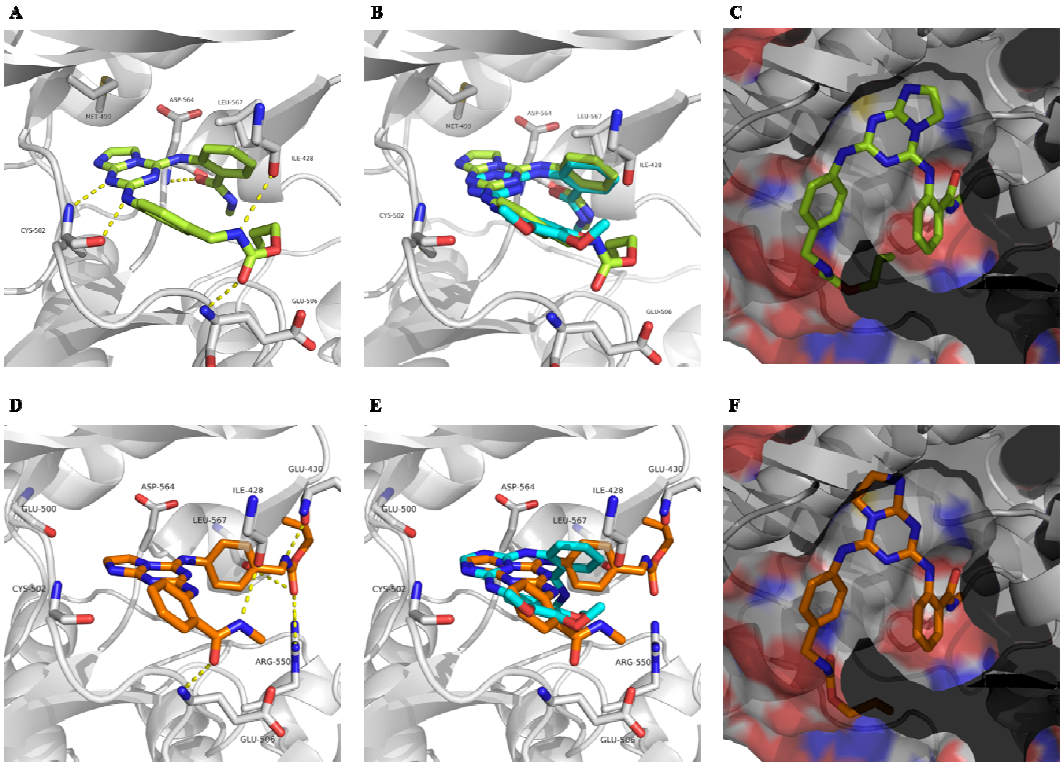


Figure S1. Docking of **19a-b** into the ATP binding pocket of FAK. A-F) The lowest energy binding modes obtained by molecular docking are shown for compound **19a** in light green (A-C) and **19b** in orange (D-E). Hydrogen bond interactions between receptor and ligands are shown as yellow dashes lines. B) Superposition of **19a** with the PHM16 ligand (cyan) indicates a very similar binding mode. C) A slice through the FAK kinase above the ligand-binding site with the protein surface shown semi-transparent indicates a good fit for **19a** into

the ATP binding pocket. D) **19b** makes no hydrogen bonds with the kinase hinge region. The hydrogen bonds formed are mostly solvent exposed. E) Superposition of **19a** and **19b** bound FAK structures. F) The **19b** compound is positioned in a hypothetical position that corresponds to the **19a** binding mode. Clashes of the imidazo ring in **19b** with Glu500 in the kinase hinge region prevent this binding mode for **19b**.

