



Universitat de Barcelona

DISRUPTING THE PROTEIN-PROTEIN RECOGNITION IN CANCER PATHWAYS BY MOLECULAR MODELING

Cristian Obiol Pardo



Departament de Química Física

Institut de Química Teòrica i Computacional

Universitat de Barcelona
Departament de Química Física
Facultat de Química
Institut de Química Teòrica i Computacional

DISRUPTING THE PROTEIN-PROTEIN
RECOGNITION IN CANCER PATHWAYS BY
MOLECULAR MODELING

Doctorado en Química Teórica y Computacional

Memoria presentada por Cristian Obiol Pardo para optar al título de Doctor por la
Universidad de Barcelona

Director: Dr. Jaime Rubio Martínez
Dept. Química Física, Facultat de Química
Universitat de Barcelona

El misterio es la cosa más bonita que podemos experimentar.
Es la fuente de todo arte y ciencia verdaderos.

A. Einstein.

Agradecimientos

A mi director de tesis. A Jaime Rubio por aceptarme desde el primer momento en su grupo, por su apoyo desde el principio cuando me topé con linux y con la dinámica de proteínas. Y por ayudarme siempre que me he quedado atascado en algún punto de mi tesis. Por eso y por más...gracias!.

A todos los compañeros de departamento y de despacho que he ido conociendo con los años. Primero a Oscar Villacañas por enseñarme muchos pequeños trucos para entender los programas y porque ha sido un referente a seguir. A Alex Rodríguez que recuerdo que andaba por allí en mis primeros días, A Emilianita d'Oria y a Miquel Lluell que también compartimos despacho y vivencias. Como no, a Iñigo García porque congeniamos desde el principio y hemos sido amigos hasta el final (como pasan los años). En aquellos días, también me pasaba por mi 'segundo' despacho...A Francesc Viñes por TODO lo que hemos compartido juntos, a Javi Carrasco por ser tan divertido (sin pretenderlo?), a Dani Torres porque te admiro como amigo y como investigador...y por el dibujo que ilustra esta sección (es esto el final ?.no es el principio). Y con los cambios de despacho, agradecer a Anabel Álvarez, a Norma Merchán, a Abel Carreras...por todas las horas que hemos compartido. A Axel por las discusiones físico-matemáticas. A Laura Delgado por revitalizar el grupo, por las muchas cosas que me ha enseñado, como el Glide y como hacer más bonitas las presentaciones. Y porque no me he podido aburrir desde que ella llegó.



A los chicos y chicas de la UPC: Patricia Gómez, Arnau Cordoní, Josep Cantó, Marta Pinto, Lourdes Roset, Francesc Corcho y J.J. Pérez. A Marta por la energía positiva que desprende y porque he sobrevivido a sus achuchones. A Patri por ser como es y por tener buen gusto para la música!.

A mis compañeros de licenciatura, Carlos, Charo, Ari, Susana...por hacer los descansos agradables, a Santi porque siempre es entretenido hablar con él, a Gema Alcarraz por la cantidad de pruebas experimentales que nos ha realizado, y porque ha dado realidad a mis resultados. A Roger por ayudarme en algunos momentos difíciles. A Oscar Blanco por todas las charlas que hemos tenido. A Isabel Solé especialmente, por poder compartir muchos momentos con ella, buenos y malos.

También quiero agradecer a los grupos experimentales que han colaborado con nosotros. Al grupo de Marta Cascante de Bioquímica y Biología Molecular UB, a Dolors Colomer y Roberto Alonso del departamento de hematopatología del Hospital Clínic de Barcelona.

A un profesor que recuerdo especialmente, a J.C. Paniagua porque no pude tener mejor profesor de QFIII y así me decidí por la química teórica.

Por último, quiero agradecer a mi familia porque ellos son parte de los que soy.

Aims and Distribution of the Thesis

Cancer is the second disease leading cause of death in industrialized countries. Although early detection and more efficient drugs are responsible of the reduction of mortality, several cancers still present difficult treatments and low survival rates. Conventional drugs only exhibit moderate therapeutic index between cancer and normal tissues but recent advances are focused to improve less-toxic treatments. Hence, new drugs must target specific signaling pathways involved in cell growth and proliferation. Concerning this aim, two mechanism involved in cancer disease, named **apoptosis** (or programmed cell death) and **pentose phosphate pathway**, have been selected in this work to search new inhibitors to target crucial proteins of both cell routes.

Overexpression of antiapoptotic genes has been correlated with tumor growth and resistance to chemotherapy, thus many efforts have been done to block the activity of **XIAP** and **Survivin**, central proteins acting in apoptosis and studied in the present work. Moreover, the two most active proteins detected in both the oxidative and non-oxidative branches of the pentose phosphate pathway, **Glucose-6-Phosphate Dehydrogenase (G6PDH)** and **Transketolase (TKT)**, have been also selected in this thesis.

Molecular Modeling methods, covering topics in protein and peptide recognition, molecular dynamics, pharmacophore generation, database searching, docking and scoring in virtual screening and binding free energy prediction, have been applied with success to discover new active molecules inhibitors of XIAP, Survivin, G6PDH and TKT proteins.

After a brief introduction of the theoretical methods applied in this work, described in Chapter 1, Chapters 2, 3 and 4 are used to collect the results obtained studying the XIAP and Survivin proteins. The Transketolase protein is discussed in Chapter 5 while Glucose-6-Phosphate Dehydrogenase is studied in Chapters 6 and 7. Finally, the general conclusions are summarized in Chapter 8.

A summary of this work written in Spanish can be seen in the last part of this manuscript.

INDEX

◦ CHAPTER I: Computational Methods, Biomolecular Simulation and Drug Design

1. BRIEF INTRODUCTION.....	2
2. CONTEXT.....	2
2.1. QUANTUM MECHANICS.....	2
2.1.1. Polielectronic Wavefunctions and Slater Determinants.....	4
2.1.2. Hartree-Fock Method.....	6
2.1.3. Restricted Closed-Shell Hartree-Fock.....	7
2.1.4. Rootham Equation and Base functions.....	8
2.1.5. AM1 Method.....	10
2.2. MOLECULAR MECHANICS AND FORCE FIELDS.....	11
2.2.1. Parametrization.....	17
2.2.2. Computational Considerations.....	17
2.2.3. Optimization Methods.....	18
2.2.4. Molecular Dynamics.....	19
2.2.5. Periodic Boundary Conditions.....	22
2.2.6. AMBER Program.....	23
2.2.7. AMBER Force Field and Parametrization.....	24
2.3. INTRODUCTION TO DRUG DESIGN.....	26
2.3.1. The Pharmacophore.....	27
2.3.2. Drug Design, Pharmacokinetics and Lipinsky Rules.....	28
2.3.3. Structure-based methods.....	34
2.3.4. 3D Database Searching.....	34
2.3.5. Docking.....	35
2.3.6. Scoring.....	36
2.3.7. MMPBSA Methodology.....	39
3. REFERENCES.....	43

INDEX

◦ **CHAPTER II: Protein-Protein Recognition as a first step towards the inhibition of XIAP and Survivin anti-apoptotic proteins**

1. BRIEF INTRODUCTION.....	48
2. CONTEXT.....	49
3. METHODS.....	52
3.1. Construction of Smac/DIABLO-XIAP complex.....	52
3.2. Construction of Smac/DIABLO-Survivin complex.....	53
3.3. Minimization.....	54
3.4. Molecular Dynamics Simulation.....	55
3.5. Free energy estimations using the MMGBSA approach.....	55
3.6. Docking and Post-Docking Methodology.....	56
4. RESULTS AND DISCUSSION.....	57
4.1. Convergence analysis.....	57
4.2. Zinc-ligand analysis.....	59
4.3. Hydrogen Bond analysis.....	60
4.4. Van der Waals interaction analysis.....	61
4.5. Electrostatic interaction analysis.....	63
4.6. Binding free energy analysis.....	65
4.7. Pharmacophore analysis.....	66
4.8. Docking and MD simulations of embelin and embelin derivatives.....	69
4.9. Binding free energy analysis for embelin and embelin derivatives complexed with XIAP.....	78
5. CONCLUSIONS.....	82
6. REFERENCES.....	83

INDEX

◦ **CHAPTER III:** Pharmacophore Exploitation of Smac/DIABLO complexed with XIAP and Survivin

1. BRIEF INTRODUCTION.....	90
2. CONTEXT.....	90
3. RESULTS.....	93
3.1. Pharmacophore and 3D Searching.....	93
3.2. Docking and Scoring Protocols.....	95
3.3. Purchased Compounds.....	96
3.4. Experimental Results.....	98
4. CONCLUSIONS.....	99
5. REFERENCES.....	100

◦ **CHAPTER IV:** Comparative Evaluation of MMPBSA and XSCORE to Compute Binding Free energy in XIAP-peptide Complexes

1. BRIEF INTRODUCTION.....	104
2. CONTEXT.....	104
3. METHODS.....	107
3.1. Construction of the XIAP-Smac/DIABLO (9 residues) Complex.....	107
3.2. Construction of the XIAP-Peptides (4 residues) Complexes.....	108
3.3. Minimization and Molecular Dynamics.....	108
3.4. XSCORE.....	109
3.5. MMPBSA.....	110
4. RESULTS.....	112
5. CONCLUSIONS.....	124
6. REFERENCES.....	126

INDEX

◦ **CHAPTER V:** Searching for a human Transketolase Inhibitor

1. BRIEF INTRODUCTION.....	131
2. CONTEXT.....	131
3. METHODS.....	134
4. RESULTS.....	137
4.1. Analysis of Interactions.....	137
4.2. Pharmacophoric Hypothesis.....	139
4.3. 3D Database Searching, Docking and Scoring.....	142
4.4. Experimental Results.....	144
4.4.1. Second Generation of Human Transketolase Inhibitors.....	145
5. CONCLUSIONS.....	148
6. REFERENCES.....	149

◦ **CHAPTER VI:** Exploring the Dimerization Interface of Glucose-6-Phosphate Dehydrogenase by Molecular Dynamics: Searching for Interface Peptide Inhibitors

1. BRIEF INTRODUCTION.....	154
2. CONTEXT.....	154
3. METHODS.....	157
3.1. Construction of the G6PDH Dimer Complex.....	157
3.2. Minimization and Molecular Dynamics of the G6PDH Dimer.....	158
4. RESULTS AND DISCUSSION.....	159
4.1. Hydrogen Bond Analysis.....	159
4.2. Electrostatic Analysis.....	162
4.3. Van der Waals Interaction Analysis.....	163
4.4. Construction of the Peptide-G6PDH Complexes.....	166
4.5. Construction of the Cyclic Peptide-G6PDH Complex.....	166
4.6. Interaction Analysis of the Peptide-G6PDH Complexes.....	168
4.7. Calculation of Binding Free energy for the Peptide-G6PDH Complexes.....	177

INDEX

5. EXPERIMENTAL RESULTS.....	179
6. FURTHER MODIFICATIONS OF THE CYCLIC PEPTIDE.....	180
7. CONCLUSIONS.....	183
8. REFERENCES.....	185

◦ **CHAPTER VII:** Pharmacophore Exploitation of Glucose-6-Phosphate Dehydrogenase: Searching for Non Peptidic Inhibitors

1. BRIEF INTRODUCTION.....	189
2. CONTEXT.....	189
3. RESULTS.....	190
3.1. Pharmacophore and 3D Searching.....	190
3.2. Docking and Scoring Protocols. Purchased Compounds.....	193
4. EXPERIMENTAL RESULTS.....	196
5. CONCLUSIONS.....	197
6. REFERENCES.....	198

◦ **CHAPTER VIII:** Final Conclusions.....199

◦ **APPENDIX:** Amino Acid Structures and Nomenclature.....202

◦ Ruptura del Reconocimiento Proteína-Proteína en Rutas Tumorales mediante Modelización Molecular.....203

1. Interés del Proyecto y Objetivos.....	204
2. Métodos.....	205
3. Estudio de las Proteínas Antiapoptóticas XIAP y Survivin.....	208
4. Estudio de la Proteína Transketolasa.....	213
5. Estudio de la Proteína Glucosa-6-Fosfato Deshidrogenasa.....	217
6. Conclusiones Generales.....	223
7. Bibliografía.....	225

CHAPTER I: Computational Methods,
Biomolecular Simulation and Drug Design



1. BRIEF INTRODUCTION

The relationship between the molecular structure of a biological system and its physicochemical properties, is the fundamental problem involved in this work.

Due to the fast development of computational resources, it is now possible to simulate a biological system at atomic level, thus the modern computational methods can be used to design new active molecules, exhibiting a specific interaction with proteins and nucleic acids.

The procedure to study a biomolecule is to find a mathematical function that describes its potential energy. Quantum Mechanics is the most correct methodology that can be used for this purpose, nevertheless, when large and complex systems are studied, Molecular Mechanics must be applied due to the high computational cost of ab-initio methods.

The Molecular Mechanics methodology is based on classical potentials and it has been widely applied with success since 1976. Although it does not describe with physical severity the molecular behaviour (especially the electronic structure), it can be used with remarkable results for biomolecular systems.

2. CONTEXT

2.1. QUANTUM MECHANICS

In this section, we will describe briefly the quantum chemistry methodology applied in the present work. Although this work is based mainly on the treatment of biological systems, at Molecular Mechanics level of theory, we also carried out ab-initio calculations in order to study simple molecules, such as organic ligands, known protein inhibitors and enzyme cofactors.

The Schrödinger equation is postulated as the basis of Quantum Mechanics:

$$H \Psi = E \Psi \quad (1)$$

Where H is the hamiltonian operator, E is the energy of the system and Ψ is the wavefunction, postulated as a complex function of Hilbert's space (a generalized Euclidean space) which contains all

the information about the system.

Taken a system of N nucleus and n electrons, the Schrödinger equation can be expressed as following:

$$H \Psi (A,i) = E \Psi (A,i) \quad (2)$$

Being $\Psi (A,i) = \Psi (W_1, W_2, \dots, W_i, \dots, W_N, w_1, w_2, \dots, w_i, \dots, w_n)$, where $A = W_1, W_2, \dots, W_i, \dots, W_N$ denotes for the spatial and spin coordinates of N nucleus and $i = w_1, w_2, \dots, w_i, \dots, w_n$ denotes for spatial and spin coordinates of n electrons. The non-relativistic hamiltonian operator of a polielectronic molecule is shown in eq. 3 (in atomic units):

$$\begin{aligned} H = & \sum \frac{-1}{2m_A} \nabla_A^2 + \sum \frac{-1}{2} \nabla_i^2 + \sum \sum \frac{Z_A Z_B}{R_{AB}} + \\ & + \sum \sum \frac{I}{R_{ij}} + \sum \sum \frac{-Z_A}{R_{iA}} = T_{\text{nuc}} + T_{\text{el}} + V_{\text{nuc}} + V_{\text{el}} + V_{\text{nuc-el}} \end{aligned} \quad (3)$$

Where T accounts for the kinetic energy operator and V denotes for the potential energy operator.

We can define the electronic hamiltonian, that includes the terms depending only on electron coordinates taken a fixed nuclear structure:

$$\begin{aligned} H_{\text{el}} = & \sum \frac{-1}{2} \nabla_i^2 + \sum \sum \frac{I}{R_{ij}} + \\ & + \sum \sum \frac{-Z_A}{R_{iA}} = T_{\text{el}} + V_{\text{el}} + V_{\text{nuc-el}} \end{aligned} \quad (4)$$

If we fix a nuclear geometry A, then H_{el} only depends on electronic coordinates and we could solve the electronic Schrödinger equation:

$$H_{\text{el}, A} \Psi_{\text{el}, A}(\mathbf{i}) = E_{\text{el}, A} \Psi_{\text{el}, A}(\mathbf{i}) \quad (5)$$

To solve the whole problem it will be very useful to describe the total wavefunction as a nuclear and

electronic wavefunctions product:

$$\Psi(A, i) = \Psi_{\text{nuc}}(A) \Psi_{\text{el}, A}(i) \quad (6)$$

Introducing eq.6 into eq. 2 and performing the following simplification [1]:

$$T_{\text{nuc}} [\Psi_{\text{nuc}}(A) \Psi_{\text{el}, A}(i)] \cong \Psi_{\text{el}, A}(i) T_{\text{nuc}} \Psi_{\text{nuc}}(A) \quad (7)$$

Which is possible because the electronic wavefunction changes very slow with nuclear coordinates. Thus we can obtain the nuclear Schrödinger equation:

$$H_{\text{nuc}} \Psi_{\text{nuc}}(A) = E_{\text{nuc}} \Psi_{\text{nuc}}(A) \quad (8)$$

With $H_{\text{nuc}} = T_{\text{nuc}} + V_{\text{nuc}} + E_{\text{el}, A}$. This simplification was introduced by M. Born and J. R. Oppenheimer to solve in two steps the Schrödinger equation, it is of excellent quality in most common chemical problems. Therefore, we can solve separately the electronic Schrödinger equation (eq. 5) and the nuclear one (eq. 8) which mathematically is an eigenvector-eigenvalue problem. The computational cost is drastically reduced using the Born-Oppenheimer approximation and it does not affect at the calculated chemical properties.

2.1.1. POLIELECTRONIC WAVEFUNCTIONS AND SLATER DETERMINANTS

If we think about a 2 electron system, and we choose a full Hilbert space of base functions $\{\psi_1(w_1), \psi_1(w_2), \psi_1(w_1)\psi_2(w_2), \dots, \psi_i(w_1)\psi_j(w_2), \dots\}$, using the superimposition theorem [1] we can describe the wavefunction of the system as a linear combination of these known base functions:

$$\Psi(w_1, w_2) = \sum \sum a_{ij} \psi_i(w_1) \psi_j(w_2) \quad (9)$$

Where the summations are extended to infinity, due to it is the Hilbert's space dimension.

Taking into account also the Pauli's antisymmetric postulate [1]:

$$\Psi (w_1, w_2) = - \Psi (w_2, w_1) \quad (10)$$

We can develop eq. 9 to:

$$\Psi (w_1, w_2) = \sum \sum a_{ij} [\psi_i(w_1)\psi_j(w_2) - \psi_j(w_1)\psi_i(w_2)] \quad (11)$$

And this can be expressed in a determinant form, which is called the Slater determinant:

$$\Psi (w_1, w_2) = \sum \sum a_{ij} | \psi_i(w_1)\psi_j(w_2) | \quad (12)$$

It is easy to realize that if the base functions are orthonormalized, then the electronic wavefunction can be described as following:

$$\Psi (w_1, w_2) = \sum \sum C_{ij} \frac{1}{\sqrt{2}} | \psi_i(w_1)\psi_j(w_2) | \quad (13)$$

$$C_{ij} = \sqrt{2} a_{ij}$$

For a polielectronic system, we can obtain an n-electronic wavefunction simply generalizing last equation:

$$\Psi (w_1, w_2, \dots, w_n) = \sum \sum \dots \sum C_{ij, \dots, l} \frac{1}{\sqrt{n!}} | \psi_i(w_1)\psi_j(w_2) \dots \psi_l(w_n) | \quad (14)$$

Which is the general expression for the electronic wavefunction of an n-electron system and the basic equation for the resolution of chemical quantum problems, nevertheless its difficult resolution, requires several simplifications. One of them will be presented in the next section.

2.1.2. HARTREE-FOCK METHOD

The most drastic approximation in order to solve eq. 14 is to reduce the summations to only one term (only one Slater determinant), and to forget the physical interpretation of the coefficients C_{ij} . This simplification is called the Hartree-Fock method (HF):

$$\phi(w_1, w_2, \dots, w_n) = \frac{1}{\sqrt{n!}} |\psi_i(w_1)\psi_j(w_2)\dots\psi_n(w_n)| \quad (15)$$

In order to obtain the best wavefunction ϕ is very important to find the best ensemble of spinorbitals $\{\psi_i(w_1)\psi_j(w_2)\dots\psi_n(w_n)\}$. For this purpose it is introduced the minimization of the variational integral:

$$W = \langle \phi | H \phi \rangle = \int \dots \int \phi^* H \phi \, dw_1 \dots dw_n \quad (16)$$

Here, it is used the Dirac notation, $\langle \phi | H \phi \rangle$, which denotes for a scalar product of functions in the Hilbert's space and ϕ^* denotes for the complex conjugated function.

Following the variational principle, the energy and wavefunction of the system will be:

$$E^{\text{HF}} = W^{\text{op}} = \langle \phi^{\text{HF}} | H \phi^{\text{HF}} \rangle \geq E^{\text{exp}} \quad (17)$$

$$\phi^{\text{HF}}(w_1, w_2, \dots, w_n) = \frac{1}{\sqrt{n!}} |\psi_i^{\text{HF}}(w_1)\psi_j^{\text{HF}}(w_2)\dots\psi_n^{\text{HF}}(w_n)| \quad (18)$$

It can be demonstrated, using Lagrange multipliers [2], that the best ensemble of base wavefunctions can be found solving the following expression:

$$f(w) \psi_i^{\text{HF}}(w_i) = \varepsilon_i \psi_i^{\text{HF}}(w_i) \quad i = 1, 2, \dots, n \quad (19)$$

Where $f(w) = h(r) + v_{re}(w)$ and $h(r) = -\nabla^2/2 + \sum -Z_A/r_A$. $f(w)$ is the Fock's operator that does not represent any observable of the system, and v_{re} is a potential energy operator, which can be described in the following form:

$$v_{re}(w) = \sum j_1(w) - k_1(w) \quad (20)$$

Being $j_1(w)$ and $k_1(w)$ the Coulomb and the exchange operators respectively. The former is a consequence of the Coulomb repulsion between electrons, while the latter is a consequence of the antisymmetric behaviour of the wavefunction, without classical interpretation. These two new operators are defined as bielectronic integrals in eq. 21 and 22:

$$\begin{aligned} \langle \psi(w_1) | j_1 \psi(w_1) \rangle &= \int_{w_1} \psi^*(w_1) \left\{ \int [\psi_1^{HF}(w_2)]^* 1/r_{12} \psi_1^{HF}(w_2) dw_2 \right\} \psi(w_1) dw_1 = \\ &= \langle \psi(w_1) \psi_1^{HF}(w_2) | 1/r_{12} \psi(w_1) \psi_1^{HF}(w_2) \rangle \end{aligned} \quad (21)$$

$$\langle \psi(w_1) | k_1 \psi(w_1) \rangle = \langle \psi(w_1) \psi_1^{HF}(w_2) | 1/r_{12} \psi_1^{HF}(w_1) \psi(w_2) \rangle \quad (22)$$

In this stage, the resolution of the Hartree-Fock equations is performed by an iterative procedure which is called a self consistent field (SCF).

2.1.3. RESTRICTED CLOSED-SHELL HARTREE FOCK

Restricted closed-shell Hartree Fock (RCHF) is the ab-initio methodology most applied in this work. It is useful for molecules in equilibrium geometries with close-shell electronic structure. We can describe now the electronic wavefunction as following:

$$\Phi(w_1, w_2, \dots, w_n) = \frac{1}{\sqrt{n!}} | \phi_1 \alpha(w_1) \phi_1 \beta(w_2) \dots \phi_{n/2} \alpha(w_{n-1}) \phi_{n/2} \beta(w_n) | \quad (23)$$

Here, spatial and spin coordinates are shown separately, α denotes for the quantum state “up” of the

8

electron spin while β denotes for the quantum state “down”.

On the other hand, if we integrate only the spin coordinates in eq. 19 we can achieve the next equation:

$$f_{\phi}(r) \phi_i(r_i) = \epsilon_i \phi_i(r_i) \quad (24)$$

with the Fock's operator defined as:

$$f_{\phi}(r) = \int_w g^*(w) f(r,w) g(w) dw \quad (25)$$

Where $g(w) = \alpha(w)$ or $\beta(w)$.

Resolution of the integral for the Fock's operator is also shown:

$$f_{\phi}(r) = h(r) + v_{\phi}(r) \quad (26)$$

Being $v_{\phi}(r) = \sum_j j_1(r) - k_1(r)$, where the summation is extended to only $n/2$ electrons and being $j_1(r)$, $k_1(r)$ the spatial operators, analogous to $j_1(w)$, $k_1(w)$ of equation 20.

Now, the properties of the α and β functions simplify the total resolution of bielectronic integrals and the resolution of the RCHF energy [1] to:

$$E^{RCHF} = \sum_i \langle \phi_i | h | \phi_i \rangle + \sum_i \sum_j [2 \langle \phi_i(r_1) \phi_j(r_2) | 1/r_{12} | \phi_i(r_1) \phi_j(r_2) \rangle - \langle \phi_i(r_1) \phi_j(r_2) | 1/r_{12} | \phi_j(r_1) \phi_i(r_2) \rangle] \quad (27)$$

Where the summations are extended to only a number of electrons of $n/2$.

2.1.4. ROOTHAM EQUATION AND BASE FUNCTIONS

The simplest form to solve the RCHF equation is to express the HF orbitals as linear combinations of known functions $\{ \chi_1(r) \dots \chi_m(r) \}$:

$$\phi_i(r) = \sum_s c_{si} \chi_s(r) \quad \begin{matrix} i = 1, \dots, n/2 \\ s = 1, \dots, n/2 \end{matrix} \quad (28)$$

And to optimize the coefficients by applying the variational principle, solving the following eigenvector-eigenvalue problem (best described as a matrix system):

$$\mathbf{f} \mathbf{c}_i = \varepsilon_i \mathbf{S} \mathbf{c}_i \quad (29)$$

Where the bold letters are referred as matrix (or vector) and \mathbf{S} is the overlap matrix, or combination of known base functions:

$$\begin{aligned} S_{st} &= \langle \chi_s | \chi_t \rangle \quad s, t = 1, \dots, m \\ f_{st} &= \langle \chi_s | f_\phi | \chi_t \rangle \end{aligned} \quad (30)$$

We can develop each element of the Fock's matrix as:

$$\begin{aligned} f_{st} = h_{st} + \sum_u \sum_v c_{ui}^* c_{vj} [& 2 \langle \chi_s(r_1) \chi_u(r_2) | 1/r_{12} \chi_t(r_1) \chi_v(r_2) \rangle - \\ & \langle \chi_s(r_1) \chi_u(r_2) | 1/r_{12} \chi_v(r_1) \chi_t(r_2) \rangle] \end{aligned} \quad (31)$$

Which is called the Rootham equation.

Up to now, nothing is explained about the best base function set that can be selected. Which are the best functions to solve the variational principle, with low computational cost and quantitative ab-initio results?. For mathematical properties reasons, gaussian functions are the answer, they have the following general expression:

$$g_{ijk} = N x_b^i y_b^j z_b^k \exp[-\alpha r_b^2] \quad (32)$$

Being r_b the distance electron-nucleus, i-j-k integer identifiers and N the normalization constant:

$$N = (2\alpha / \pi)^{3/4} [(8\alpha)^{i+j+k} \frac{i!}{2i!} \frac{j!}{2j!} \frac{k!}{2k!}]^{1/2} \quad (33)$$

If $i+j+k = 0$ we have a s-type orbital, if $i+j+k = 1$ we have three p-type orbitals and if $i+j+k = 2$ we have 6 d-type orbitals (but only 5 linearly independent).

Moreover, to increase flexibility to reproduce the wavefunction, we can use a linear combination of gaussian functions:

$$\chi_s = \sum d_{us} g_u \quad (34)$$

χ_s is the contracted gaussian-type function (CGTF), d_{us} are the contraction coefficients and g_u are called primitive gaussian functions.

The most applied base function in this work is the so-called 6-31G basis set, which represents each inner-shell orbital with one contracted gaussian function that is a linear combination of 6 primitive gaussian functions, and each valence orbital with two basis functions, one contracted that is a linear combination of 3 primitives and 1 primitive gaussian function.

Basis sets such as 6-31G* are said to be polarized basis sets.

2.1.5. AM1 METHOD

AM1 or Austin Model 1 [3] was published in 1985 as a semiempirical method for electronic structure which could improve the MNDO method by reducing the repulsion of atoms at close distances.

Like other semiempirical methods, AM1 introduces some simplifications in order to reduce the computational cost:

- A simple Fock's operator.
- The addition of experimental parameters.
- The neglect of differential diatomic overlap (NDDO).
- The treatment of valence electrons only.

For the AM1 method, the hamiltonian operator takes the form :

$$H_{\text{val}} = \sum [-1/2 \nabla_i^2 + V(i)] + \sum \sum 1/r_{ij} = \sum H_{\text{val,core}}(i) + \sum \sum 1/r_{ij} \quad (35)$$

Where it is only extended to the number of valence electrons and where $V(i)$ is the potential energy of

each valence electron in the field of nuclei and inner-shell electrons.

This method employs Slater type orbitals (instead of gaussian type orbitals) that are shown in a general expression in eq. 36:

$$f_{\text{STO}} = [(2 \xi / a_0)^{n+1/2} / (2n)!^{1/2}] r_a^{n-1} \exp [- (\xi r_a) / a_0] Y_l^m(\sigma, \varphi) \quad (36)$$

Where ξ is the charge parameter, r_a is the distance nucleus-electron and $Y_l^m(\sigma, \varphi)$ are the spherical harmonics. Moreover, it is introduced the following simplifying approximation:

$$\langle f_z f_y | f_m f_n \rangle = \iint [f_z(r_1) * f_y(r_1) f_m(r_2) f_n(r_2)] / r_{12} dv_1 dv_2 = \delta_{zy} \delta_{mn} \langle f_z f_y | f_m f_n \rangle \quad (37)$$

Being $\langle f_z f_y | f_m f_n \rangle$ the bielectronic integrals and δ the Kronecker delta.

2.2. MOLECULAR MECHANICS AND FORCE FIELDS

Molecular Mechanics studies molecules by applying a simple potential function and if it is used correctly can achieve accurate results. This methodology is widely used in several biochemical and biophysical problems, such as conformational analysis of proteins and nucleic acids, ligand-receptor interactions, peptide folding, drug design, etc.

Force field is the name of the potential function when we use Molecular Mechanics, and it is usually based on the following terms:

Valence terms, associated with the bond and angle movements.

Non-valence terms, associated with non-bonded interactions, like van der Waals interactions and electrostatic interactions.

Cross terms, related to the coupling of valence terms.

Moreover, each force field needs for a great number of constants, which are called parameters.

A Force field can be classified depending of three characteristics, firstly, the mathematical form of each term, secondly the number of cross terms included and thirdly the information used to develop the parameters.

If we choose the first and second criteria we can classify the force field as :

Class I fields: without cross terms and using harmonic functions. For instance, we can cite AMBER [4] and CHARMM [5] as typical examples.

Class II fields: they contain more complex functions and cross terms, such as the MM3 [6] force field.

Class III fields: they add to a class II field some properties, such as the polarizability.

If we choose the third characteristic, a force field can be called as:

Experimental field: when the majority of the parameters are extracted from experimental data (IR spectroscopy, Raman spectroscopy, NMR, thermochemical data ...).

Quantum-mechanical field: when the parameters come from ab-initio or density functional methods. For instance the QMFF [7] can be called as a quantum-mechanical force field.

Equation (38) shows a general expression of a force field, E_{str} denotes for the stretching movements of a molecule, E_{bend} denotes for the bending energy, E_{tors} accounts for the torsional term, E_{vdw} accounts for the van der Waals interactions, E_{ele} accounts for the electrostatic interactions and finally E_{cross} accounts for the cross terms.

$$E_{\text{total}} = \sum E_{\text{str}} + \sum E_{\text{bend}} + \sum E_{\text{tors}} + \sum E_{\text{vdw}} + \sum E_{\text{ele}} + \sum E_{\text{cross}} \quad (38)$$

a) Stretching term

The stretching potential energy of a pair of bonded atoms can be expressed as a Taylor expansion around a small equilibrium distance R_0 , truncating in the quadratic terms (harmonic approximation):

$$E_{\text{str}} = E(0) + dE/dR (R-R_0) + 1/2 d^2E/dR^2 (R-R_0)^2 \quad (39)$$

Taking into account that $E(0)$ is a scale reference and that in a minimum zone the first derivative is null, equation (39) can be reduced as following:

$$E_{\text{str}} = \frac{1}{2} K (R - R_0)^2 \quad (40)$$

Where K is the bond force constant, thus R_0 and K are the parameters of this term (and they are different for each atom pair).

There are two alternative and more accurate ways to calculate the stretching term, one simple alternative is to take more terms in the Taylor expansion (with the expensive cost to obtain more parameters) while the other is based on the use of the Morse function, which describes far from equilibrium geometries:

$$E_{\text{morse}} = D \{1 - \exp [a(R - R_0)]\}^2 \quad (41)$$

Here, D is the bond dissociation energy and a is related to the bond force constant.

b) Bending Term

Bending potential energy (energy variation for bond angle) among three bonded atoms has a similar expression, if we use the harmonic approximation is described as:

$$E_{\text{bend}} = \frac{1}{2} K (\Theta - \Theta_0)^2 \quad (42)$$

Where Θ accounts for the bond angle, Θ_0 is the equilibrium bond angle and K is the bond angle force constant. If we need to study far from equilibrium zones we can use the third order function, to treat changes of $\Theta_0 \pm 70^\circ$ correctly.

c) Torsional Term

This term is included in the force field to describe the energy variation related to the change of dihedral angle, among four bonded atoms (Figure 1). It is fundamental to treat the conformational analysis.

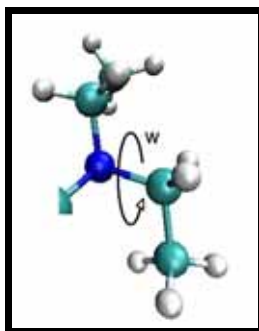


Figure 1: Torsional movement.

Usually, this term is calculated by a Fourier expansion as:

$$E_{\text{tors}} = V_n/2 (1 + \cos nw) \quad (43)$$

Being V_n the rotational energy barrier, w the dihedral angle and n the rotation periodicity ($n=1$ if it is 360° periodical, $n=2$ if it is 180° periodical ...).

Another contribution to the torsional term is the out of the plane bending. Eq. 44 is used in its harmonic approximation to describe this movement:

$$E_{\text{oop}} = K/2 \chi^2 \quad (44)$$

Being χ the angle in the next figure:

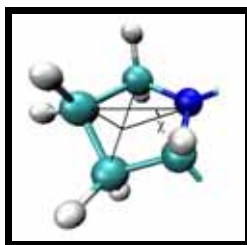


Figure 2: Out of the plane torsion.

d) Van der Waals Term

This term is employed in the force field to compute the van der Waals interactions that appear in the repulsion and attraction of non-bonded atoms. The interactions are repulsive at close distances, becomes zero at large distances and are attractive at average distances.

The origin of these forces is fundamentally based on the electronic correlation, so it has a quantum nature itself. Nevertheless, a well-known classical function employed to describe van der Waals forces is the 12-6 Lennard Jones potential:

$$E_{\text{vdw,L-J}} = \epsilon \left[\left(\frac{R_0}{R} \right)^{12} - 2 \left(\frac{R_0}{R} \right)^6 \right] \quad (45)$$

Where R_0 accounts for the distance of the function minimum and ϵ is the minimum depth. It can be seen this potential in Figure 3.

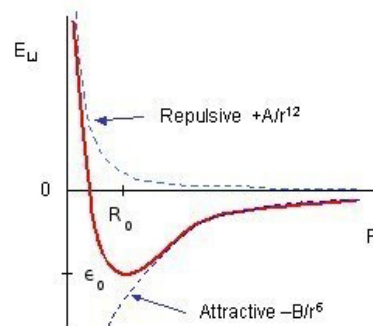


Figure 3: Lennard-Jones potential.

Another function to describe the van der Waals interactions, with higher quality is called the Hill potential:

$$E_{\text{vdw, Hill}} = A e^{-BR} - C/R^6 \quad (46)$$

Where A , B and C are treated as parameters of the force field.

e) Electrostatic Term

The second non-bonded term is related to the different charge distribution of a molecule, which creates positive and negative charged zones. Following the Coulomb law:

$$E_{el} = \frac{q_a q_b}{\epsilon R_{ab}} \quad (47)$$

Where ϵ is the dielectric constant and q_a q_b are the parameters, which are usually assigned as partial quantum-mechanical charges.

One important set of charges employed in this work is the AM1-BCC (Austin Model 1 Bond Charge Corrections) [8] which quickly and efficient generates high-quality atomic charges for organic molecules. This charge set produces atomic charges that emulate the HF/6-31G* electrostatic potential of a molecule, simply adding some bond charge corrections. Firstly it is generated a AM1 population charges and then it is added bond charge corrections. BCC has been parametrized fitting more than 2,700 molecules to the HF/6-31G* electrostatic potential.

f) Cross Term

Cross terms are included in the force field to take into account the coupling of internal coordinates: bonding distances, bonding angles and dihedral angles. For instance, when a bonding angle decreases, the bonding distances increase to minimize the steric repulsion. However a few number of cross terms are necessary to reproduce the structural properties.

As a representative example, the MM3 [6] force field calculates the stretching-bending coupling with the following expression:

$$E_{\text{strt-bend, MM3}} = 2.51 [(R - R_0) + (R' - R'_0)] (\theta - \theta_0) \quad (48)$$

Where the variables R , R' and θ are shown in Figure 4.

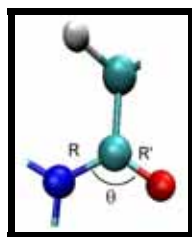


Figure 4: *Stretching-bending cross term.*

2.2.1. PARAMETRIZATION

The force field parametrization is an extremely difficult task. A first problem is the election of experimental and theoretical data which can defer from different authors.

For example, we should calculate the number of parameters required for the MM2 force field:

It contains 71 atom types, each one with two parameters of van der Waals (R_0 and ϵ), to a total of 142 parameters.

It contains 30 types of bonded atoms pair, so $30 \cdot 29 / 2! = 435$ stretching terms, with two parameters each term (K and R_0), we have a total of 870 parameters.

It contains $30 \cdot 29 \cdot 28 / 3! = 4,060$ bending terms, with two parameters K and θ_0 each one, we have a total of 8,120 parameters.

It contains $30 \cdot 29 \cdot 28 \cdot 27 / 4! = 27,405$ torsional terms, with three parameters each one (V_1 , V_2 and V_3), we have a total of 82,215 parameters !.

2.2.2. COMPUTATIONAL CONSIDERATIONS

As we have seen in the last section, a force field is constituted by a high number of parameters. In order to reduce complexity and computational cost, different approximations are implemented to simplify the force field. For instance, it can be reduced the dependence of a bending term to only the central atom, or it can be reduced the dependence of a torsional term to only the two central and bonded atoms.

On the other hand, the van der Waals and electrostatics terms are very time-consuming, being, in number, the most important part of the force field.

The simplification for the van der Waals term is carried out by taking only the interactions involved by

the atoms of a local region (around 9 \AA), which is called cutoff distance. The list of neighbour atoms is updated at a concrete time interval. This approximation is only applied to short-range interactions.

Long-range interactions, such as the electrostatic terms, are more difficult to simplify. Fortunately, a computational method which is called the particle-mesh Ewald method [9] is implemented in several molecular mechanics programs. The fundamental idea of this method is to substitute a slow converging electrostatic term of coulomb, which in a mathematical point of view is called a conditional summation, for two fast converging summations, using a gaussian distribution of neutralizing charge around one of a pair of interaction charges. The first term is calculated in real space by reducing one of the pair charges with the gaussian distribution while the second term is calculated by increasing one of them, and using the reciprocal space.

2.2.3. OPTIMIZATION METHODS

Several problems in computational chemistry are based on a function optimization. Usually one can find energy minimums, which are points with gradient zero and positive hessian matrix eigenvalues. In Molecular Mechanics, the force field minimization can be used to determine the structure of local minimums and thus, favourable conformations of the system.

The potential function to describe a protein system is composed by a complex multidimensional landscape of minimum points, and optimizing a structure consists in achieve a reasonable local minimum, but not always the global one.

We will explain here the two optimization methods employed by the AMBER program that only require the calculation of the first derivative (called first order algorithms).

a) Steepest descent method

Taking a function $F(X)$ and performing a Taylor expansion around a point X_n . For a close point $X_{n+1} = X_n + d$, eq. 49 can be obtained.

$$F(X_{n+1}) = F(X_n + d) = F(X_n) + F'(X_n) d + \frac{1}{2} F''(X_n) d^2 \quad (49)$$

An optimum election of the displacement $d = -k F'(X_n)$, with a positive and small k gives us:

$$F(X_{n+1}) = F(X_n + d) = F(X_n) - k(F'(X_n))^2 + \frac{1}{2} k^2 (F'(X_n))^2 F''(X_n) \quad (50)$$

Moreover, if we simplify the equation being k^2 very small, then:

$$F(X_{n+1}) = F(X_n + d) = F(X_n) - k(F'(X_n))^2 < F(X_n) \quad (51)$$

Thus, the following step X_{n+1} would decrease the function accordingly with the harmonic approximation. The steepest descent method is useful in a first minimization procedure but the convergence is not completely satisfactory when the structure is close to a minimum point.

b) Conjugated gradient method

This method follows an improved displacement:

$$d_n = -g_n + b d_{n-1} \quad (52)$$

Being g the function gradient and b a specific value that depends of the minimization method. This method is more efficient, but the combined use of both methods is preferred. Firstly we can minimize fast the function calculating a steepest descent routine and then change to conjugated gradient, when the structure is close to the minimum point.

On the other hand, alternative methodologies are more useful to search for a global minimum, to perform an exhaustive conformational sampling or to achieve other minimum point with less energy. For instance Molecular Dynamics, Monte Carlo methods and Simulated Annealing are used for these purposes. The former is presented in the next section.

2.2.4. MOLECULAR DYNAMICS

Molecular Dynamics is employed to study a time-dependent system with deterministic equations, using the classical Newton's laws.

In the context of Molecular Dynamics, trajectories are simulated by solving a system of differential equations based on the second Newton's law:

$$F_i = m_i d^2x_i/dt^2 \quad (53)$$

Where the forces are evaluated using the force field potential gradient:

$$F_i = - dE/dx_i \quad (54)$$

To follow the positions in every time step, we must integer the eq. 53 by numerical algorithms. One of the most popular ones is called the Verlet algorithm [10], it is described in the next equations to obtain the position, the velocity and the acceleration:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2} a(t) \Delta t^2 \quad (55)$$

$$v(t + \Delta t) = v(t) + a(t)\Delta t + \frac{1}{2} d^3r/dt^3 \Delta t^2 \quad (56)$$

$$a(t + \Delta t) = a(t) + d^3r/dt^3 \Delta t + \frac{1}{2} d^4r/dt^4 \Delta t^2 \quad (57)$$

If we add two time intervals $t + \Delta t$ and $t - \Delta t$ to the position equation, then:

$$r(t + \Delta t) = 2 r(t) - r(t - \Delta t) + a(t) \Delta t^2 \quad (58)$$

Therefore, to find the particle in the position $r(t + \Delta t)$, we only need to calculate the position in the two last steps and the acceleration.

To solve this system of equations we need an initial configuration, usually it is prepared by assigning a Maxwell-Boltzmann distribution of velocities (eq. 59) accordingly to a working temperature (eq. 60), taking also into account that there is no overall momentum in the system (eq. 61).

$$p(v_{ix}) = \left(\frac{m_i}{2\pi K_B T} \right)^{\frac{1}{2}} \exp\left[-\frac{1}{2} \frac{m_i v_{ix}^2}{2K_B T}\right] \quad (59)$$

$$T = \frac{1}{3N} \sum \frac{p_i^2}{2m_i} \quad (60)$$

$$P = \sum m_i p_i = 0 \quad (61)$$

Where $p(v_{ix})$ is the probability to obtain the atom i with the velocity v_x , m_i is the atom mass, K_B is the Boltzmann's constant, T is the temperature, N is the total number of particles, p_i is the linear momentum of each particle and P is the total momentum of the system.

Our Molecular Dynamics simulations run within the Canonical Ensemble (NVT): this is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms, N , a fixed volume, V , and a fixed temperature, T . The use of an ensemble together with the Ergodic hypothesis (eq. 62) enables Molecular Dynamics to calculate experimental properties.

$$\langle P \rangle_{\text{ensemble}} = \langle P \rangle_{\text{time}} \quad (62)$$

In the NVT ensemble, the temperature is controlled by using the following expression in a Molecular Dynamics simulation:

$$\Delta T = \frac{1}{2} \sum \frac{2}{3} \frac{(\lambda v_i)^2}{NK_B} - \frac{1}{2} \sum \frac{2}{3} \frac{v_i^2}{2K_B T} = (\lambda^2 - 1) T(t) \quad (63)$$

$$\lambda = \sqrt{\left(\frac{T_{\text{new}}}{T(t)} \right)} \quad (64)$$

Being v_i the velocity of the particle i , N the total number of particles, T the temperature, K_B the Boltzmann's constant. λ is the velocity scaling factor that is the parameter used here to alter the temperature.