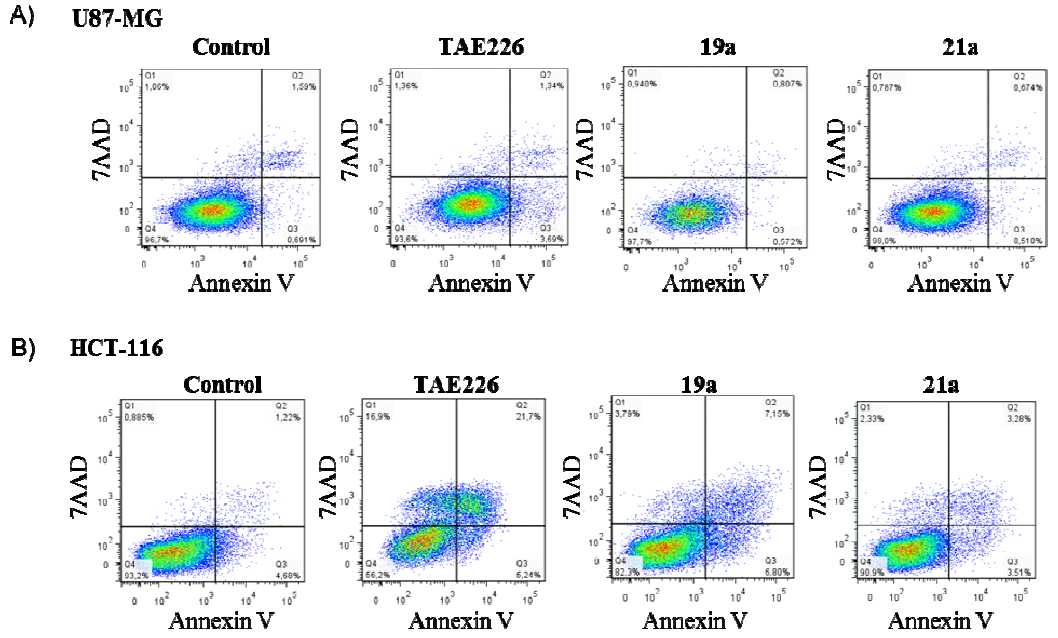


Figure S2: Compounds **19a**, **21a** and TAE-226 treatments significantly induce cell apoptosis in HCT-116 ($p < 0.001$): cells were treated for 48h with 0.1 μM **19a**, or 0.1 μM **21a** or 0.1 μM TAE-226, stained with Annexin V and 7-AAD, and analyzed by FACS calibur. Cells were then classified as alive with no apoptosis (i.e., 7-AAD negative, Annexin V negative), dead (i.e., 7-AAD positive), or apoptotic (i.e., Annexin V positive). One representative experiment is shown.

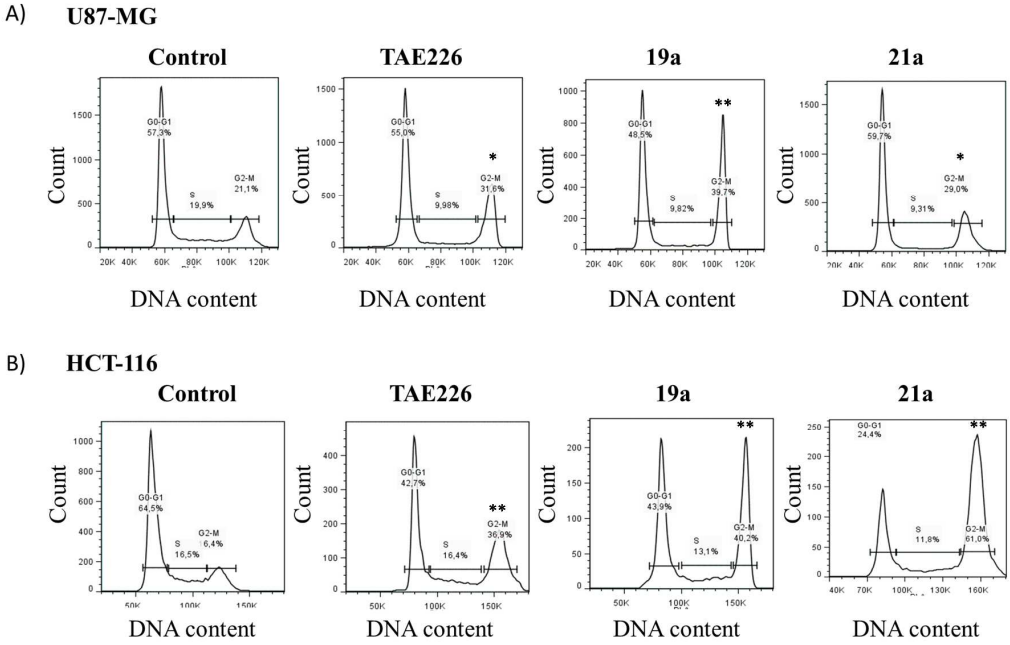


Figure S3: The cell cycle distribution was studied by flow-cytometry: cells were plated at $10x10^4$ per well in a 6-well culture dish, incubated overnight, and then exposed to 0.1 μM **19a** or 0.1 μM **21a** or 0.1 μM TAE-226 along with a vehicle treatment control for 48 h. Cell cycle distribution was represented by histograms. Asterisks denote statistically significant differences (*p < 0.01; **p < 0.001) as determined by unpaired Student's t test. One representative experiment is shown.