Usually, Molecular Dynamics simulations are time-consuming to study long evolutions of biological systems. For instance, Table 1 shows typical time intervals involved in some protein movements. The slowest movement determines the total simulation time that should be performed.

<i>Movement</i>	<i>Time interval</i>
Turn motion	10^{-15} to 10^{-1} s
Side chain motion	10^{-15} to 10^{-1} s
Alpha helix motion	10^{-9} to 1 s
Protein folding/unfolding	10^{-7} to 10^4 s

Table 1: Usual protein movements.

One simplification to this problem, to reduce the computational cost, is to freeze the bonding distances, especially those that involve hydrogen atoms, due to these freedom degrees are not really important in the general behaviour of the system. The SHAKE algorithm [11] is usually used for this purpose.

The first macromolecular Molecular Dynamics, published in 1977, only treated a system with 500 atoms through a simulation time of 0.0092 ns [12], nevertheless the fast growth of computer power in 30 years allows the treatment of incredible big systems (such as the complete simulation of the tobacco mosaic virus, with 1 million atoms [13]) or very long simulation time (such as 500,000 ns = 500 μ s of the Villin headpiece protein [14]).

2.2.5. PERIODIC BOUNDARY CONDITIONS

A dynamical simulation without periodic boundary conditions (PBC) is not realistic because the system is surrounded by surfaces, and atoms close to the boundary would have less neighbours than atoms inside. This effect is negligible in a macroscopic system because of the big number of particles inside the system ($\sim 10^{23}$) compared with the number of them near to boundary, but it has to be reduced in a microscopic system.

A solution to this problem is to use periodic boundary conditions. When using PBC, particles are enclosed in a box, and we can imagine this box replicated to infinity by rigid translation in all the three cartesian directions, completely filling the space. Figure 5 shows the application of PBC to the system.

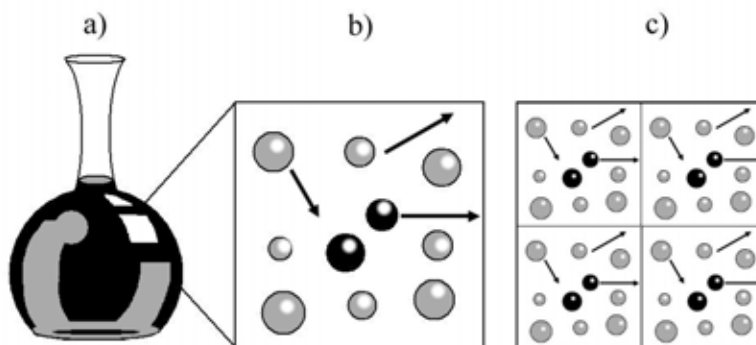


Figure 5: a) Macroscopic system, b) microscopic system without PBC and c) microscopic system with PBC.

Actually the number of interactions does not increase using PBC due to the minimum image criterion, which is based on selecting only the closest images of a particle to calculate its interactions.

2.2.6. AMBER PROGRAM

AMBER (<http://amber.scripps.edu>) is a general Molecular Dynamics package to simulate proteins, nucleic acids, sugars and organic molecules. Figure 6 shows a schematic representation of the most useful programs inside AMBER.

Several programs are included in this package, but we can classify them into three categories:

-Preparatory programs

LEAP: general program to update a system.

ANTECHAMBER: designed to treat non habitual molecules, such a organic ligands.

-Simulation Programs

SANDER and PMEMD: based programs to perform Molecular Dynamics and NMR refinement.

GIBBS: free energy perturbation energy program, to calculate free energy of different system states.

ROAR: Quantum-Mechanics/Molecular-Mechanics program.

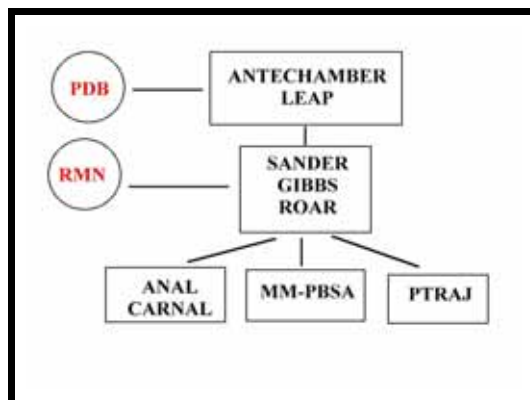


Figure 6: Schematic representation of AMBER package.

-Analysis programs

ANAL: to analyze the force field contributions to the total energy of a single structure.

PTRAJ: to analyze the dynamics trajectories and coordinates throughout a simulation.

CARNAL: It complements the PTRAJ program.

MMPBSA: Molecular Mechanics Poisson Boltzmann program, which computes the free energy of binding.

2.2.7. AMBER FORCE FIELD AND PARAMETRIZATION

The AMBER force field is shown in eq. 65:

$$E_{\text{tot}} = \sum k (R - R_0)^2 + \sum k (\Theta - \Theta_0)^2 + \sum \frac{V_n}{2} [1 + \cos(n\omega - \phi)] + \sum \sum [(A_{ij}/R_{ij})^{12} - (B_{ij}/R_{ij})^6] + \sum \sum \frac{q_i q_j}{R_{ij}} \quad (65)$$

It is formed by harmonic stretching and bending terms, a torsional term, a 12-6 Lennard Jones potential

and a simple electrostatic term.

Concerning the parameters of the force field, they have been changed along the different program versions, traditional ones use partial charges centered on each atom, such as ff86 [15], ff94 [16], ff96 [17], ff98 [18] and ff99 [19] while more modern parameters include polarizabilities and other modifications, such as ff02 [20] and ff02EP [20].

We will describe here some of the force field parameters of the AMBER program.

a) The 1984 and 1986 force fields [15]

They are described in early papers. These parameters are not recommended any more, but still are useful to perform vacuum simulations of proteins and nucleic acids in a distance dependent dielectric medium.

b) The 1994 and 1998 force fields

Contained in the ff94 parameter set [16], they are called 'second generation' parameters. They are especially derived for solvated systems computing charges obtained at Hartree-Fock 6-31 G* level of theory.

The ff94 parameters have been used in the present work, because they have been extensively tested.

The ff96 [17] differs from ff94 in that the torsions for ϕ and ψ angles were modified in response to ab initio calculations, finally ff98 [18] differs from ff94 in the parameters involving the glycosidic torsions in nucleic acids.

c) The 1999 and 2002 force fields

The ff99 [19] represents a new direction of parameters, pointing towards organic and bioorganic systems. Moreover they include the atomic polarizabilities.

The ff02 [20] parameters are a polarizable variant of ff99 using charges calculated at B3LYP/cc-pVTZ//HF/6-31G* level of theory and finally ff02EP [20] adds additional point charges to mimic electron lone pairs of O, N and S atoms.

2.3. INTRODUCTION TO DRUG DESIGN

Although there are a lot of drugs in the pharmaceutical industry, the development of new and more efficient drugs is still a difficult and exciting work. At the present, the structural knowledge of biomolecules is very useful for a rational drug design scenario. Now, even with theoretical studies it is possible to design new molecules to evaluate with success in experimental stages and with a specific activity against a particular target. Some guidelines of drug design will be described in this section.

A drug can be defined as a substance with a beneficial effect for an altered biological system. In a molecular point of view, this is translated into an interaction between the drug and the target. Targets can be lipids (when the drug has affinity for the cell bilayer), nucleic acids (when the molecule interacts by disrupting the DNA chains) and proteins, which are the most important targets.

When the protein acts as an enzyme, the drug can be classified in the following forms:

Reversible and competitive inhibitor: the drug structure is very similar to the natural ligand, thus the protein recognizes both molecules (Figure 7).

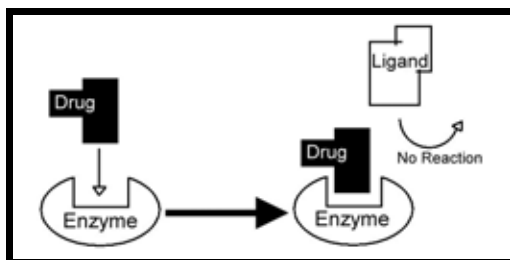


Figure 7: Competitive inhibition.

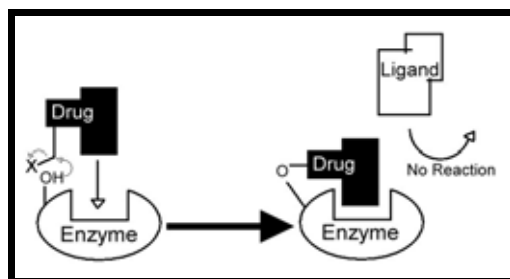


Figure 8: Non competitive inhibition.

Non competitive and irreversible inhibitor: the drug interacts covalently with the protein and blocks the action of the natural ligand even with a big excess of it (Figure 8).

Allosteric inhibitor: the drug can bind to a different protein pocket, but its interaction distort the active center in such a way that the enzyme is inhibited (Figure 9).

When the protein is a membrane receptor, the drug can be classified as:

Agonist: if the drug binds to the receptor activating it in the same way as the natural ligand.

Antagonist: if the molecule binds to the receptor and the biological effect is different to the natural ligand effect.

Partial agonist: when the drug only produces a partial receptor activation.

Inverse agonist: when the drug produces a total receptor inactivation.

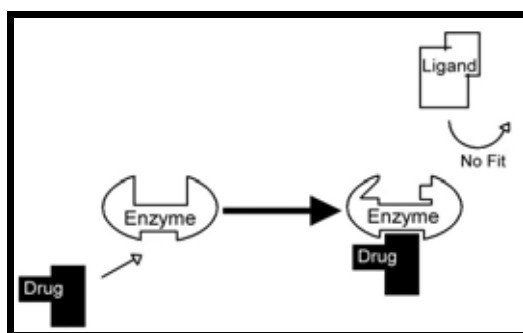


Figure 9: Allosteric inhibition.

2.3.1. THE PHARMACOPHORE

First defined by P. Ehrlich in 1909, the modern definition of a pharmacophore is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response. These features can be basic and acid centers, charged centers, hydrophobic volumes, hydrogen acceptors, hydrogen donors...

A pharmacophore also includes informations related to distances, angles and planes among all these groups. Figure 10 shows the 5-point pharmacophore of adrenaline, taken as an example, when this molecule interacts with its receptor.

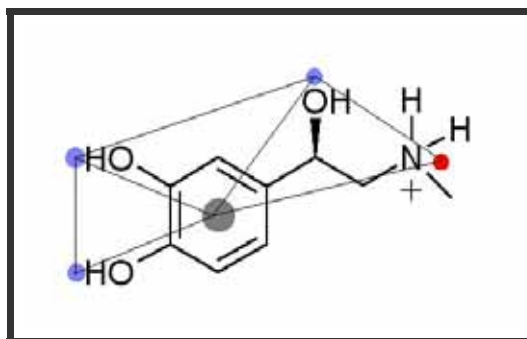


Figure 10: Adrenaline pharmacophore.

2.3.2. DRUG DESIGN, PHARMACOKINETICS AND LIPINSKY RULES

There are several strategies which can be used to improve the interaction between the drug and its target [21], most useful ones include:

- Variation of substituents, - Extensions of the structure, - Ring expansions/contractions, - Ring variations, - Ring fusions, - Isosteres, - Simplification of the structure and - Rigidification of the structure.

These guidelines of a drug design process, will be described in this section.

a) Variation of substituents

Varying easily accessible substituents is a common method to modify binding interactions. For example, the alkyl substituents of ethers, amines, esters and amides are easily varied as shown in Figure 11. The alkyl substituent already present can be removed and replaced with another substituent. Nevertheless alkyl groups which are part of the skeleton of the molecule are not easily modified without a full synthesis of the compound.

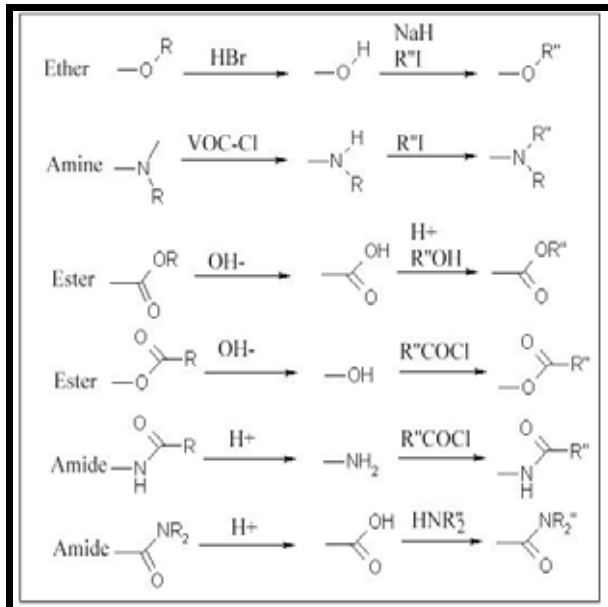


Figure 11: Alkyl modifications.

On the other hand, alkyl groups as methyl, ethyl, propyl, butyl, isobutyl and tertbutyl are often used to investigate the steric hindrance of hydrophobic pockets. For example, isoprenaline (Figure 12) is an analogue of adrenaline (Figure 10) where a methyl group has been replaced by an isopropyl, resulting in an enhancing of activity in beta-adrenergic receptors over alfa-adrenergic receptors.

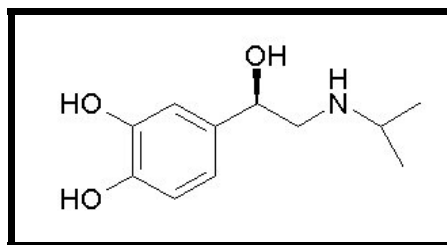


Figure 12: Isoprenaline.

In addition, if a drug contains an aromatic ring, it is relatively easy to vary the substituent position to find a better recognition pattern (Figure 13).

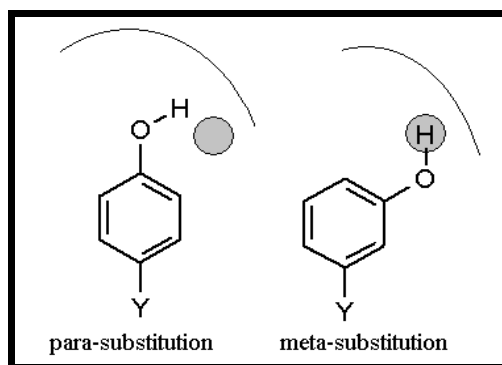


Figure 13: Aromatic substitution.

b) Extensions of structure

The drug extension strategy is based on the addition of an extra functional group to increase the number of interaction points. We can add extra hydrogen bonding, ionic interactions or extra alkyl groups to find hydrophobic regions. Figure 14 shows an example of the development of an angiotensin converting enzyme (ACE) inhibitor with application as antihypertensive agent [21], in which a new phenyl group recognizes a hydrophobic pocket.

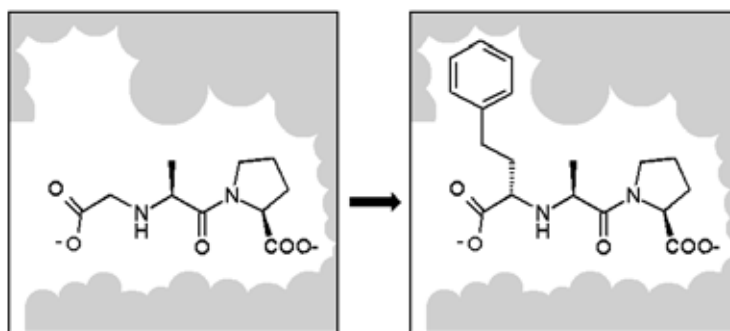


Figure 14: ACE inhibitor and chain extension.

c) Ring expansions/contractions

Expanding or contracting a ring puts the binding groups in a different position relative to each other and may increase the activity (Figure 15).

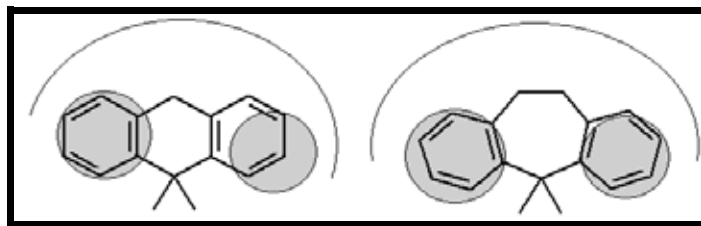


Figure 15: Ring expansion.

d) Ring variations

This is referred to a replacement of the original ring with a range of other heteroaromatic rings of different size and heteroatom position. One advantage of altering an aromatic ring to a heteroaromatic is the introduction of an extra potential hydrogen bond.

e) Ring fusion

Extending a ring by using a ring fusion can sometimes result in increased interactions or selectivity. Moreover, with this strategy a drug can be less flexible and more selective (presenting also a less entropic binding penalty).

f) Isosteres

Isosteres [21] or different functional groups with highly similar physicochemical properties, have often been used in drug design to vary the character of the molecule with respect to size, polarity or electronic distribution.

For example, fluorine is considered a hydrogen isoster because it has virtually the same size. However, it is a more electronegative atom and can be used to vary electronic properties without having any steric effect.

g) Simplification of the structure

Simplification is commonly used on complex lead compounds arising from natural sources. Essential groups of the drug are reminded while non-essential ones can be removed without losing activity, Figure 16 shows an hypothetic strategy [21] to simplify a structure:

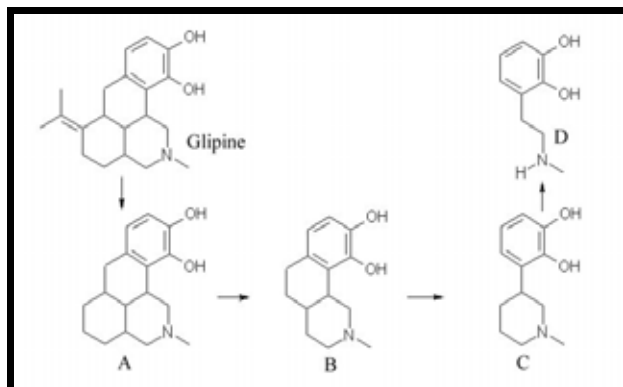


Figure 16: Glipine analogs.

h) Rigidification of the structure

The strategy of rigidification is to lock the drug into a more rigid conformation such it cannot take undesirable conformations. This tactic is used to increase activity or selectivity. An example of a rigidification of an inhibitor of platelet aggregation [21] is shown in Figure 17.

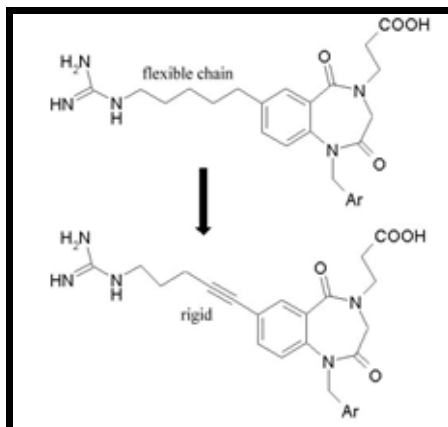


Figure 17: Rigidification of a lead compound.

Thus, physical and chemical properties of drugs are very important, but compounds with the best binding are not necessarily the best drugs to use in clinical stages.

Some pharmacokinetics issues [21] to take also into account are: the hydrophobic/hydrophilic balance, the ionizable state, the size and the number of hydrogen bonding interactions.

a) Hydrophilic/ hydrophobic balance

A drug must have the correct hydrophilic/hydrophobic balance. Without this balance, drugs have several disadvantages. Too polar drugs are easily excreted by the kidneys and do not cross the cell membrane. On the other hand, too lipophilic drugs show poor solubility in water and poor absorption in the gastrointestinal tract.

b) Ionization state

A drug strongly ionized has difficulty crossing the cell bilayer. An alternative to solve this problem is to design a drug with its ionizable groups temporally masked or to take advantage of the cell's own carrier proteins, which are designed to carry ionic molecules (such as sugars, amino acids or metal ions).

On the other hand, a partially ionized state allows to cross the cell bilayer in the non-ionized form, while the ionized form gives the drug solubility and good binding interactions with its receptor.

c) Size

Most useful drugs have a molecular weight less than 500. However, size itself is not a barrier for absorption, a more limitation factor is polarity. Nevertheless, it is often the case that large molecules are poorly absorbed, not because of their size, but because they have larger number of polar groups.

d) Number of hydrogen bonding interactions

The more hydrogen bonding groups on a molecule, the less it will be absorbed. Usually, compounds with more than 5 hydrogen bond donors or 10 hydrogen bond acceptors are poorly absorbed.

As a summary of the section 2.3.2., and based on a statistical treatment of the commercial drugs available in 2001, Lipinsky et al. [22] described a simple rule, which is called 'the rule of 5', that a potential compound should comply in order to penetrate across the cell bilayer. They stated that poor absorption or permeation are more likely when:

There are more than 5 H-bond donors (expressed as the sum of OHs and NHs).

The molecular weight is over 500.

The log P (octanol-water partition coefficient) is over 5.

There are more than 10 hydrogen bond acceptors (expressed as the sum of Ns and Os).

Hence, in this work we have used these rules (when it was possible) to assure an acceptable activity at cell level.

2.3.3. STRUCTURE-BASED METHODS

By means of experimental methods, such as the X ray diffraction and the NMR spectroscopy, it is possible to obtain, with high resolution, the structure of a protein. For instance, the Protein Data Bank which is a public database contains around 30,000 solved proteins.

Usually there are also a lot of proteins complexed with their natural ligands or well-know inhibitors. This structural information is necessary in the so called structure-based drug design, which is the methodology used in this work.

2.3.4. 3D DATABASE SEARCHING

Several public and private organizations have databases of their chemical compounds. In these databases we can find a molecule which complains the conformational and the chemical group restrains imposed by a pharmacophore. For this purpose, these databases usually contain fast algorithms to search for the most favourable conformation of their compounds.

In this work, we have used the CATALYST™ program [23] together with our pharmacophores of reference to find for commercial molecules which could be actives in biological assays. At the present, the CATALYST™ database contains the following companies:

Available Chemical Directory (ACD): 266,812 compounds, Derwent World Drug Index: 63,307 compounds, BioByte (Pomona College): 39,383 compounds, National Cancer Institute (NCI): 238,819 compounds, Maybridge: 59,652 compounds, SPECS: 255,000 compounds and ChemDiv New Chemistry: 95,209 compounds.

2.3.5. DOCKING

Docking is the strategy used to predict the binding place, conformation and affinity of the drug into its biological target. It is very useful in the virtual screening of databases, because it is a fast methodology to select the most promising molecules to target a biomolecule.

There are several methodologies to perform a docking procedure, the most applied ones use molecular dynamics information [24], genetic algorithms [25] or some information about the receptor flexibility [26]. In the present work, we have primary used our home-made program Dock-Dyn [24]. Some of the basic characteristics of this program are described here:

Dock-Dyn is a program of directed docking, in a first step it tests if the ligand satisfies a reference pharmacophore (extracted for instance of a known inhibitor) which can be composed of 6 types of interaction points: HA (hydrogen acceptor), HD (hydrogen donor), P+ (positive charge), P- (negative charge), HI (hydrophobic point), AR (aromatic ring). These points are described with their x-y-z coordinates (imposed as a restriction) and a certain uncertain interval, both values obtained by a Molecular Dynamics simulation.

In a second step, those ligands with a correct pharmacophore, so similar to the reference one, are superimposed to the reference pharmacophore placed inside the receptor to evaluate their molecular recognition. In this moment, we take into account the receptor and the van der Waals radii of the atoms and we only accept those molecules without steric hindrance. To obtain more accepted molecules, we can relax the van der Waals interaction by reducing the atomic radii of the atoms, multiplying them by a reduction factor. In addition, a minimization routine can be also carried out.

Finally, accepted conformations of each ligand are saved and evaluated by their RMS (root mean square) deviation with respect to the reference pharmacophore coordinates.

Later, they can be also evaluated by using a scoring methodology, this will be explained in the next section.

2.3.6. SCORING

The scoring process [27] is based on the evaluation and order of the different accepted conformations during a docking procedure. It is of capital importance in order to discriminate those ligand with more features to become actives among all the database selected compounds. Moreover, due to the high number of conformations to evaluate, a fast algorithm is preferred.

Recently, 11 scoring functions were compared [28] using a significative number of ligands and receptors, being the XSCORE [29], one of the best performing ones. We will discuss about XSCORE, which is the scoring function selected in this work to evaluate the poses generated from the docking procedure. XSCORE, is so called an empirical scoring function, because it was calibrated using 200 ligand-receptor complexes (and 800 ones in its last version) with experimental information such as dissociation or association constants.

The program calculates the binding free energy related to the receptor-ligand complex formation, by means of the eq. 66.

$$\Delta G = \Delta G_{\text{vdw}} + \Delta G_{\text{h-bond}} + \Delta G_{\text{deform}} + \Delta G_{\text{hydro}} + \Delta G_0 \quad (66)$$

ΔG_{vdw} : this term denotes the van der Waals interaction between the fragments, using a 8-4 Lennard Jones potential, which is softer than the habitual 12 -6 potential:

$$\text{VDW} = \sum \sum \text{VDW}_{ij} = \sum \sum [(d_{ij,0}/d_{ij})^8 - 2 (d_{ij,0}/d_{ij})^4] \quad (67)$$

Here, i denotes for ligand atoms, j denotes for receptor atoms and $d_{ij,0}$ is the sum of van der Waals radius of atom i and j , while $d_{ij} = r_i + r_j$ is the distance between the ligand atom i and the protein atom j .

$\Delta G_{\text{h-bond}}$: this term accounts for the number of hydrogen bonds between the fragments by means of a simple geometric depending function:

$$\text{HB} = \sum \sum \text{HB}_{ij} = f(d_{ij}) f(\theta_{1,ij}) f(\theta_{2,ij}) \quad (68)$$

Where the distance function $f(d)$ and the angular functions $f(\theta_1)$, $f(\theta_2)$ are written in the following forms and illustrated in Figure 18:

$$f(d) = 1.0 \text{ for } d \leq r_i + r_j - 0.7 \text{ \AA}$$

$$f(d) = 1/0.7 (r_i + r_j - d) \text{ for } r_i + r_j - 0.7 \text{ \AA} < d \leq r_i + r_j$$

$$f(d) = 0.0 \text{ for } d > r_i + r_j$$

$$f(\theta_1) = 1.0 \text{ for } \theta_1 \geq 120^\circ$$

$$f(\theta_1) = 1/60 (\theta_1 - 60) \text{ for } 120^\circ > \theta_1 \geq 60^\circ$$

$$f(\theta_1) = 0.0 \text{ for } \theta_1 < 60^\circ$$

$$f(\theta_2) = 1.0 \text{ for } \theta_2 \geq 120^\circ$$

$$f(\theta_2) = 1/60 (\theta_2 - 60) \text{ for } 120^\circ > \theta_2 \geq 60^\circ$$

$$f(\theta_2) = 0.0 \text{ for } \theta_2 < 60^\circ$$

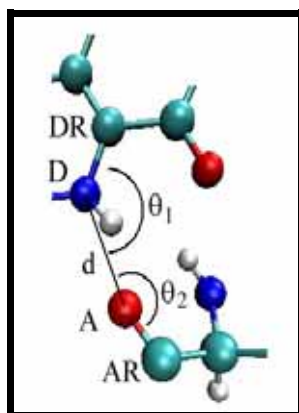


Figure 18: Illustration of the three geometric parameters used in characterizing a hydrogen bond, DR: donor root atom, D: donor atom, A: acceptor atom and AR: acceptor root atom.

ΔG_{deform} : It denotes for the negative entropic factor, related to the loss of ligand freedom degrees (number of ligand rotatable bonds) when it interacts into the receptor:

$$RT = \sum RT_i \quad (69)$$

Where $RT_i = 0$ if the atom i is not involved in any rotor, $RT_i = 0.5$ if the atom i is involved in one rotor, $RT_i = 1.0$ if atom i is involved in two rotors and $RT_i = 0.5$ if atom is involved in more than two rotors.

ΔG_{hydro} : It evaluates the complex desolvation process (or hydrophobic effect). The program calculates this factor by using three different algorithms, the first one (HS or surface effect) takes into account the ligand solvent accessible surface which becomes buried when the ligand interacts with the receptor (eq. 70). The second one, (HC or contact effect) estimates the interaction of hydrophobic atoms, with a simple distance dependent function (eq. 71). Finally, the third one (HM or match effect) calculates how hydrophobic is the close region of each hydrophobic atom of the ligand (eq. 72) :

$$HS = \sum SASA_i \quad (70)$$

$$HC = \sum \sum f(d_{ij}), \quad (71)$$

$$f(d_{ij}) = 1.0 \text{ for } d \leq r_i + r_j - 0.5 \text{ \AA}$$

$$f(d_{ij}) = 1/1.5 (r_i + r_j - d) \text{ for } r_i + r_j + 0.5 \text{ \AA} < d \leq r_i + r_j + 2.0 \text{ \AA}$$

$$f(d_{ij}) = 0.0 \text{ for } d > r_i + r_j + 2.0 \text{ \AA}$$

$$HM = \sum HM_i \log P_i \quad (72)$$

Being P_i the octanol/water partition coefficient and HM_i is set to 1 if a hydrophobic atom i is placed in a hydrophobic environment, otherwise is set to 0.

ΔG_0 : this term appears by adjusting the function to experimental data, it must include entropic effects not taken into account by the other XSCORE terms.

Each term is multiplied by a constant factor, which was found by adjusting the values of the different terms to experimental binding free energies. Due to this parametrization, XSCORE performs very fast and it was identified a small average deviation of 2 kcal/mol among all the studied systems. One of the failures of the method, which was confirmed in our work, is the difficult to discriminate between good and poor inhibitors of a specific target, when they have similar structures and interaction points.

2.3.7. MMPBSA METHODOLOGY

MMPBSA (Molecular Mechanics Poisson Boltzmann Solvation Area) [30] is used in drug design to calculate in an accurate manner, the free energy of binding between ligand-protein, protein-protein or nucleic acid complexes. Coupling together molecular dynamics and continuum models, this method was firstly described in the year 2000, and it has been used with success for several biological protein and nucleic acid complexes [31-33].

Free energy of binding is not calculated directly but by using a thermochemical pathway to introduce the solvation:

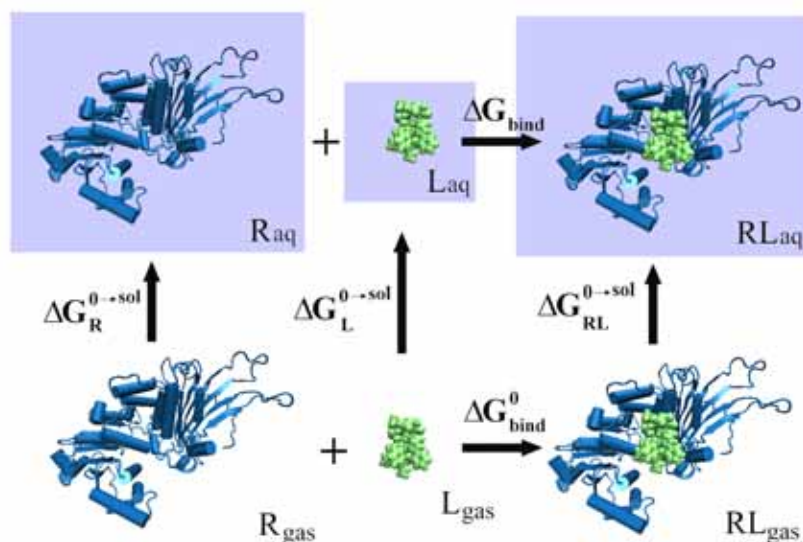


Figure 19: Thermochemical pathway related to free energy calculation in MMPBSA.

Where R accounts for the receptor, L accounts for the ligand and RL accounts for the ligand-receptor complex.

Using the pathway of Figure 19, free energy of binding can be expressed as:

$$\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^0 + \Delta G_{\text{RL}}^{0 \rightarrow \text{sol}} - \Delta G_{\text{R}}^{0 \rightarrow \text{sol}} - \Delta G_{\text{L}}^{0 \rightarrow \text{sol}} \quad (73)$$

ΔG_{bind}^0 denotes for the free energy of binding in vacuo, and $\Delta G^{0 \rightarrow \text{sol}}$ accounts for the solvation free energy of R, L and RL.

In vacuo free energy can be calculated as following:

$$\Delta G_{\text{bind}}^0 = \Delta H_{\text{bind}}^0 - T \Delta S_{\text{bind}}^0 \quad (74)$$

Being ΔH_{bind}^0 the internal energy variation of the system, extracted from the molecular mechanics force field and ΔS_{bind}^0 is the entropic factor, which is calculated in this work by means of the normal mode analysis (NMA) that requires the diagonalization of the hessian matrix, a normal mode extraction and a statistical mechanics treatment. Total entropy is calculated by adding the rigid body entropy (rotational and translational modes) to the vibrational entropy. This contribution is obtained by the eq. 75 once the normal modes are calculated.

$$T S_{\text{vib}} = \sum \left\{ \left[\frac{h\nu_i}{K_B T} \right] \ln \left[\exp\left(\frac{h\nu_i}{K_B T}\right) - 1 \right] - K_B T \ln \left[1 - \exp\left(-\frac{h\nu_i}{K_B T}\right) \right] \right\} \quad (75)$$

Being T the temperature, h the Planck's constant, K_B the Boltzman's constant and ν_i each normal mode frequency.

On the other hand, the solvation contribution can be decomposed and evaluated using the eq. 76:

$$\Delta G^{0 \rightarrow \text{sol}} = \Delta G_{\text{elec}}^{0 \rightarrow \text{sol}} + \Delta G_{\text{np}}^{0 \rightarrow \text{sol}} \quad (76)$$

Where the former is the electrostatic contribution to solvation, and the latter is the non polar contribution to solvation. The electrostatic term is computed by solving the Poisson-Boltzmann (PB) equation, when the solvent is referred as a continuum dielectric medium:

$$\nabla [\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] - K' \sinh[\phi(\mathbf{r})] = -4\pi \rho(\mathbf{r}) \quad (77)$$

$$K' = 8\pi N_A e^2 I / 1000 \varepsilon K_B T$$

Here, $\varepsilon(\mathbf{r})$ accounts for the dielectric constant of each point, $\phi(\mathbf{r})$ is the electrostatic potential in each

point, $\rho(r)$ is the solute charge density in each point, N_A is the Avogadro's constant, e is the elementary charge, I is the ionic strength, K_B is the Boltzmann constant and T is the temperature of the system.

These equations add together the electrostatic potential function due to the charge density in a medium with a non constant dielectric, and the Boltzmann distribution of mobile ions (Na^+ and Cl^-) into the solution, which originates the $\sinh[\phi(r)]$ term.

We can linearize the Poisson-Boltzmann equation expressing the hyperbolic sinus as a Taylor expansion truncated in the first term. This approximation is useful in low charged systems (like the proteins) and it is not so accurate in highly charged systems (like DNA). Thus, the linear Poisson-Boltzmann equation is shown:

$$\nabla [\epsilon(r) \nabla \phi(r)] - K' [\phi(r)] = -4\pi \rho(r) \quad (78)$$

This equation, even being simplified, has no analytical solution in non-symmetric systems, and it is usually solved by the numerical finite difference method [34]. This method of differential equation resolution requires for a grid construction around the system to assign a charge, dielectric constant and ionic strength to each point of the grid. With this method we can solve the electrostatic potential in every grid point. Finally taking the electrostatic potential, the electrostatic contribution to the free energy of solvation is expressed as following:

$$\Delta G_{\text{elec}}^{0 \rightarrow \text{sol}} = \frac{1}{2} \sum q_i (\phi_i^{\text{sol}} - \phi_i^0) \quad (79)$$

Where q_i is the charge in each point and ϕ_i^{sol} , ϕ_i^0 are the potentials when we define the solvent, with relative dielectric constant of 80 and the potential in the same point when we calculate the energy in vacuo, using a value of 0 for the relative dielectric constant.

Moreover, a simplified version of the linear Poisson-Boltzmann method can be obtained by using the Generalized Born (GB) equation [35]:

$$\Delta G_{\text{elec GB}}^{0 \rightarrow \text{sol}} = - \frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \sum \frac{q_i q_j}{f_{GB}} \quad (80)$$

Where ϵ is the dielectric constant of the solvent, q_i and q_j are the electrostatic charges of particles and f is a functional depending of distance between particles (r_{ij}) and defined radii (a_i , a_j) called effective Born radii (eq. 81). These radii characterize their degree of burial inside the solute and this is critical for an accurate estimation of the Generalized Born energy.

$$f_{GB} = (r_{ij}^2 + a_{ij}^2 e^{-D})^{1/2} \quad (81)$$

$$D = \left(\frac{r_{ij}}{2a_{ij}} \right)^2, \quad a_{ij} = \sqrt{a_i a_j} \quad (82)$$

The Tsui et al. [36] parametrization has been applied in this work to solve the GB equation.

On the other hand, the non polar contribution to the free energy of binding, $\Delta G_{np}^{0 \rightarrow sol}$, has been solved by using the next expression:

$$\Delta G_{np}^{0 \rightarrow sol} = \gamma \text{SASA} - b \quad (83)$$

Being γ equal to 0.00542 kcal/mol \AA^2 , b equal to 0.92 kcal/mol and being the SASA the accessible surface area, computed in the present work throughout the LCPO (linear combination of pairwise overlaps) method [37].

Eq. 83 is an empirical expression, accordingly with a correlation between the free energy of solvation and the SASA of some simple molecules [38]. Although very crude for biological systems this expression is widely used to compute this contribution.

Finally, it is worth to remark that the MMPBSA method is applied to a representative set of receptor-ligand structures throughout a dynamic simulation. The results are averaged and thus they include a statistical component. Nevertheless, recent applications of the MMPBSA to a single and minimized structure also show accurate predictions [39]. This alternative to the usual MMPBSA protocol is more appropriate to study a large number of ligands.

3. REFERENCES

1. Química Quàntica, Paniagua J. C., Alemany P., Llibres de l'Índex. Universitat. **2000**.
2. Modern Quantum Chemistry, Szabo A., Ostlund N.S., New York : Mc Graw-Hill. **1989**.
3. Dewar, M. J. S.; Zoebisch, E.G.; Healy, E. F.; Steward, J. J .P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
4. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheathan III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. D.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Sligh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, Univ. California, San Francisco, 2002.
5. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM : a program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem* **1983**, *4*, 187-217.
6. Allinger, N. L.; Yuh, Y. H.; Lii, J. Molecular mechanics, the MM3 force field for hydrocarbons. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8566.
7. Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.R.; Hwang, M. J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfish, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. *J. Comp. Chem.* **2001** *15*, 1782-1800,
8. Jakalian, A; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model. Parametrization and validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641,
9. Allen, M. P.; Tildesley ,D. J. Computer simulations of liquids. Clarendon, Oxford University Press. London (1987).
10. Verlet, L. Computer experiments on classical fluids, *Phys. Rev.* **1967**, *159*, 98-103,
11. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of

motion of a system with constraints : molecular dynamics of n-alkanes. *J. Comp. Phys.* **1977**, *23*, 327-341.

12. Stillinger, F. H.; Rahman, A. Improved simulation of liquid water by molecular dynamics. *J. chem. Phys.* **1974**, *60*, 1545–1551.

13. Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure.* **2006**, *14*, 437-449.

14. Lucent, D.; Vishal, V.; Pande, V. S. Protein folding under confinement: a role for solvent. *PNAS.* **2007**, *104*, 10430-10434.

15. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all-atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **1986**, *7*, 230-252.

16. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D.M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J.W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

17. Kollman, P. A.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The development and application of a minimalistic organic/biochemical molecular mechanics force field using combination of ab initio calculations and experimental data. *Comp. Sim. Bio. Sys.* **1997**, *3*, 83-96.

18. Cheatman III, T. E.; Cieplak, P.; Kollman, P. A. A modified version of the cornell et al. Force field with improved sugar pucker phases and helical repeat. *J. Bio. Struct. Dyn.* **1999**, *16*, 845-861.

19. Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules ? . *J. Comp. Chem.* **2000**, *21*, 1049-1074.

20. Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models of organic and biological systems going beyond the atom centered two body additive approximation : aqueous solution free energies of methanol and n-methyl acetamide, nucleic acid base and amide hydrogen bonding and chloroform/water partition coefficients of nucleic acid bases. *J. Comp. Chem.* **2001**, *22*, 1048-1057.

21. Patrick, G. L. An introduction to medicinal chemistry, Oxford University Press, **2001**.

22. Lipinsky, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Delivery Reviews* **2001**, *46*, 27-43.

23. CATALYST™ (Accelrys Inc. USA).

24. Rubio-Martínez, J.; Pinto, M.; Tomás, M. S.; Pérez, J. J. Dock_dyn: a program for fast molecular docking using molecular dynamics information. University of Barcelona and Technical University of Catalonia. Barcelona, 2005.

25. Cechini, M.; Kolb, P.; Majeux, M.; Caflish, A. Automated docking of highly flexible ligands by genetic algorithms : a critical assessment. *J. Comp. Chem.* **2004**, *25*, 412-422.

26. Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementary. *Perspt. Drug Des. Discov.* **2000**, *20*, 171-190.

27. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery : methods and applications. *Nature Rev.,Drug Discov.* **2004**, *3*, 935-949.

28. Wang , R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287-2303.

29. Wang, R.; Lai, L.; Wang, S.; Further development and validation of empirical functions for structure-based binding affinity prediction. *J. Comp. Aid. Mol. Des.* **2002**, *16*, 11-26.

30. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srivasan, J.; Case, D. A.; Cheatham III, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.

31. Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode of HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221-5230.

- 32.** Gohlke, H.; Case, D. A. Converging free energy estimates: MMPBSA studies on the protein-protein complex Ras-Raf. *J. Comp. Chem.* **2004**, *25*, 238-250.
- 33.** Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D. Investigations of neuroaminidase substrate recognition using molecular dynamics and free energy calculations. *J. Med. Chem.* **2003**, *46*, 5628-5637.
- 34.** Nicholls, A.; Honing, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* **1991**, *12*, 4335-4445.
- 35.** Bashford, D.; Case, T. A. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *224*, 473-486.
- 36.** Tsui, V.; Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Nucleic Acid Sci.* **2001**, *56*, 275-291.
- 37.** Weiser, J.; Shemkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comp. Chem.* **1999**, *20*, 217-230.
- 38.** Sitkoff, D.; Sharp, K.; Honing, B. Accurate calculations of hydration free energies using macroscopic solvents. *J. Phys. Chem.* **1994**, *98*, 1978-1988.
- 39.** Kuhn, B.; Gerber, P.; Schulz-Gash, T.; Stahl, M. Validation and use of MMPBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48*, 4040-4048.

Published: J. Mol. Recognit. 2008, 21, 190-204.

CHAPTER II: Protein-protein recognition as a first step
towards the inhibition of XIAP and Survivin anti-apoptotic proteins



1. BRIEF INTRODUCTION

Apoptosis, also called programmed cell death, is a conserved mechanism inherent to all cells that sentences them to death when they receive the appropriate external stimuli. Inhibitor of apoptosis proteins (IAPs) are a family of regulatory proteins that suppress such cell death. XIAP is the most commonly studied member of the IAP family. It binds to and inhibits Caspases, an important family of apoptotic proteases. In addition, XIAP over-expression has been detected in numerous types of cancer. Smac/DIABLO, a mitochondrial protein that binds to IAPs and promotes Caspase activation, has the opposite action to XIAP and can be considered a key protein in the regulation of IAPs. Survivin, the smallest IAP protein, has received a lot of attention due to its specific expression in many cancer cell lines. It has been shown to interact with Smac/DIABLO, even though the structure of this complex has not yet been reported.

We analysed the protein-protein interactions appearing in the Smac/DIABLO-XIAP and Smac/DIABLO-Survivin complexes fully, using molecular dynamics simulations. This information is a first step towards the design of Smac/DIABLO peptidomimetics that could be used as innovative therapeutic agents for the treatment of malignancy.