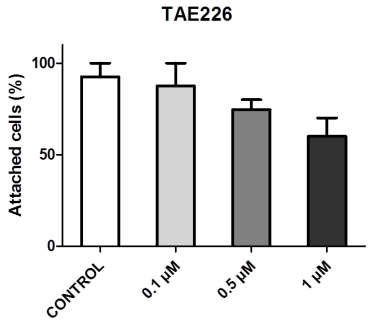


Figure S4: Effects of TAE-226 on the cell attachment in U-87MG cell lines.

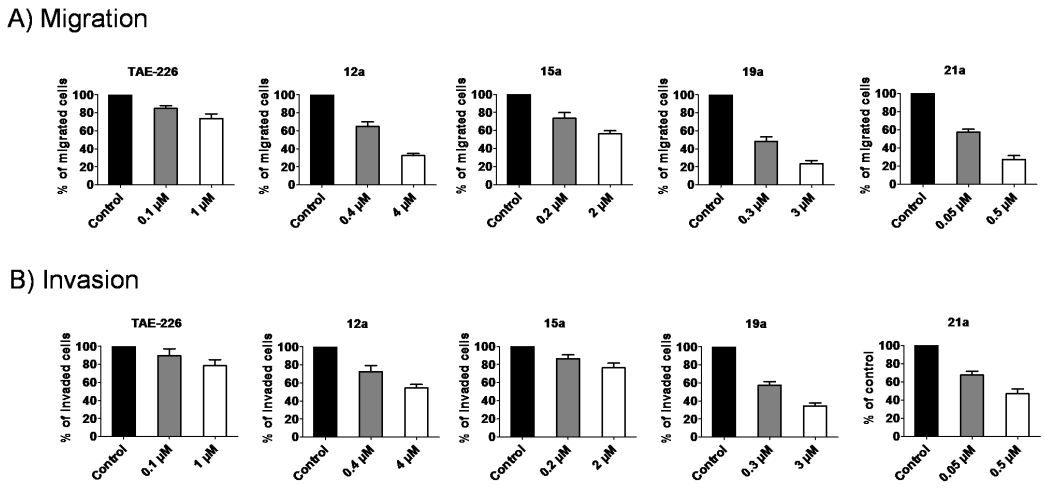


Figure S5: Effects of compounds **TAE-226, 12a**, **15a**, **19a** and **21a** on the migration/invasion in U-87MG cell lines, using boyden chamber assays. Cells were treated for 30 min in the presence of **12a**, or **15a**, or **19a**, or **21a** or TAE226. Cells were allowed to migrate/invade 24 hours. The migrated/invaded cells were fixed, stained, and five representative fields were photomicrographed for counting cells. (A) The average proportion of migrating cells (% of control) in five fields was represented by the histograms for the five compounds. (B) The average proportion of invading cells (% of control) in five fields was represented by the histograms for the five compounds.

Materials and Methods

Migration and invasion assays. Cells were serum straved for 24 h, trypsinized, resuspended to a concentration of 100,000 cells/mL in serum-free medium, and then treated for 30 min with DMSO and each of compounds (**12a**, **15a**, **19a**, **21a** and TAE-226). Next, 500 µL of the treated cells (50 000 cells) was added to the upper Transwell chambers coated with or without collagen 0.1%. Media containing 15% serum was added to the bottom well of the Transwell chambers. After 24, cells were fixed with paraformaldehyde 4% and stained with Toluidine Blue 0.1% in PFA 4%. After washing with PBS, the Optical Density was analyzed with a microplate reader (Bio-Rad) at 620 nm to determine cell migration/invasion.