Our results complement the experimental interactions described for the first complex and provide a detailed description for the second. We show that Smac/DIABLO interacts in a similar way with both targets through its amino terminal residues. In addition, we identify a pharmacophore formed by eight stable protein-protein interactions for the XIAP complex and seven stable protein-protein interactions for the Survivin complex, which describe the whole contact surface.

This information is used to suggest the binding mode of embelin, the first non-peptidic inhibitor of XIAP, and two of its derivatives. Molecular docking and molecular dynamics simulations were also carried out to describe ligand and receptor flexibility.

Finally, an MMGBSA protocol was used to obtain a more quantitative description of the binding in all the complexes studied.

2. CONTEXT

Apoptosis [1] is a complex biochemical mechanism by means of which the organism keeps its number of cells balanced (Figure 1). There are two pathways that initiate apoptosis. The first pathway, named the extrinsic pathway, is mediated by membrane receptors (CD95, TNF receptors and TRAIL receptors) that activate an enzymatic cascade of Caspases. Caspase-8 and Caspase-10, named initiator Caspases, act in the first step. They activate, by proteolytic networks, Caspase-3 and Caspase-7, named effector Caspases because they cause the morphological changes in a cell during apoptosis.

IAPs (inhibitor of apoptosis proteins) [2] are a family of regulatory proteins related to this mechanism of cell death. Eight subtypes of human IAPs have been discovered and XIAP is considered an outstanding member of this family. It inhibits Caspase-3, Caspase-7 and Caspase-9 through the BIR (baculoviral IAP repeat) domains. Specifically, XIAP interacts with Caspase-9 through the BIR3 domain and interacts with Caspase-3 and Caspase-7 through the BIR2 domain and the BIR1-BIR2 interdomain zone.

The second pathway, named the intrinsic pathway, promotes apoptosis with the appropriate activation of mitochondrial proteins, such as Smac/DIABLO, HtrA2 and Cytochrome c. Cytochrome c induces the activation of Caspase-9, while HtrA2 and Smac/DIABLO bind to and inhibit XIAP [3].

At present, the structures of XIAP with Caspase-3 [4], Caspase-7 [5-6], Caspase-9 [7] and the complex with Smac/DIABLO [8-9] have been reported. These last two studies showed, by means of NMR and X-ray techniques, that only four amino terminal residues (AVPI) of Smac/DIABLO are required for its function. The suggested interactions reported in these articles are shown in Table 1. These interactions basically involve the above-mentioned four residues of Smac/DIABLO and some conserved residues of the BIR3 domain of XIAP. It should be noted that this short peptide sequence (AVPI) maintains this high number of interactions with the receptor, as confirmed by the solved structures.

Many cancer cell lines such as hepatocellular carcinoma [10] or myeloid leukaemia [11] over-express XIAP. Furthermore, Smac/DIABLO sensitises the cell to apoptotic stimulus in melanoma [12], ovary cancer [13] and prostate cancer [14], among others. For these reasons, development of small peptidomimetics of Smac/DIABLO is a promising new way of treating cancer [15-16], characterized by apoptotic resistance.

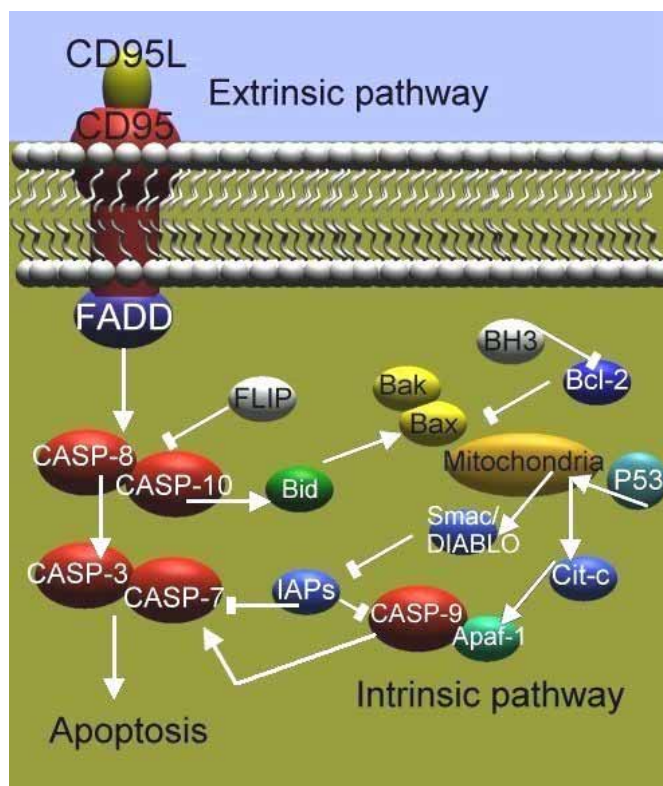


Figure 1: Main events during the apoptosis process. This work is focused on the Smac/DIABLO protein and its interactions with the IAPs, XIAP and Survivin.

XIAP	Smac/DIABLO	Interaction type
E314	Amino terminal of A1	Electrostatic
W310 (+ L307 and Q319)	Methyl of A1	Van der Waals
Methyl of T308	Side chain of V2	Van der Waals
Indole ring of W323	Side chain of P3	Van der Waals
K297 to K299 zone	Side chain of I4	Van der Waals
W323	Side chain of V2	Van der Waals
Y324	Side chain of P3	Van der Waals
Indole NH group of W323	Carbonyl of A1	Hydrogen bond
Carbonyl of T308	NH group of V2	Hydrogen bond
NH group of T308	Carbonyl of V2	Hydrogen bond
Carbonyl group of G306	NH group of I4	Hydrogen bond
NH ₂ group of Q319	Carbonyl of A1	Hydrogen bond

Table 1: Suggested interactions of the Smac/DIABLO-XIAP complex from experimental data (Liu et al., 2000; Wu et al., 2000).

Along these lines, various studies trying to find small molecules that bind the BIR3 pocket of XIAP have been reported recently. Kipp *et al.* [17] used a fluorescence assay to test the binding of a library of tetrapeptide molecules based on the amino terminal sequence of Smac/DIABLO. Glover *et al.* [18] used a high-throughput fluorescence polarization assay to screen two different sets of compounds from the National Cancer Institute database with only relative success. Oost *et al.* [19] generated a series of capped tripeptides, containing unnatural amino acids, which bind with nanomolar affinity. Tripeptide BIR3 inhibitors with unnatural amino acids have also been identified by Park *et al.* [20] and Sun *et al.* [21]. Li *et al.* [22] described a dimeric molecule that interacts simultaneously with the BIR2 and BIR3 domains of XIAP. Finally, Nikolovska-Coleska *et al.* [23] screened 8,221 natural compounds from traditional Chinese medicinal herbs, finding five active molecules. One of them, called embelin, may represent the first promising non-peptidic lead compound that targets the BIR3 domain of XIAP. Recently, some modifications of embelin have been reported, including a molecule that is more active than the original compound [24].

These results show that the design of Smac/DIABLO mimics is a difficult and interesting challenge for structure-based drug design and high-throughput screening methods.

Survivin, the smallest IAP protein, is also related to various tumours [25], such as lung, breast, colon and stomach tumours, leukaemia or melanoma and its specific expression in damaged tissues makes it a promising target. In addition, the X-ray and NMR structures of Survivin have been solved [26-28], but no structure for the Survivin complex with Smac/DIABLO has been reported to date, although its existence has been proved [27-29]. Up to now, no small molecule inhibitors of Survivin have been reported. However, a comparative study of the binding affinity of various pentapeptides, derived from the five amino terminal residues of Smac/DIABLO, to the BIR domain of Survivin and to the BIR3 domain of XIAP has been reported [27]. Although NMR data suggest a similar binding mode of these peptides to both receptors, the best reported binding affinity for Survivin is 6 M, while single-digit nanomolar peptides were identified for XIAP.

Protein–protein recognition is essential in most biological processes. With greater computer resources, molecular modeling has become an alternative way of treating protein-protein interactions at atomic level. Our aim in this study is to provide insights into the disruption and stability of these apoptotic complexes.

3. METHODS

All the calculations described in the present study were carried out at the molecular mechanics level, using the parm94 force field [30], as implemented within the AMBER7 suite of programs [31]. The solvent was considered explicitly and the cut-off distance was kept to 9 Å to compute the non-bonded interactions. All simulations were carried out under periodic boundary conditions. Long-range electrostatic interactions were treated using the particle-mesh-Ewald method (PME) [32]. The cationic dummy approach [33] was used for the treatment of the zinc atom present in XIAP and Survivin proteins. Zinc and four dummy atoms were used to impose the tetrahedral orientation required for the coordinated residues. The zinc atom is bound covalently to the dummies and interacts with the protein only through van der Waals forces, while the dummies interact with the protein only through electrostatic forces. The force field parameters required for this construction are detailed in the above-mentioned reference.

This simple approach was employed with success in farnesyltransferase [33], metalloproteinase [34], phosphotriesterase [35] and beta-lactamase [36] and solves the problem of maintaining the tetrahedral coordination of zinc throughout a molecular dynamics simulation without a loss of protein flexibility.

All the figures in the present study were prepared with the VMD graphics program [37], with the exception of Figure 6, for which Chimera software [38] was used. The electrostatic potential shown in Figure 6 was calculated with the Delphi II package [39].

3.1. Construction of Smac/DIABLO-XIAP complex

The initial 3D structure of the human Smac/DIABLO-XIAP complex was taken from the Protein Data Bank (PDB entry 1G3F). The complex is formed by the BIR3 domain of XIAP (residues 240 to 357) and nine amino terminal residues of Smac/DIABLO (residues 1 to 9). Dummies were added to the zinc atom and the four residues of XIAP that coordinate this atom (C327, C300, C303 and H320) were deprotonated, according to the experimental NMR structure. A cubic box of 8,243 TIP3P [40] water molecules was added to the system.

The system was neutralized during the simulation by using a uniform plasma, as implemented in the

AMBER7 package.

A visual representation of the Smac/DIABLO-XIAP interface is shown in Figure 2A.

3.2. Construction of Smac/DIABLO-Survivin complex

The initial 3D structure of the human Survivin dimer was taken from the Protein Data Bank (PDB entry 1E31). The protein consists of a BIR-like domain and a long alpha helix domain (residues 5 to 140 in subunit A and 5 to 142 in subunit B). Cobalt atoms were removed and, similarly to the XIAP receptor, dummies were added to zinc and the four coordinated residues (C57, C60, C84 and H77) were deprotonated. Then, the Smac/DIABLO protein was placed into the BIR-like domain of each subunit according to a XIAP(BIR3)-Survivin(BIR-like) sequence alignment (Figure 3A). To do this, the backbone atoms of five conserved residues of XIAP (L307, W310, P312, P316 and E318) were superimposed on the backbone atoms of the equivalent five residues in Survivin (L64, W67, P69, P73 and E75). Thus, the coordinates of the nine residues of Smac/DIABLO in the experimental Smac/DIABLO-XIAP complex were adapted to the BIR-like domain of Survivin. Figure 3B shows a three-dimensional representation of the central residues of the BIR3 domain of XIAP and the equivalent residues for the Survivin protein. Superimposed residues were selected in order to force a good description of the binding pocket. After this procedure, a cubic box of 27,125 TIP3P [40] waters was added and the system was neutralized, as described. A visual representation of the Smac/DIABLO-Survivin interface is shown in Figure 2B.

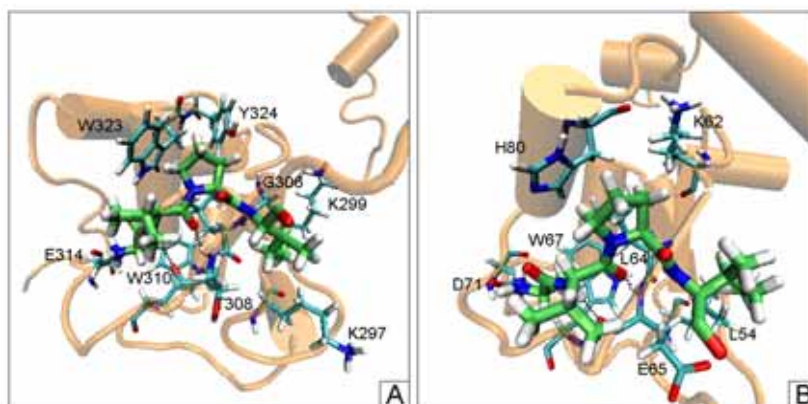


Figure 2: Most important interactions of the Smac/DIABLO-XIAP complex (A) and the Smac/DIABLO-Survivin complex (B). For visual simplicity, only the first four residues of Smac/DIABLO are shown and carbon atoms of this sequence are remarked in light green.

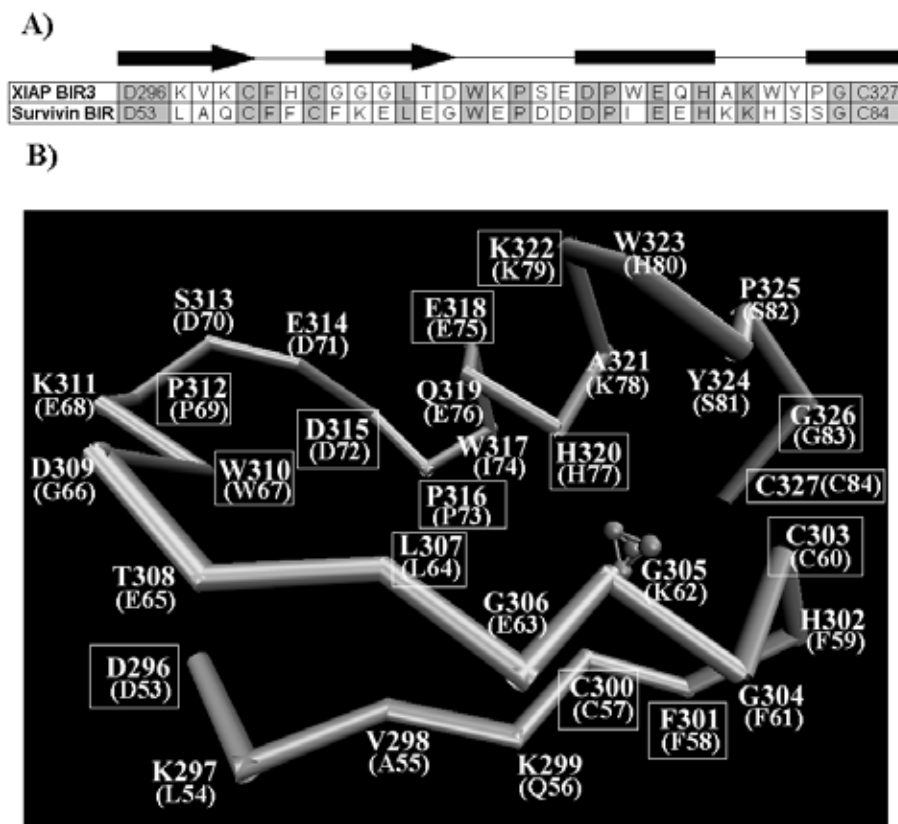


Figure 3: A) Sequence alignment of XIAP BIR 3 and Survivin BIR, dark boxes show conserved residues. B) Visual representation of the central residues in BIR3 domain of XIAP and the sequence alignment with Survivin (in brackets). Conserved residues in both proteins are remarked with a box.

3.3. Minimization

The complexes were energy-minimized in order to remove possible steric stress by a multi-step procedure. First, water molecules were allowed to relax while the rest of the system was kept frozen. Second, side chains of XIAP or Survivin (except the four zinc-coordinated residues) and Smac/DIABLO were relaxed together with water molecules. Third, all atoms except zinc, dummies and the four coordinate residues of XIAP or Survivin were relaxed; and finally all atoms were allowed to

move. The steepest descent method, followed by the conjugated gradient method, were used as minimization algorithms and the final energy gradients, achieved after 100,000 iterations, were 0.6 kcal/mol for the XIAP complex and 2.0 kcal/mol for the Survivin complex, which are reasonable gradients for a local minimum.

3.4. Molecular Dynamics Simulation

Molecular dynamics (MD) simulations of the complexes were performed at a constant temperature of 300 K by coupling the system to a thermal bath using Berendsen's algorithm [41], as implemented in AMBER7 [31], with a time coupling constant of 0.2 ps. Time step was set to 1 fs, and the list of nearest neighbour atoms was updated every 15 steps. A cut-off of 9 Å for non-bonded interactions was used, constraining the length of bonds involving hydrogen atoms by means of the SHAKE algorithm [42] to achieve a rapid energy convergence.

MD simulation began by heating the minimized structure to 300 K over a period of 100 ps at a constant rate of 30 K/10 ps with the protein atoms being kept frozen. The second step was a 40 ps pressure-constant period to increase the density of the system, in which protein atoms were restricted in movement. The third step involved a 150 ps volume-constant period with only the zinc, dummies and the four coordinated residues of XIAP or Survivin kept frozen. Finally, a 1 ns dynamics calculation was performed, in the canonical ensemble, with no restrictions on the system and a constant temperature of 300 K.

3.5. Free energy estimations using the MMGBSA approach

Theoretical binding free energies were calculated within the one-trajectory protocol using the MMGBSA (Molecular Mechanics Generalized Born Surface Area) approach [43], described as:

$$G_{binding} = G_{complex} - (G_{receptor} + G_{ligand})$$

$$G_{molecule} = \langle E_{MM} \rangle + \langle G_{solvation}^{polar} \rangle + \langle G_{solvation}^{nonpolar} \rangle - TS$$

$$\langle E_{MM} \rangle = \langle E_{internal} \rangle + \langle E_{electrostatic} \rangle + \langle E_{vdw} \rangle$$

$$G_{solvation}^{nonpolar} = \gamma SASA$$

This methodology was applied to 100 snapshots extracted from the production time of the molecular dynamics for each system studied. $\langle \rangle$ denotes the average value of the selected set of structures along the molecular dynamics trajectory. Internal energy ($E_{internal}$) includes bond, angle and dihedral energies, and $E_{electrostatic}$ and E_{vdW} are intermolecular electrostatic and van der Waals energies, respectively. All these terms are calculated by molecular mechanics *in vacuo*. Electrostatic contribution to the free energy of solvation ($G_{solvation}^{polar}$) was calculated by a continuum representation of the solvent within the Generalized Born model, with the Tsui *et al.* [44] atom parameter set and the MEAD program [45]. External dielectric constant was set to 80.0, while internal dielectric constant was set to 1.0. Non-polar contribution to solvation ($G_{solvation}^{nonpolar}$) was obtained by using a simple linear relationship with the SASA (Solvent Accessible Surface Area), where γ gets the value of 0.0072 kcal/molÅ², and employing the LCPO method [46]. Finally, entropic effects were computed with a normal mode analysis, as implemented in the NMODE module of AMBER7 software package.

3.6. Docking and Post-Docking Methodology

Two docking programs were used to determine the binding mode of embelin and its derivatives. The use of different docking programs is required in order to avoid differences in conformation generation and scoring function, which can be translated in obtaining different binding poses for the same ligand.

The first program was Dock_Dyn [47], our home-made pharmacophore-directed program of docking. It takes as input a reference pharmacophore obtained by monitoring protein-protein interactions along a molecular dynamics. Specifically, we take those molecular features or fragments belonging to the Smac/DIABLO protein whose interaction with the XIAP protein persists throughout the simulation. Pharmacophore coordinates were defined using their mean positions and their maximum and minimum deviations during the production time. Once the reference pharmacophore is characterized, pharmacophoric points are assigned to the ligands and conformational flexibility is introduced by rotating all their dihedral angles. For each of the compounds studied, a different number of interaction points, belonging to the Smac/DIABLO-XIAP pharmacophore, was used. Finally, the program minimizes the RMS (root mean square) between the selected reference pharmacophoric points and all the possible combinations of equivalent pharmacophoric points assigned to the ligand. This value is

used as a scoring function and thus the selected poses are those that best satisfy the restrictions imposed by the reference pharmacophore. This methodology is especially adapted to obtain peptidomimetics, i.e. small molecules that simulate peptide-protein interactions. Thus, best poses are selected by means of a geometrical criterion.

To avoid any possible bias introduced by our pharmacophore-directed program, GLIDE [48], a general non-pharmacophore-dependent docking of Schrödinger suite (Maestro version 80110), was also used. All docking calculations were performed using the XP (or extra-precision) mode. Ligands were prepared using the Ligprep application with the OPLS-2005 force field [49]. The ten best poses generated, selected by the GlideScore function, were always retained for visual analysis.

Both programs use the NMR structure of XIAP as a rigid receptor and a flexible ligand approach.

Embelin and embelin derivatives were minimized at AM1 level using the Gaussian03 package [50] before the docking procedure, to assure a correct initial structure.

Once the best poses from both methods were identified, a post-docking process was carried out to introduce receptor flexibility and evaluate the binding free energies of each structure. Thus, different minimization steps and a 2 ns MD simulation were performed, using the protocol cited previously for the case of Survivin and XIAP complexed with Smac/DIABLO. The force field parameters, needed to perform the MD simulation, for embelin and its derivatives were obtained within the general amber force field (GAFF) [51]. Finally, after total energy was stabilized (around the last 500 ps), an MMGBSA protocol calculated the binding free energy for the different binding modes.

4. RESULTS AND DISCUSSION

4.1 Convergence Analysis

Smac/DIABLO-XIAP and Smac/DIABLO-Survivin complexes were equilibrated after the first 100 ps of molecular dynamics. This rapid convergence is usually ascribed to the SHAKE algorithm and to the use of a good initial minimized structure for the MD simulation. Accordingly, the last 900 ps were considered as production time and used for the analysis of interactions.

RMS convergence was also achieved rapidly during the production time, with reduced fluctuations as

an indication of low structural protein mobility. Average RMS deviations for the first four residues of Smac/DIABLO were 0.4 Å and 0.2 Å for the XIAP complex (Figure 4A) in relation to the experimental and minimized structures, respectively, and 0.7 Å for the minimized Survivin complex. The initial Smac/DIABLO-Survivin complex was not taken into account in this analysis because the superimposition model needs a minimization step. Finally, average backbone RMS deviations, in relation to the experimental and minimized receptor structures, were 1.8 Å and 1.7 Å (Figure 4B) for the XIAP protein and 1.0 Å for the Survivin protein, in relation to the minimized system.

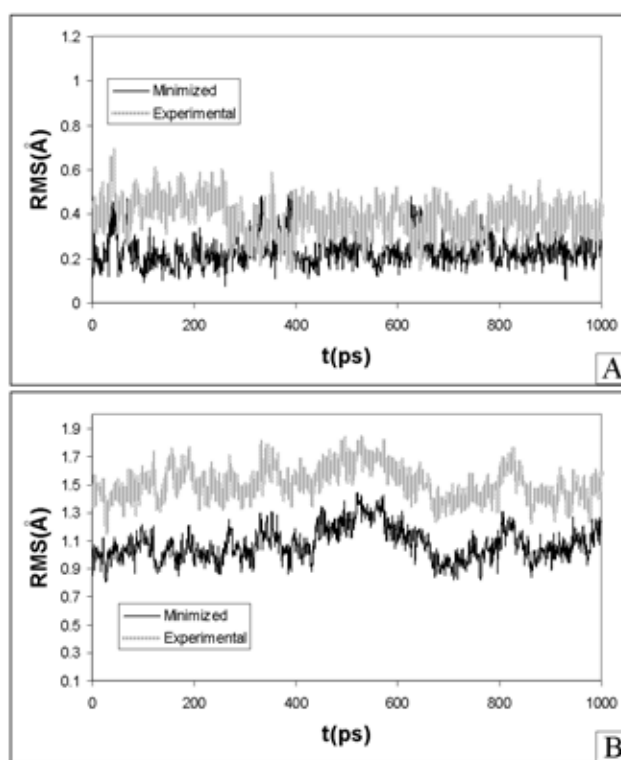


Figure 4: A) Evolution of the RMS deviation, for backbone atoms, of the first four residues (AVPI) of Smac/DIABLO. B) Evolution of the RMS deviation, for backbone atoms, of residues 299 to 329 of XIAP. RMS deviation with respect to the experimental structure (PDB code 1G3F) is shown in dashed lines and RMS deviation with respect to the minimized system is shown in black.

When one focuses on the C-terminal part of Smac/DIABLO, the RMS deviation does not converge due to the lack of interactions with the receptors and the consequent high mobility of this zone. This is a

first test to confirm that only the initial AVPI sequence of Smac/DIABLO interacts with both receptors. This result suggests that the AVPI sequence can be seen as a conserved tetrapeptide motif. In fact, Caspase-9, which competes with Smac/DIABLO for the same XIAP binding pocket, shows the highly similar sequences of ATPF in human and AVPY in mouse.

4.2 Zinc-ligand Analysis

To confirm the correct behaviour of the cationic dummy approach to describe the XIAP protein, the zinc-ligand structure was analysed. Table 2 lists the experimental and average molecular dynamics zinc-coordinated atom distances. All four distances confirm the experimental NMR structure, and the small fluctuation (around 0.1 Å) should be due to the flexibility of the BIR3 domain. As a first conclusion, the simple model employed for the description of this zone showed good results. Table 2 also shows the same distances for the Survivin complex.

	Experimental distance / Å	Average simulation distance / Å	RMS / Å
XIAP			
Zn – S (C300)	2.10	2.12	0.04
Zn – S (C303)	2.10	2.13	0.04
Zn – S (C327)	2.10	2.14	0.04
Zn – N (H320)	2.22	2.07	0.05
Survivin			
Zn – S (C57)	2.31	2.12	0.04
Zn – S (C60)	2.32	2.13	0.04
Zn – S (C84)	2.39	2.13	0.04
Zn – N (H77)	2.26	2.04	0.04

Table 2: Average simulation distances and experimental distances of the zinc-coordinated atoms in XIAP and Survivin.

Theoretical zinc-ligand distances for XIAP and Survivin are very similar. However, experimental distances are slightly different. These differences could be attributed to experimental distances for Survivin, unlike XIAP, being obtained from the crystal structure of Survivin alone. Then, the presence of Smac/DIABLO residues could introduce some modifications that are taken into account in our MD simulations.

4.3. Hydrogen Bond Analysis

The CARNAL program of AMBER7 package was used to identify the intermolecular hydrogen bonds of the complexes. We found only three hydrogen bonds with optimum geometric parameters in the XIAP complex, and only two in the Survivin complex, during the production time of the MD simulations. Tables 3 and 4 list the relevant data of distances and angles for the hydrogen bonds. The hydrogen bond pattern is equivalent in both receptors except for the last hydrogen bond, established between residue G306 of XIAP and residue I4 of Smac/DIABLO. There is an equivalent residue in Survivin (E63 with I4), which showed non-typical geometrical parameters, i.e. an average hydrogen bond distance of 2.56 Å and an average angle of only 105.6°.

These results corroborate the analysis of the experimental structure [8-9] except for two additional hydrogen bonds shown in Table 1, W323 with A1 and Q319 with A1, which do not show typical parameters for hydrogen bond distances or angles through the simulation. In fact, even for the NMR structure, the structural parameters of these hydrogen bonds are not optimum. Thus, the hydrogen bond distance with W323 is 2.31 Å but with an angle of only 133.0° and the hydrogen bond distances with the Q319 side chain amide hydrogens are 3.88 Å and 3.91 Å, with angles of only 83.1° and 81.5°, respectively.

The positively charged amino terminal group of Smac/DIABLO forms several salt bridges with close negatively charged residues, E314 and D309 in the XIAP complex, and, D71 and E76 in the Survivin complex. We prefer to consider these interactions as electrostatic, rather than genuine, hydrogen bonds.

XIAP	Smac/DIABLO	Distance / Å	RMS / Å	Angle / °	RMS / °	% ¹ Occupation
T308 NH	OC V2	1.93	0.13	159.3	10.83	83.78
T308 CO	HN V2	1.97	0.19	150.9	14.12	64.33
G306 CO	HN I4	2.17	0.31	155.5	12.13	59.33

Table 3: Average data of hydrogen bonds throughout the production time of MD simulation in the XIAP complex. ¹Occupation: Represents the per cent of simulation time that the hydrogen bond is optimum (maximum distances between N and O of 3.3 Å and hydrogen bond angle N-H ... O of 180 ± 20°).

Survivin	Smac/DIABLO	Distance / Å	RMS / Å	Angle / °	RMS / °	% ¹ Occupation
E65 NH	OC V2	2.07	0.17	161.5	8.52	85.03
E65 CO	HN V2	1.92	0.16	153.6	13.70	72.06

Table 4: Average data of hydrogen bonds throughout the production time of MD simulation in the Survivin complex. ¹Occupation: Represents the per cent of simulation time that the hydrogen bond is optimum (maximum distances between N and O of 3.3 Å and hydrogen bond angle N-H ... O of 180 ± 20°).

4.4. Van der Waals interaction Analysis

The ANAL program of AMBER7 package was used to identify van der Waals interactions. The van der Waals interaction energy between each residue of Smac/DIABLO and XIAP or Survivin was analysed throughout the production time to find the most favourable interaction. Average energies calculated with 9 snapshots (1 structure each 100 ps) of MD simulation are summarized in Figure 5A. No cut-off was used.

There is a clear difference between the first four residues (AVPI) and the remaining sequence, showing that only this sequence interacts with the receptors, as experimental data have previously suggested. The analysis was expanded by calculating the van der Waals energy between different residue pairs in the XIAP complex, as suggested by the experimental structure [8-9]. The most important contacts were A1 with W310 (average energy -2.6 kcal/mol), A1 with L307 (average energy -1.2 kcal/mol), A1 with Q319 (average energy -1.2 kcal/mol), V2 with T308 (average energy -0.9 kcal/mol), V2 with W323

(average energy -1.9 kcal/mol), P3 with W323 (average energy -3.1 kcal/mol), P3 with Y324 (average energy -1.2 kcal/mol) and I4 with K297 to K299 (average energy -2.9 kcal/mol). One can see that W310 and W323 are the most important hydrophobic residues of the BIR3 domain of XIAP. The importance of these two residues was noticeable through the W323A and W310A mutations [8]. Thus, for the Smac/DIABLO-XIAP complex the binding affinity (K_D) is 0.74 M, whereas for the XIAP mutant complexes W323A and W310A affinities are 42 M and up to $1,000$ M, respectively. Some of the most important synthetic modifications of the tetrapeptide AVPI focus on the enlargement of the hydrophobic interactions of residues P3 and I4 [19].

Similar to the XIAP complex, the average energy of the equivalent residues interacting with Survivin was calculated (see Figure 3A for alignment). The most important contacts were A1 with W67 (average energy -2.6 kcal/mol), A1 with L64 (average energy -1.4 kcal/mol), A1 with E76 (average energy -0.6 kcal/mol), V2 with E65 (average energy -0.3 kcal/mol), V2 with H80 (average energy -0.9 kcal/mol), P3 with H80 (average energy -2.3 kcal/mol), P3 with S81 (average energy -0.1 kcal/mol) and I4 with L54 to Q56 (average energy -1.3 kcal/mol). It is worth commenting that Y324 of XIAP has a non-hydrophobic residue, S81, equivalent in Survivin. Thus the important hydrophobic interaction with P3 of Smac/DIABLO is lost, with an interaction energy decreased to only -0.1 kcal/mol. Also, residue W323 of XIAP is replaced by residue H80 in Survivin, producing a noticeable decrease of hydrophobic interactions with V2 and P3 residues of Smac/DIABLO. Moreover, the XIAP hydrophobic groove formed by carbon side chains of K297 to K299 has no equivalent in the Survivin domain. Thus, the hydrophobic interaction with I4 of Smac/DIABLO is decreased by 1.6 kcal/mol.

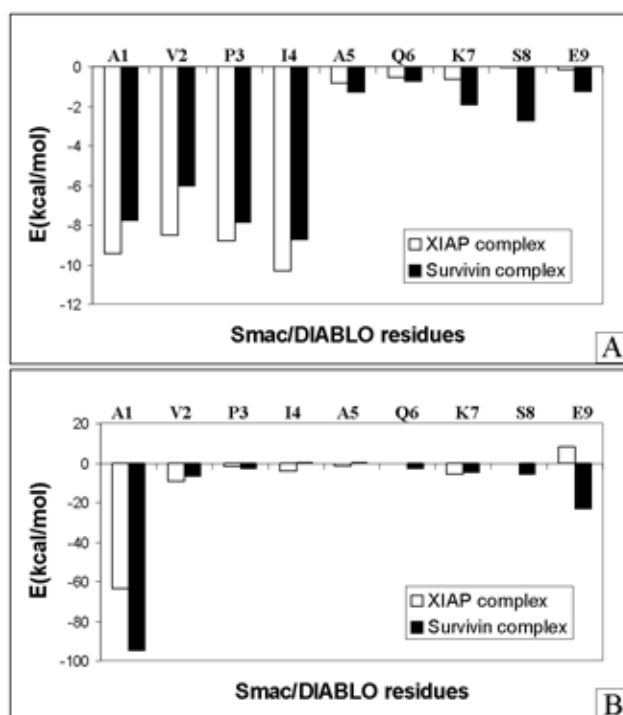


Figure 5: A) Average van der Waals energy, taking 9 snapshots, of each residue of Smac/DIABLO extracted from the production time. B) Average electrostatic energy, taking 9 snapshots, of each residue of Smac/DIABLO extracted from the production time.

Concluding these analyses, the first four residues of Smac/DIABLO are all hydrophobic and the BIR3 XIAP and BIR-like Survivin domain show fit pockets for the side chains of these residues. Thus, van der Waals recognition in these complexes seems to be particularly important. Moreover, it is worth noting that many hydrophobic interactions are lost when going from XIAP to Survivin, which confirms its relative binding affinity with the Smac/DIABLO peptide [27].

4.5. Electrostatic interaction Analysis

Electrostatic interactions were calculated similarly to van der Waals interactions, using the ANAL program of AMBER7 package throughout the production time. Interactions for each residue of

Smac/DIABLO with respect to XIAP and Survivin were calculated. The average energies of 9 snapshots (1 structure each 100 ps) are shown in Figure 5B. No cut-off was used. As can be seen, only the first residue has marked electrostatic energy, due to the positive charge of the amino terminal atom of the peptide. The high value obtained for this contribution reveals its important role in protein recognition, even more important for the Survivin complex because the BIR-like domain of Survivin has a higher number of negatively charged residues (E63, E65, E68, D70, D71, D72, E76) than the BIR3 domain of XIAP.

For the Smac/DIABLO-XIAP complex, the electrostatic energy between the first residue of Smac/DIABLO and two close charged residues of XIAP, E314 and D309, was also calculated. The average energies obtained from the simulation were -33.0 and -18.8 kcal/mol, respectively. This result shows that most of the electrostatic interaction of the BIR3 domain is due to these two residues. Only the first interaction, between the first residue of Smac/DIABLO and E314 of XIAP, was suggested by the analysis of the experimental structure. In addition, this was later confirmed by the reduction in activity of the E314S mutation [8]. Similar results were obtained for the Smac/DIABLO-Survivin complex, with D71 (-33.8 kcal/mol) and E76 (-25.2 kcal/mol) being the most important residues. D71 is necessary for Smac/DIABLO binding to Survivin [29]. To complement this discussion, Figure 6 gives a surface coloured by the electrostatic potential of both receptors that shows the much more negative electrostatic potential of the Survivin BIR domain.

The Smac/DIABLO 9 residues peptide can be seen as a molecule directed by its dipolar momentum, as this electrostatic analysis suggests. This may be the reason for the greater affinity of Smac/DIABLO 9 residues than full-length Smac/DIABLO [8], for which dipolar momentum can be less pronounced.

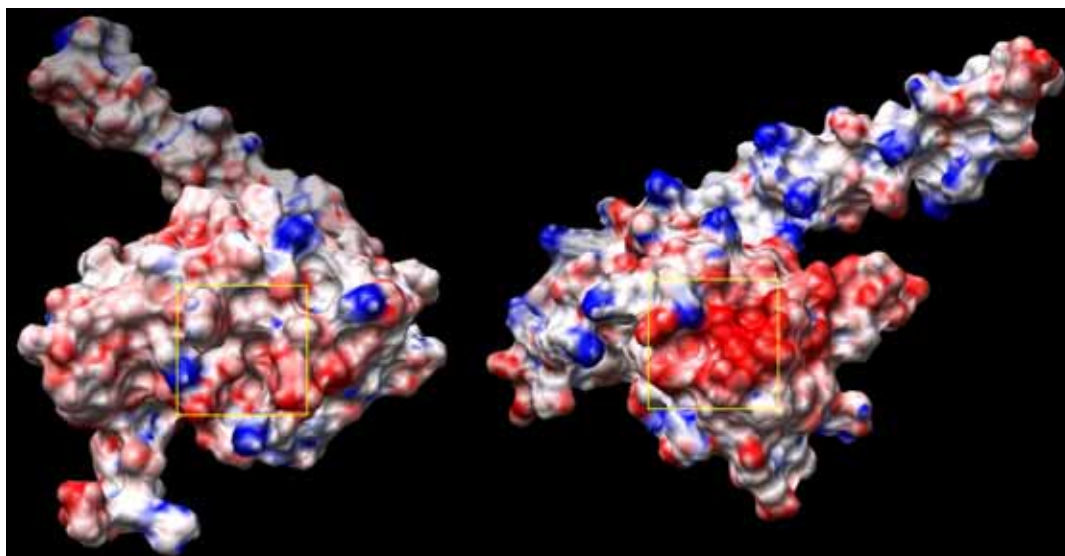


Figure 6: Electrostatic surface potential of XIAP (left) and Survivin (right). Electronegative potential is coloured red, electropositive potential is coloured blue and neutral potential is coloured white. BIR domain of both proteins is remarked with a box.

4.6. Binding free energy Analysis

To complement the interaction analysis performed for both complexes, we also used the MMGBSA protocol to calculate the binding free energy of the Smac/DIABLO peptide bound to the XIAP and Survivin proteins. Table 5 shows (in kcal/mol units) the relevant data from these analyses.

As we can see, the van der Waals balance upon binding (ΔE_{VDW}) and the non-polar contribution to solvation (ΔG_{SUR}) are similar in both systems.

This corroborates the analysis of interactions for both the complexes described above, taking into account that a similar recognition pattern was found between the Smac/DIABLO peptide and both receptors.

Actually the most important difference between both receptors is the electrostatic balance (ΔE_{ELE}), which is larger for Survivin (as can also be inferred from Figure 6). This is translated to a higher desolvation penalty (ΔG_{GB}), which makes the Smac/DIABLO-Survivin complex less stable than the Smac/DIABLO-XIAP complex. After the inclusion of the entropic contribution to binding free energy, which is very similar in both systems, MMGBSA results predict a difference in binding of 2.29

kcal/mol, in corroboration of the experimental affinity difference of 2.39 kcal/mol. Due to intrinsic limitations of the method [52], absolute binding free energies are overestimated.

	<i>Smac/DIABLO-XIAP</i>	<i>Smac/DIABLO-Survivin</i>
ΔE_{ELE}	-179.57	-373.13
ΔE_{VDW}	-36.29	-40.41
ΔG_{SUR}	-4.73	-5.88
ΔG_{GB}	177.08	378.91
ΔG_{SOL}	172.35	373.03
ΔG_{ELE}	-2.49	5.78
ΔG_{TOT}	-43.50	-40.51
$-T\Delta S$	25.83	25.13
ΔG_{MMGBSA}	-17.67	-15.38
ΔG_{EXP}	-8.78 ^a	-6.39 ^b

Table 5: Binding free energy, averaged over 100 snapshots, for the studied complexes. All values are shown in kcal/mol units. ΔE_{ELE} and ΔE_{VDW} account for the electrostatic and van der Waals in vacuo binding enthalpic contribution. ΔG_{SUR} accounts for the non polar contribution to solvation, ΔG_{GB} is the polar contribution to solvation. ΔG_{SOL} denotes for the $\Delta G_{\text{SUR}} + \Delta G_{\text{GB}}$. ΔG_{ELE} accounts for the $\Delta E_{\text{ELE}} + \Delta G_{\text{GB}}$ addition. ΔG_{TOT} accounts for the sum of polar and nonpolar solvation contribution to the binding free energy. $-T\Delta S$ accounts for the entropic balance (truncating the receptor to only those atoms within a cutoff of 12 Å from the ligand and using 10 snapshots). Final theoretical free energy of binding at 300 K is denoted as ΔG_{MMGBSA} while the experimental one is denoted as ΔG_{EXP} . a) From the affinity constant of Liu et al. [8]. b) From the affinity constant of Sun et al., [27].

4.7. Pharmacophore Analysis

Grouping together all the conclusions from the last analyses, we propose a pharmacophore for the ligand in the Smac/DIABLO-XIAP complex formed by eight interaction points (Figure 7). These points are one electrostatic interaction at amino terminal atom, three hydrogen bonds and four van der Waals zones corresponding to the side chains of the AVPI sequence. Regarding the Survivin complex, the pharmacophore does not show the hydrogen donor of point 7 of Figure 7. This could contribute to the

explanation of the lesser affinity of Smac/DIABLO in complex with Survivin suggested previously [27].

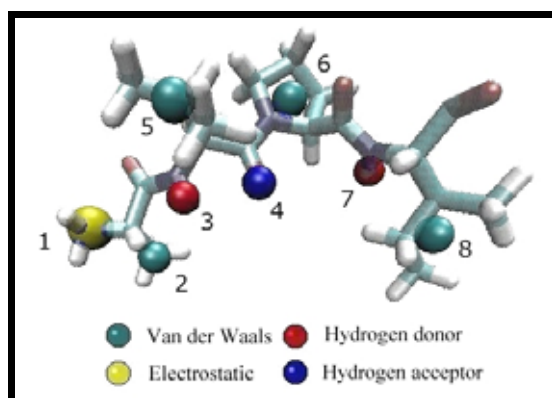


Figure 7: *Pharmacophore of Smac/DIABLO.* Points 5, 6 and 8 are assigned to the center of mass of the side chains.

It is important to note that all hydrogen bonds involved in the proposed pharmacophore belong to peptide bonds. Thus, they cannot be used to explain the binding affinity of different tetrapeptides to XIAP. In addition, as peptide bonds interact only with the carbonyl group at point 4 or only with the amino group at points 3 and 7, a complete peptide bond is never used. Thus, a peptidomimetic molecule of Smac/DIABLO does not need this type of bond. This is interesting because peptide bonds are enzymatically hydrolyzed before a drug can reach its therapeutic target.

On the basis of only the number of pharmacophoric points obtained for each system, the possibility of designing selective inhibitors of XIAP and Survivin seems, in principle, a difficult task. However, all the interaction analyses we have carried out, including the more quantitative results of the MMGBSA methodology, indicate a better binding of Smac/DIABLO to XIAP than to Survivin. This conclusion fully confirms the only experimental results reported that compare the relative affinity of different peptides to the BIR3 domain of XIAP and to the BIR domain of Survivin [27]. Sun *et al.* [27] found that the binding affinity of the AVPFY peptide, taken as an example, is 7 M to Survivin, but is 0.06 M to BIR3 of XIAP, showing that this peptide, which satisfies our derived pharmacophore, binds to XIAP and also, though less strongly, to Survivin. It is clear that a pharmacophoric description can describe compounds with M activity, but does not distinguish them at a more quantitative level. In fact, the

AVPIF peptide, which only has a different hydrophobic residue at the fourth position, has binding affinity values of 60 M and 0.36 M to Survivin and to the BIR3 domain of XIAP, respectively. Reported attempts [27] to improve the binding affinity of the AVPFY peptide for Survivin, through mono-substitutions with hydrophobic, natural or non-natural amino acids, have failed, indicating that these types of changes do not compensate for the desolvation penalty of the binding process. To the best of our knowledge, no experimental evidence indicating that replacements of hydrophobic residues by charged amino acids could improve the best binding affinity that has been reported.

However, the suitability of obtaining a selective inhibitor in front of a pan-IAP inhibitor is now a matter of discussion [53]. Therefore, it is clear that a compound able to interact with both proteins would be of great interest.

Regarding pharmacophore description and applicability, we recently studied six tetrapeptides based on modifications of the Smac/DIABLO amino terminal sequence and they exhibited the same recognition pattern [54]. In addition, other peptidomimetic molecules bind to the BIR3 domain of XIAP, maintaining a similar binding mode [18-22]. These facts suggest that our proposed pharmacophore can be used to search for small molecules that are mimics of the Smac/DIABLO protein. In this sense, our future work will be based on virtual screening of databases with available organic compounds, which may give some of the most important interaction points.