7.3 Annex 3

Table VS – MD:

POCKET	Glide Score	MD Comment	average RMSD ± sd [Å] (final RMSD)	rank
42_2	**1**	**Firmly bound during 400ns**	1.933 ± 0.812 (2.353)	**1**
	2	Moves (not so slightly) in the pocket	2.795 ± 0.768 (2.400)	**3**
	3	Slight Movements and slight changes of orientation	2.002 ± 0.654 (2.913)	**2**
	319	Moves not so slightly in the pocket	4.669 ± 3.372 (8.759)	4
	465	Undocks	3.132 ± 3.807 (17.066)	-
01_1	**14**	Moves and changes orientation	4.014 ± 1.770 (5.312)	-
	1	Undocks	5.018 ± 2.597(9.292)	-
	12	Minor reorientation	2.695 ± 0.969(3.221)	**6**
	29	**Firmly bound**	3.400 ± 0.758(2.288)	**5**
	83	Minor reorientation	3.703 ± 1.303(3.109)	**16**
	182	Changes orientation slightly, but still docked	3.287 ± 1.344(4.025)	**17**
	295	Minor reorientation	1.308 ± 0.416(0.898)	**7**
31_1	**23**	**Firmly bound**	1.731 ± 0.467 (1.329)	**8**
	25	Undocks	3.044 ± 1.894 (8.820)	-
	40	Changes orientation, but remains docked	5.623 ± 1.897 (7.373)	**9**
03_1	**5**	Can't be minimized!!	---	-
	13	changes orientation slightly	2.138 ± 0.827(1.804)	**15**
	14	moves slightly	1.851 ± 0.796(2.740)	**14**
	17	**Firmly bound**	1.756 ± 0.449(2.220)	**10**
	40	moves slightly, but still docked (270 ns)		**18**
	126	Changes orientation, but remains docked	2.243 ± 0.935(3.486)	**11**
14_1	**3**	**Firmly Bound**	2.038 ± 0.365(1.557)	**12**
	7	**Firmly Bound**	2.490 ± 0.609(1.73)	**13**
	8	Undocks	24.436 ± 12.674(28.860)	-
	12	Undocks	9.176 ± 15.595(17.137)	-
	22	Undocks	10.983 ± 16.895(48.807)	-
	36	Undocks	40.435 ± 23.151(101.725)	-

This Table shows from the left to the right: the pockets selected for VS; the molecules selected after Glide VS; molecules behaviour after MD simulations; the molecules' RMSD average and its standard deviation; the ranking for further validation where 1 is the best candidate.

7.4 Annex 4

Other publications:

1. Davide Provasi, **Marta Camacho-Artacho**, Ana Negri, Juan Carlos Mobarec, Marta Filizola.
   Ligand-Induced Modulation of the Free-Energy Landscape of G Protein-Coupled Receptors Explored by Adaptive Biasing Techniques.
   PLOS Computational Biology. Oct 2011. 7(10): e1002193.
   DOI: 10.1371/journal.pcbi.1002193

2. **Marta Camacho-Artacho**, Valentina Abet, Luis M. Frutos, Federico Gago, Julio Alvarez-Builla and Carolina Burgos.
   Unusual Approach to 3-Aryl-2-aminopyridines through a Radical Mechanism: Synthesis and Theoretical Rationale from Quantum Mechanical Calculations.
   The Journal of Organic Chemistry. Jan 2011. *76* (5), pp 1452–1455.
   DOI: 10.1021/jo102122h

3. George Khelashvili, Kevin Dorff, Jufang Shan, **Marta Camacho-Artacho**, Lucy Skrabanek , Bas Vroling, Michel Bouvier, Lakshmi A. Devi , Susan R. George, Jonathan A. Javitch, Martin J. Lohse, Graeme Milligan, Richard R. Neubig, Krzysztof Palczewski, Marc Parmentier, Jean-Philippe Pin, Gerrit Vriend, Fabien Campagne and Marta Filizola.
   GPCR-OKB: The G Protein Coupled Receptor Oligomer Knowledege Base. *Bioinformatics.* Jul 2010. 26(14), pp 1804–1805.
   DOI: 10.1093/bioinformatics/btq264

“Hagas lo que hagas ámalo, como amabas la cabina del Paradiso cuando eras niño.”
(Cinema Paradiso, 1988)