Recently, a non-peptidic compound, named embelin, and some derivatives of it have been reported as new XIAP inhibitors [24]. Unlike the other inhibitors reported, these compounds were not discovered following a procedure designed to obtain Smac/DIABLO mimetics. As such, knowledge of their binding modes may contribute to more insights into XIAP inhibition. Figure 8 shows embelin and two embelin derivatives selected in this study. Compound 1 has the lowest activity of all the embelin derivatives (K_i of 10.4 M); embelin or compound 2 has moderate activity (K_i of 0.40 M); whilst compound 3 has the highest activity (K_i of 0.18 M).

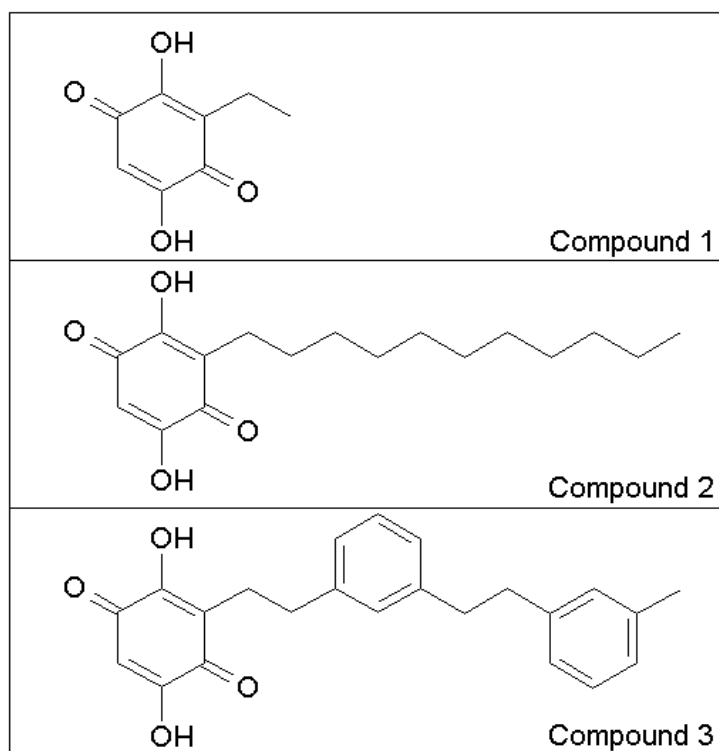


Figure 8: Structures of embelin (compound 2) and selected embelin derivatives (compounds 1 and 3).

Accordingly, we used docking together with MD simulations and binding free energy calculations for the protein-ligand complexes, to propose the binding mode of embelin and its derivatives shown in Figure 8 and to correlate their experimental affinities.

4.8. Docking and MD simulations of embelin and embelin derivatives

A docking procedure, as commented in Methods, was performed for embelin and two of its derivatives. The first docked compound was compound 1 (Figure 8), which consists of a benzoquinone ring and an ethyl group.

Figures 9A, 9C, 9E and 9G show the best docked poses using the GLIDE XP program. Pose 9A is the best scored pose, while the pose in Figure 9G is the lowest scored pose.

Since compound 1 is a small molecule, lacking the long hydrophobic tail of embelin, we used only two pharmacophoric constraints to run the Dock_Dyn program. Both points belong to the benzoquinone

ring and correspond to the hydrogen donor point 3 and the hydrogen acceptor point 4 of the pharmacophore. The best pose obtained using this methodology was almost identical to the best scored pose obtained with the GLIDE program (Figure 9A), providing the first evidence that the information acquired from the study of protein-protein recognition could be used as a starting-point in the discovery of non-peptidic compounds containing similar interactions.

Thus, the pose in Figure 9A shows two hydrogen bonds with T308 similar to points 3 and 4 (residue V2 of Smac/DIABLO) of the pharmacophore. In addition, the benzoquinone forms π - π stacking with W323, which was also an important residue for Smac/DIABLO recognition. L307 and Y324 interact with the ethyl group of compound 1 by means of van der Waals forces.

The next poses (Figures 9C, 9E and 9G), ordered by their scores, were selected from the ten best scored poses for visual analysis, in an attempt to retain great diversity.

The pose in Figure 9C is similar to the first pose, but the π - π stacking with W323 seems to be weaker than in the best pose, as does the interaction with Y324 due to the position of the ethyl group.

However, poses in Figures 9E and 9G form two hydrogen bonds, the first with T308, similar to poses in Figures 9A and 9C, and the second with the charged D309. The ethyl group interacts with W323 only in the pose in Figure 9E and L307 maintains van der Waals contacts with the benzoquinone ring of both poses.

Because Dock_Dyn and GLIDE only include in the docking process ligand flexibility, the four poses were used as a starting point to perform molecular dynamics simulations of the complexes, thus introducing protein flexibility too. The combined use of docking and molecular dynamics simulations is a common methodology in drug design [55].

Figures 9B, 9D, 9F and 9H show the final structures of the complexes after the MD simulation. The first Figure, 9B, shows both hydrogen bonds lost: the interaction is mainly established by W323, decreasing the interactions with L307 and Y324.

However, the structure in Figure 9D showed a similar disposition to the docking pose in Figure 9A, remaining almost unaltered throughout the MD simulation. In fact, π - π stacking with W323 was optimized, the two hydrogen bonds with T308 were maintained and the good van der Waals contact between the ethyl group and L307 was also unaltered. A detailed analysis of this structure shows that the most important difference between Figures 9A and 9D is a bond rotation within the ethyl group.

It seems that the structure in Figure 9D maintains better binding after the MD simulation than the pose

in Figure 9A, as this latter pose seems a local minimum, reached due to a constrained initial position of the ethyl group that cannot be well positioned without protein reorganization. In conclusion, although both docking methods located the same best pose, the description of the best structure can only be achieved by the inclusion of protein flexibility.

Moreover, the good binding description and the low fluctuation observed throughout this simulation gives confidence to the second final structure (Figure 9D), which only needs a small reorientation of both the ligand and the protein during the MD simulation.

Figure 9F shows the structure after the MD simulation of pose in Figure 9E. Both hydrogen bonds were lost and the general binding interactions decreased, showing only a subtle van der Waals contact with W323.

Finally, the pose in Figure 9G was reoriented to increase the interaction between the ethyl group and the W323 residue, while the hydrogen bonds with T308 and D309 were stable during the MD simulation.

Given each recognition pattern, the pose optimized through the MD simulation seems the best structure (Figure 9D). This conclusion will be supported in the next section by adding the more quantitative MMGBSA results of binding free energy.

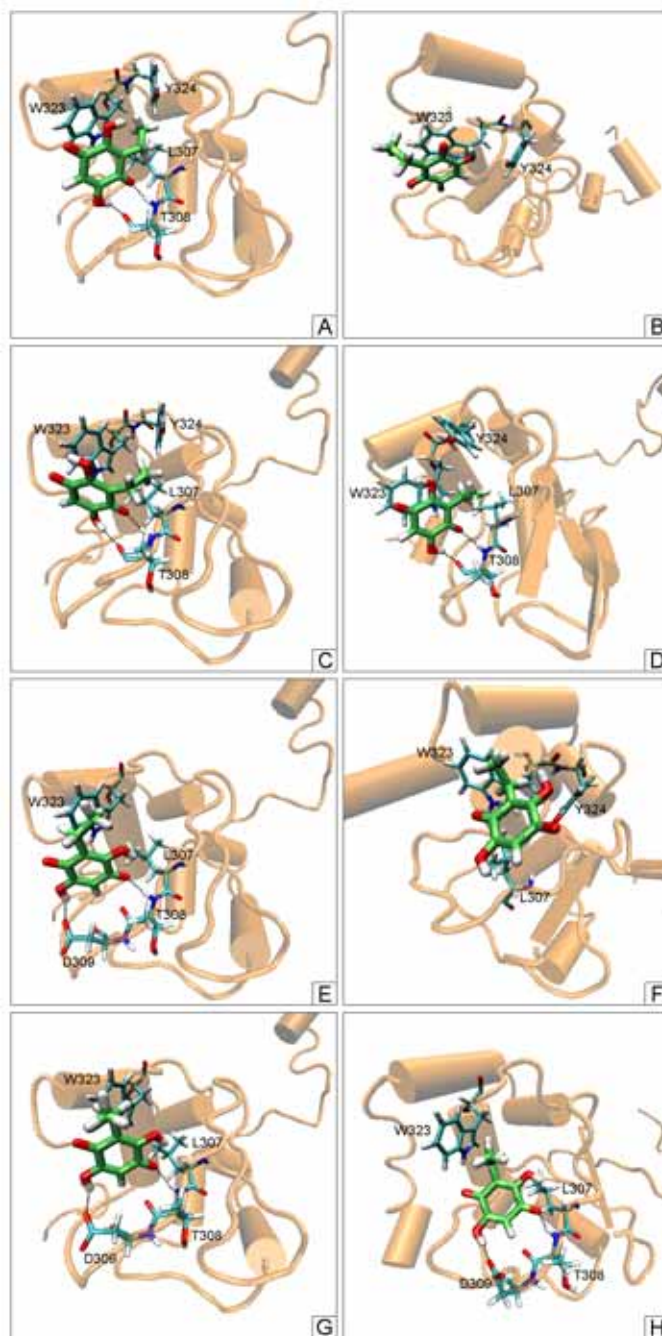


Figure 9: Docking poses (left) and Molecular Dynamics structures (right) for the compound 1-XIAP complex. Carbon atoms of the ligand are remarked in light green. A) Best *GLIDE* pose 1 and best *Dock_Dyn* pose, B) MD structure from the docking pose shown in A, C) *GLIDE* pose 2, D) MD structure from the docking pose shown in C, E) *GLIDE* pose 3, F) MD structure from the docking pose shown in E, G) *GLIDE* pose 4, H) MD structure from the docking pose shown in G.

The second compound studied was embelin (Figure 8), which is composed of the benzoquinone fragment and a long, hydrophobic and flexible tail. Figures 10A, 10C and 10E show the best poses obtained by the docking procedure with both programs.

The first pose, shown in Figure 10A, was obtained with our home-made program by applying 3 pharmacophoric constraints, which can be recognized in embelin. These are points 3, 4 and 8 of the Smac/DIABLO pharmacophore. Hydrophobic point 8 was preferred to dock the molecule instead of point 6, in order to achieve a more extended conformation of the hydrophobic tail. Thus, the compound in this binding mode interacts with T308 forming two hydrogen bonds (like V2 of Smac/DIABLO) and it interacts with K297 and K299 with the aliphatic tail of embelin (like I4 of Smac/DIABLO). L307 and W323 are residues also involved in this binding mode.

The pose in Figure 10A was obtained by attempting to reproduce the maximum number of protein-protein interactions. Thus, we suppose that embelin is a mimetic of the Smac/DIABLO protein and should maintain similar types of interactions with the receptor.

On the other hand, GLIDE's best poses are shown in Figures 10C and 10E, ordered in line with their scores. The former is recognized by forming two hydrogen bonds with E314 and T308, and placing the tail close to the hydrophobic environment of W323 and Y324. As the program does not force any special orientation to the hydrophobic tail, this fragment moves out of the receptor hydrophobic zone defined by the I4 residue of the Smac/DIABLO protein. Finally, Figure 10E shows a completely different orientation, placing the benzoquinone fragment near to K299 and the tail close to L307 and T308, interacting by van der Waals forces. Two hydrogen bonds, with K299 and V298, are also detected.

In addition, Figures 10B, 10D and 10F show the final structures of the complexes after the MD simulation. The initial docked pose, Figure 10A, was modified throughout the MD simulation by moving the benzoquinone head close to W323 and Y324, forming a hydrogen bond with the latter and establishing π - π stacking with W323, but losing the two hydrogen bonds with T308. This behaviour of embelin was similar to the reorientation of compound 1 observed, shown in Figure 9B. Moreover, the experimental NMR spectrum of the embelin-XIAP complex showed that the most prominent changes of XIAP during its recognition were noticed in W323 and Y324 signals [23]. This gives confidence to the binding mode of embelin shown in Figure 10B.

However, our supposition that embelin reproduces the described protein-protein interactions seems, in

some way, contradicted. We can explain the final structure in Figure 10B if we assume that for this compound, interactions of the long hydrophobic tail are more important than the hydrogen bonds found, and then that in the case of embelin, optimization of the hydrophobic contacts is preferred, instead of maintenance of the hydrogen bonds. Thus, pharmacoporic points 3 and 4 are not used in the structure of Figure 10B.

This assumption is also supported by that with this reorganization, embelin obtains a new π - π stacking with W323, maintaining the hydrophobic tail contacts. Since compound 1 has no hydrophobic tail, it achieves the same π - π stacking without losing the hydrogen bonds (Figure 9D), which means that the initial pose is almost unaltered throughout the dynamics simulation.

In addition, the fact that embelin has no positive charge or a methyl group (similar to A1 of Smac/DIABLO) near the atoms involved in the hydrogen bonds contributes to the observed displacement from the initial coordinates, when protein flexibility is introduced during the MD simulation.

The final MD structures from GLIDE's poses are shown in Figures 10D and 10F. Regarding the first structure, the hydrophobic tail of embelin was also changed to a more extended conformation (with less steric stress), but lost important van der Waals interactions with W323 and Y324. Experimental NMR tests confirmed the role of W323 and Y324 in XIAP complexed with embelin [23], so the structure shown in Figure 10D cannot explain the experimental results. In addition, a new hydrogen bond with D209 was found, but two hydrogen bonds were lost. The hydrophobic tail is now in slight contact with the lateral chain of L307, but remains out of the hydrophobic pocket suggested by the Smac/DIABLO residues.

In the MD structure obtained from the pose in Figure 10E, the hydrogen bond with V298 was maintained, but the hydrophobic chain was almost fully solvent-exposed, which is an unfavourable binding feature.

Taking into account the interaction pattern, it seems that the binding mode seen in Figure 10B is the best one. Nevertheless, as we stated before, binding free energy was analysed to find the best binding mode found by this computational protocol. This information is included in the next section.

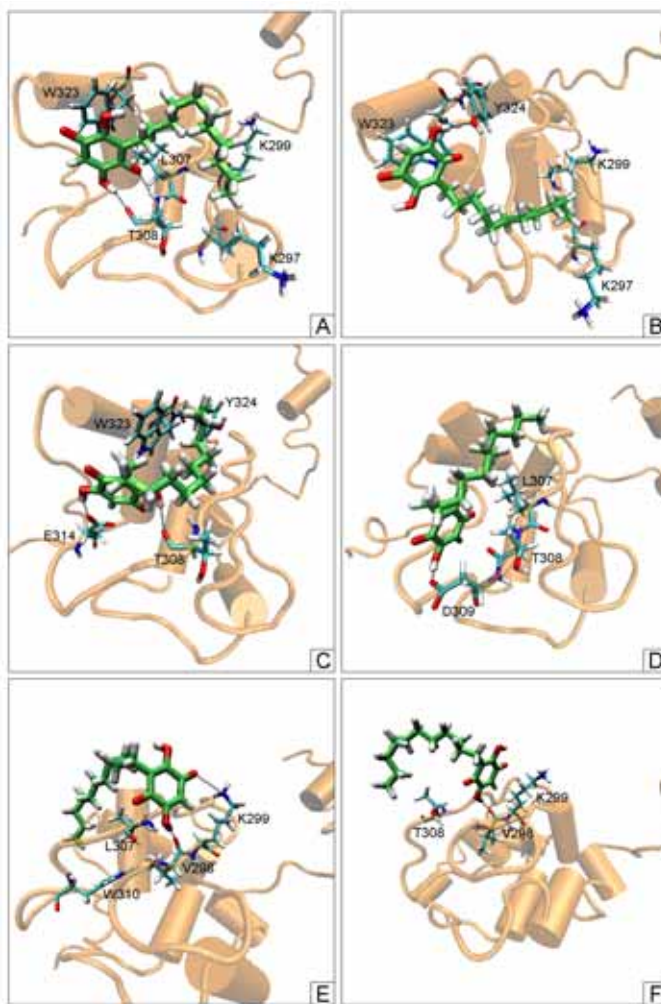


Figure 10: Docking poses (left) and Molecular Dynamics structures (right) for the embelin-XIAP complex. Carbon atoms of the ligand are remarked in light green. A) Best Dock_Dyn pose, B) MD structure from the docking pose shown in A, C) GLIDE pose 1, D) MD structure from the docking pose shown in C, E) GLIDE pose 2, F) MD structure from the docking pose shown in E.

Finally, compound 3 (Figure 8), which is the most active embelin derivative found so far, was also studied. Similarly, Figure 11 shows the best poses from the binding analysis. Figure 11A shows the best pose using our docking program, Dock_Dyn, and applying the pharmacophore constraints of points 3, 4, 6 and 8. Thus, the binding mode of Figure 11A maintains two hydrogen bonds with T308 (similar to

V2 of Smac/DIABLO) and a hydrophobic interaction with K297 and K299 (similar to I4 of Smac/DIABLO). In addition, residues W323, Y324 and L307 interact by means of van der Waals forces with the second ring of this embelin derivative (similar to P3 of Smac/DIABLO).

Figures 11C, 11E and 11G show the best poses obtained by the GLIDE program. The binding modes show different hydrogen bonds, one in pose 1 with T308, two in pose 2 with T308 and E314 and three in pose 3 with D309, Q319 and W323. Regarding van der Waals contacts, the second ring of this embelin derivative interacts with L307 and W323 in the three GLIDE poses, while the terminal ring of the molecule interacts with Y324, although slightly. As happened for embelin, the hydrophobic tail is not located in the hydrophobic pocket suggested by the Smac/DIABLO interactions.

In addition, Figures 11B, 11D, 11F and 11H show the final binding modes after the MD simulation of each initial pose. Figure 11B shows the final binding mode of compound 3 from the best scored pose of our program. It should be noted that the two hydrogen bonds with T308 were maintained for the entire simulation, but the rigid chain is moved close to aromatic residues W323 and Y324, to optimize ring-ring interactions with the terminal benzene ring of this embelin derivative. The second ring continues to interact with L307.

This is different behaviour from embelin, whose polar head moved close to W323 and Y324, losing the hydrogen bonds, and whose hydrophobic chain was in contact with K297 and K299. It seems that as van der Waals contacts direct the protein-ligand interactions in this kind of molecules, the aliphatic chain of embelin interacts tightly with the aliphatic chains of K297 and K299, while the two aromatic rings added in compound 3 interact in a favourable manner with the aromatic residues W323 and Y324, instead of K297 and K299. Since residue W323 is now interacting with the hydrophobic tail of compound 3, the benzoquinone fragment does not have to move away from its initial position to establish π - π stacking with W323. Structures in Figures 11D, 11F and 11H show, in general, a worse interaction pattern. Initial interactions in the docking structures were changed, especially hydrogen bonds, indicating that the GLIDE program does not detect better docking poses. Van der Waals contacts are also reduced and even the final aromatic ring of the derivative is fully solvent-exposed in structures of Figures 11F and 11H.

To complement these analyses, MMGBSA results for each binding mode will be discussed in the next section.

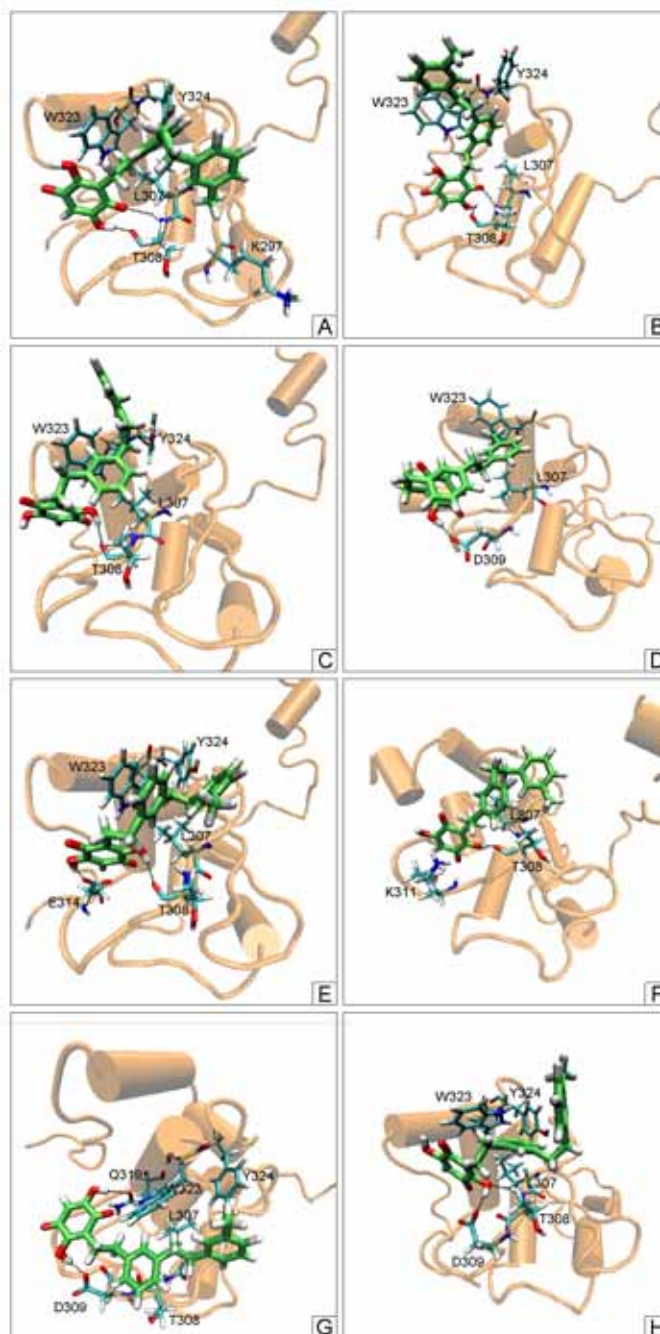


Figure 11: Docking poses (left) and Molecular Dynamics structures (right) for the compound 3-XIAP complex. Carbon atoms of the ligand are remarked in light green. A) Best Dock_Dyn pose, B) MD structure from the docking pose shown in A, C) GLIDE pose 1, D) MD structure from the docking pose shown in C, E) GLIDE pose 2, F) MD structure from the docking pose shown in E. G) GLIDE pose 3, H) MD structure from the docking pose shown in G.

4.9. Binding free energy analysis for the embelin and embelin derivatives complexed with XIAP

An MMGBSA protocol was performed as explained in the Methods section to discriminate between the different binding modes of compound 1, compound 2 (embelin) and compound 3 and to correlate their experimental affinities. Tables 6, 7 and 8 summarize (in kcal/mol units) the information averaged from 100 structures of each stabilized MD.

<i>Compound 1-XIAP</i>	<i>Structure 1</i> <i>(Dock-Dyn, GLIDE)</i>	<i>Structure 2</i> <i>(GLIDE)</i>	<i>Structure 3</i> <i>(GLIDE)</i>	<i>Structure 4</i> <i>(GLIDE)</i>
ΔE_{ELE}	-4.37	-12.56	-2.07	-34.14
ΔE_{VDW}	-13.04	-15.40	-5.69	-7.01
ΔG_{SUR}	-1.56	-1.79	-0.70	-1.78
ΔG_{GB}	11.78	17.94	5.57	36.00
ΔG_{SOL}	10.22	16.15	4.87	34.22
ΔG_{ELE}	7.42	5.38	3.50	1.85
ΔG_{TOT}	-7.18	-11.81	-2.88	-6.93
$-T\Delta S$	14.58	13.21	13.81	16.11
ΔG_{MMGBSA}	7.40	1.40*	10.93	9.18
ΔG_{EXP}	-6.83 ^a			

Table 6: Binding free energy, averaged over 100 snapshots, for the compound 1-XIAP complex. All values are shown in kcal/mol units. ΔE_{ELE} and ΔE_{VDW} account for the electrostatic and van der Waals in vacuo binding enthalpic contribution. ΔG_{SUR} accounts for the non polar contribution to solvation, ΔG_{GB} is the polar contribution to solvation. ΔG_{SOL} denotes for the $\Delta G_{\text{SUR}} + \Delta G_{\text{GB}}$. ΔG_{ELE} accounts for the $\Delta E_{\text{ELE}} + \Delta G_{\text{GB}}$ addition. ΔG_{TOT} accounts for the sum of polar and nonpolar solvation contribution to the binding free energy. $-T\Delta S$ accounts for the entropic balance (truncating the receptor to only those atoms within a cutoff of 12 Å from the ligand and using 5 snapshots). Final theoretical free energy of binding at 300 K is denoted as ΔG_{MMGBSA} while the experimental one is denoted as ΔG_{EXP} . a) From the affinity constant of Chen et al. [24]. *) Denotes for the best result.

For the first complex, which consists of the simplest embelin derivative (compound 1) and XIAP, the structure from pose 2 (Figure 9D) has the highest affinity, as suggested before, although the binding free energy is positive. Though this result can be considered a limitation of the method [52], its value correctly separates the structure of pose 2 from the others, which are 6.0-9.5 kcal/mol less stable. This binding mode has the highest van der Waals balance during binding together with an important electrostatic interaction. Electrostatic energy of structures from poses 1 and 3 is drastically reduced to -4.4 and -2.1 kcal/mol and the structure reached from pose 4 has an unfavourable desolvation penalty of 34.2 kcal/mol, which affects the final value of the calculated binding free energy. Entropic effects, highly similar in all binding modes, are not determinant factors.

Table 7 summarizes the MMGBSA information for compound 2-XIAP complex. In this case, the binding mode shown in Figure 10B has the closest affinity, though close to the affinity of GLIDE's structure reached throughout the dynamics simulation of pose 1. The second structure of GLIDE is clearly less active, with van der Waals energy reduced by 10 kcal/mol with respect to the other binding modes. The best binding is achieved with an almost hydrophobic binding mode, with the highest van der Waals contribution and the lowest electrostatic and desolvation penalty contributions, -3.5 and 10.4 kcal/mol, respectively. On the other hand, the addition of a long aliphatic chain increases van der Waals energy over the results of the compound 1-XIAP complex (shown in Table 6). Entropic effects are again not important in selection of the binding mode with the highest affinity. As commented above, experimental information confirmed the role of W323 and Y324 in the binding mode of embelin. Consequently, the structure shown in Figure 10B is suggested as the most reliable one, in corroboration with the MMGBSA results.

Finally, Table 8 shows the MMGBSA data for the complex formed by compound 3, which is the most active embelin derivative. The pose docked with the Dock_Dyn program and optimized throughout the MD simulation, Figure 11B, becomes the most active, with a calculated binding free energy of -10.2 kcal/mol. Other binding modes are clearly secondary. The closest binding free energy is reduced to -2.6 kcal/mol. Similar to what was observed in the best structure of the other complexes, this structure has the highest van der Waals contribution and the highest contribution among the other embelin derivatives. Therefore, the addition of two aromatic rings gives this binding advantage. This is a determinant factor for binding free energy. In addition, entropic effects are favourable for this binding

mode (14.3 kcal/mol) while this factor increases for structures optimized from GLIDE's poses 1, 2 and 3.

<i>Compound 2-XIAP</i>	<i>Structure 1 (Dock_Dyn)</i>	<i>Structure 1 (GLIDE)</i>	<i>Structure 2 (GLIDE)</i>
ΔE_{ELE}	-3.47	-12.07	-13.14
ΔE_{VDW}	-26.75	-26.55	-16.75
ΔG_{SUR}	-2.73	-3.15	-2.40
ΔG_{GB}	10.43	20.38	19.72
ΔG_{SOL}	7.70	17.23	17.32
ΔG_{ELE}	6.95	8.31	6.57
ΔG_{TOT}	-22.52	-21.39	-12.58
$-T\Delta S$	18.58	18.14	17.09
ΔG_{MMGBSA}	-3.94*	-3.25	4.51
ΔG_{EXP}	-8.78 ^a		

Table 7: Binding free energy, averaged over 100 snapshots, for the compound 2-XIAP complex. All values are shown in kcal/mol units. ΔE_{ELE} and ΔE_{VDW} account for the electrostatic and van der Waals in vacuo binding enthalpic contribution. ΔG_{SUR} accounts for the non polar contribution to solvation, ΔG_{GB} is the polar contribution to solvation. ΔG_{SOL} denotes for the $\Delta G_{\text{SUR}} + \Delta G_{\text{GB}}$. ΔG_{ELE} accounts for the $\Delta E_{\text{ELE}} + \Delta G_{\text{GB}}$ addition. ΔG_{TOT} accounts for the sum of polar and nonpolar solvation contribution to the binding free energy. $-T\Delta S$ accounts for the entropic balance (truncating the receptor to only those atoms within a cutoff of 12 Å from the ligand and using 5 snapshots). Final theoretical free energy of binding at 300 K is denoted as ΔG_{MMGBSA} while the experimental one is denoted as ΔG_{EXP} . a) From the affinity constant of Chen et al. [24]. *) Denotes for the best result.

<i>Compound 3-XIAP</i>	<i>Structure 1 (Dock_Dyn)</i>	<i>Structure 1 (GLIDE)</i>	<i>Structure 2 (GLIDE)</i>	<i>Structure 3 (GLIDE)</i>
ΔE_{ELE}	-12.29	-18.44	-21.09	-11.73
ΔE_{VDW}	-35.37	-19.62	-25.95	-12.84
ΔG_{SUR}	-4.00	-2.77	-3.42	-2.25
ΔG_{GB}	27.20	26.13	30.85	21.36
ΔG_{SOL}	23.19	23.37	27.43	19.11
ΔG_{ELE}	14.90	7.69	9.76	9.63
ΔG_{TOT}	-24.47	-14.70	-19.61	-5.47
$-T\Delta S$	14.29	22.09	17.02	21.78
ΔG_{MMGBSA}	-10.18*	7.39	-2.59	16.31
ΔG_{EXP}	-9.25 ^a			

Table 8: Binding free energy, averaged over 100 snapshots, for the compound 3-XIAP complex. All values are shown in kcal/mol units. ΔE_{ELE} and ΔE_{VDW} account for the electrostatic and van der Waals in vacuo binding enthalpic contribution. ΔG_{SUR} accounts for the non polar contribution to solvation, G_{GB} is the polar contribution to solvation. ΔG_{SOL} denotes for the $\Delta G_{SUR} + \Delta G_{GB}$. ΔG_{ELE} accounts for the $\Delta E_{ELE} + \Delta G_{GB}$ addition. ΔG_{TOT} accounts for the sum of polar and nonpolar solvation contribution to the binding free energy. $-T\Delta S$ accounts for the entropic balance (truncating the receptor to only those atoms within a cutoff of 12 Å from the ligand and using 5 snapshots). Final theoretical free energy of binding at 300 K is denoted as ΔG_{MMGBSA} while the experimental one is denoted as ΔG_{EXP} . a) From the affinity constant of Chen et al. [24]. *) Denotes for the best result.

Finally, it should be noted that although the absolute values of binding free energies are not well reproduced by the MMGBSA methodology, relative values indicate, with very remarkable correlation, the correct behaviour of the studied compounds. Although an experimental structure of XIAP with embelin and its derivatives would be desirable, the computational protocol applied adds some insights into the binding mode of these molecules and clearly connects protein-protein recognition with drug design. This will be helpful in the design of new, more potent XIAP inhibitors.

5. CONCLUSIONS

Molecular dynamics simulations of Smac/DIABLO-XIAP, Smac/DIABLO-Survivin, embelin-XIAP and two embelin derivatives-XIAP apoptotic complexes were performed with the aim of improving understanding of the hydrogen bond, van der Waals and electrostatic contacts in protein-protein and protein-ligand structures. As expected, for the first complex, the majority of interactions already observed from NMR or crystal structure were conserved throughout the MD simulation. However, the simulation showed small changes, such as the removal of two hydrogen bonds of the first residue of Smac/DIABLO and the finding of an important electrostatic contact between D309 and the same first residue. W323A, W310A and E314S mutations were correctly explained from the simulation.

For the second complex, the ligand showed a similar recognition pattern: our simulation was the first theoretical and structural study of the Survivin-Smac/DIABLO complex. Due to the great similarity encountered between the suggested pharmacophores of XIAP and Survivin, we argue that designing a Smac/DIABLO mimetic that differentiates both targets would be a difficult task. However, our analyses indicate that due to the different electrostatic and hydrophobic character of the binding groove of the BIR3 domain of the XIAP protein and the BIR domain of the Survivin protein, their binding free energy with a Smac/DIABLO mimetic can be very different. This prediction corroborates available experimental results.

Eight (or seven) points of the four amino terminal residues form the pharmacophore of Smac/DIABLO and it is possible to design small molecules with some of these contacts that mimic this protein. For compounds which were not Smac/DIABLO mimetics, embelin and two embelin derivatives were also studied. The joint use of docking, MD simulations and MMGBSA calculations allowed us to suggest their binding mode.

Results indicate that taking into account protein-protein recognition can be a good first step towards the understanding of binding modes of known drugs and towards the design of new ones.

Finally, we used the cationic dummy approach for the treatment of the force field parameters of the zinc atom of XIAP and Survivin domains with good results, indicating its correct use to model metalloproteins containing a tetrahedrally coordinated zinc.

6. REFERENCES

1. Salgado, J.; Garcia-Saez, A.; Malet, G.; Mingarro, I.; Perez-Paya, E. Peptides in apoptosis research. *J. Pept. Sci.* **2002**, *8*, 543-560.
2. Salvesen, G. S.; Duckett, C. S. IAP proteins: blocking the road to death's door. *Mol. Cell Biol.* **2002**, *3*, 401-410.
3. Martins, L. M.; Iaccarino, I.; Tenev, T.; Gschmeissner, S.; Totty, N. F.; Lemoine, N. R.; Savopoulos, J.; Gray, C. W.; Creasy, C. L.; Dingwall, C.; Downward, J. The serine protease Omi/HtrA2 regulates apoptosis by binding XIAP through a reaper-like motif. *J. Biol. Chem.* **2002**, *277*, 439-444.
4. Riedl, S.; Renatus, M.; Schwarzenbacher, R.; Zhou, Q.; Sun, C.; Fesik, S.; Liddington, R.; Salvesen, G. Structural basis for the inhibition of Caspase-3 by XIAP. *Cell* **2001**, *104*, 791-800.
5. Chai, J.; Schiozaki, E.; Srinivasula, S.; Wu, Q.; Datta, P.; Alnemri, E.; Shi, T. Structural basis of Caspase-7 inhibition by XIAP. *Cell* **2001**, *104*, 769-780.
6. Suzuki, Y.; Nakabayashi, Y.; Kazuko, N.; Reed, J.; Takahashi, R. X-linked inhibitor of apoptosis protein (XIAP) inhibits Caspase-3 and -7 in distinct modes. *J. Biol. Chem.* **2001**, *276*, 27058-27063.
7. Srinivasula, S.; Ramesh, H.; Saleh, A.; Datta, P.; Shiozaki, E.; Chai, J.; Lee, R.; Robbins, P.D.; Fernandes-Alnemri, T.; Shi, Y.; Alnemri, E. A conserved XIAP-interaction motif in Caspase-9 and Smac/DIABLO regulates Caspase activity and apoptosis. *Nature* **2001**, *410*, 112-116.
8. Liu, Z.; Sun, C.; Olejniczak, E.; Meadows, R.; Betz, S.; Oost, T.; Herrmann, J.; Wu, J.; Fesik, S. Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* **2000**, *408*, 1004-1008.
9. Wu, G.; Chai, J.; Suber, T.; Wu, J.; Du, C.; Wang, X.; Shi, Y. Structural basis of IAP recognition by Smac/DIABLO. *Nature* **2000**, *408*, 1008-1012.
10. Shiraki, K.; Sugimoto, K.; Yamanaka, Y.; Yamaguchi, Y.; Saitou, Y.; Ito, K.; Yamamoto, N.; Yamanaka, T.; Fujikawa, K.; Murata, K.; Kano, T. Overexpression of X-linked inhibitor of apoptosis in human hepatocellular carcinoma. *Int. J. Mol. Med.* **2003**, *12*, 705-708.

11. Tamm, I.; Kornblau, S.M.; Segall, H.; Krajewski, S.; Welsh, K.; Kitada, S.; Scudiero, D. A.; Tudor, G.; Qui, Y. H.; Monks, A.; Andreeff, M.; Reed, J. C. Expression and prognostic significance of IAP-family genes in human cancers and myeloid leukemias. *Clin. Cancer Res.* **2000**, *6*, 1796-1803.
12. Zhang, X. D.; Zhang, X. Y.; Gray, C. P.; Nguyen, T.; Hersey, P. Tumor necrosis factor-related apoptosis-inducing ligand-induced apoptosis of human melanoma is regulated by Smac/DIABLO release from mitochondria. *Cancer Res.* **2001**, *61*, 7339-7348.
13. McNeish, I. A.; Bell, S.; McKay, T.; Tenev, T.; Marani, M.; Lemoine, N. R. Expression of Smac/DIABLO in ovarian carcinoma cells induces apoptosis via a Caspase-9-mediated pathway. *Exp. Cell Res.* **2003**, *286*, 186-198.
14. Ng, C. P.; Bonavida, B. X-linked Inhibitor of apoptosis (XIAP) blocks Apo2 ligand/tumor necrosis factor-related apoptosis-inducing ligand-mediated apoptosis of prostate cancer cells in the presence of mitochondrial activation: sensitization by overexpression of second mitochondria-derived activator of Caspase/direct IAP-binding protein with low pI (Smac/DIABLO). *Mol. Cancer Ther.* **2002**, *1*, 1051-1058.
15. Reed, J. C. Apoptosis-based therapies. *Nat. Rev. Drug Discov.* **2002**, *1*, 111-121.
16. Igney, F. H.; Krammer, P. H. Death and anti-death: tumour resistance to apoptosis. *Nat. Rev. Cancer.* **2002**, *2*, 277-288.
17. Kipp, R. A.; Case, M. A.; Wist, A. D.; Cresson, C. M.; Carrell, M.; Griner, E.; Wiita, A.; Albiniak, P. A.; Chai, J.; Shi, Y.; Semmelhack, F.; McLendon, G. L. Molecular targeting of inhibitor of apoptosis proteins based on small molecule mimics of natural binding partners. *Biochemistry* **2002**, *41*, 7344-7349.
18. Glover, C. J.; Hite, K.; DeLosh, R.; Scudiero, D. A.; Fivash, M. J.; Smith, L. R.; Fisher, R. J.; Wu, J.; Shi, Y.; Kipp, R. A.; McLendon, G. L.; Sausville, E. A.; Shoemaker, R. H. A high-throughput screen for identification of molecular mimics of Smac/DIABLO utilizing a fluorescence polarization assay. *Anal. Biochem.* **2003**, *320*, 157-169.
19. Oost, T. K.; Sun, C.; Armstrong, R. C.; Al-Assaad, A.; Betz, S. F.; Deckwerth, T. L.; Ding, H.; Elmore, S. W.; Meadows, R. P.; Olejniczak, E. T.; Oleksijew, A.; Oltersdorf, T.; Rosenberg, S. H.; Shoemaker, A. R.; Tomaselli, K. J.; Zou, H.; Fesik, S. W. Discovery of potent antagonists of the

antiapoptotic protein XIAP for the treatment of cancer. *J. Med. Chem.* **2004**, *47*, 4417-4426.

20. Park, C. M.; Sun, C.; Olejniczak, E. T.; Wilson, A. E.; Meadows, R. P.; Betz, S. F.; Elmore, S. W.; Fesik, S. W. Non-peptidic small molecule inhibitors of XIAP. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 771-775.

21. Sun, H.; Nikolovska-Coleska, Z.; Chen, J.; Chao-Yie Yang, C. Y.; Tomita, Y.; Pan, H.; Yoshioka, Y.; Krajewski, K.; Roller, P. P.; Wang, S. Structure-based design, synthesis and biochemical testing of novel and potent Smac peptido-mimetics. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 793-797.

22. Li, L.; Thomas, R. M.; Suzuki, H.; De Brabander, J. K.; Wang, X.; Harran, P. G. A small molecule Smac mimic potentiates TRAIL- and TNF - mediated cell death. *Science* **2004**, *305*, 1471-1474.

23. Nikolovska-Coleska, Z.; Xu, L.; Hu, Z.; Tomita, Y.; Li, P.; Roller, P.; Wang, R.; Fang, X.; Guo, R.; Zhang, M.; Lippman, M.; Yang, D.; Wang, S. Discovery of embelin as a cell-permeable, small-molecular weight inhibitor of XIAP through structure-based computational screening of a traditional herbal medicine three-dimensional structure database. *J. Med. Chem.* **2004**, *47*, 2430-2440.

24. Chen, J.; Nikolovska-Coleska, Z.; Wang, G.; Qiu, S.; Wang, S. Design, synthesis, and characterization of new embelin derivatives as potent inhibitors of X-linked inhibitor of apoptosis protein. *Bioorg. Med. Chem. Lett.* **2006**, *18*, 5805-5808.

25. Altieri, D. C. Validating Survivin as a cancer therapeutic target. *Nat. Rev. Cancer* **2003**, *3*, 46-54.

26. Verdecia, M. A.; Huang, H.; Dutil, E.; Kaiser, D.A.; Hunter, T.; Noel, J. P. Structure of the human anti-apoptotic protein Survivin reveals a dimeric arrangement. *Nat. Struct. Biol.* **2000**, *7*, 602-608.

27. Sun, C.; Nettlesheim, D.; Liu, Z.; Olejniczak, E. T. Solution structure of human Survivin and its binding interface with Smac/Diablo. *Biochemistry* **2005**, *44*, 11-17.

28. Chantalat, L.; Skoufias, D.A.; Kleman, J. P.; Jung, B.; Dideberg, O.; Margolis, R. L. Crystal structure of human survivin reveals a bow tie-shaped dimer with two unusual alpha-helical extensions. *Mol. Cell* **2000**, *6*, 183-189.

29. Song, Z.; Yao, X.; Wu, M. Direct interaction between survivin and Smac/DIABLO is essential for

the anti-apoptotic activity of survivin during taxol-induced apoptosis. *J. Biol. Chem.* **2003**, *278*, 23130-23140.

30. Cornell, W. D.; Cieplak, P.; Bayly, C. L.; Goud, I. R.; Mertz Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

31. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. D.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Sligh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, University of California, San Francisco. **2002**.

32. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

33. Pang, Y.; Xu, K.; El Yazla, J.; Prendergast, F. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **2000**, *9*, 1857-1865.

34. Pang, Y. P. Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment. *J. Mol. Model.* **1999**, *5*, 196-202.

35. Pang, Y. P. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins* **2001**, *45*, 183-189.

36. Oelschlaeger, P.; Schmid, D. R.; Pleiss, J. Insight into the mechanism of the IMP-1 metallo- β -lactamase by molecular dynamics simulations. *Protein Eng.* **2003**, *16*, 341-350.

37. Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, *14*, 33-38.

38. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

39. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based

construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128–137.

40. Jorgensen, W. L.; Chandross, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

41. Berendsen, H. J. C.; Postman, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

42. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Chem.* **1977**, *23*, 327-341.

43. Bashford, D.; Case, T. A. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129-152.

44. Tsui, V.; Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Nucleic Acid. Sci.* **2001**, *56*, 275-291.

45. Bashford, D.; Gerwert, K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **1992**, *224*, 473-486.

46. Weiser, J.; Shemkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217-230.

47. Rubio-Martinez, J.; Pinto, M.; Tomás, M. S.; Pérez, J. J. Dock_Dyn: A program for fast molecular docking using molecular Dynamics information. University of Barcelona and Technical University of Catalonia. Barcelona, **2005**.

48. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Methods and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.

- 49.** Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field from proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.
- 50.** Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. Gaussian, Inc., Wallingford CT, **2004**.
- 51.** Wang, J.; Wolf, R. M.; Caldwell, J.W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- 52.** Pearlman, D. A. Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a cogeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796-7806.
- 53.** Schimmer, A. D.; Dalili, S.; Riedl, S. J. Targeting XIAP for the treatment of malignancy. *Cell Death Differ.* **2006**, *13*, 179-188.
- 54.** Obiol-Pardo, C.; Rubio-Martinez, J. Comparative evaluation of MMPBSA and XSCORE to compute binding free energy in XIAP-peptide complexes. *J. Chem. Inform. Model.* **2007**, *47*, 134-142.
- 55.** Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531-568.

CHAPTER III: Pharmacophore exploitation of
Smac/DIABLO complexed with XIAP and Survivin



1. BRIEF INTRODUCTION

In the third chapter, the exploitation of the Smac/DIABLO pharmacophore will be discussed in order to obtain new active molecules inhibitors of XIAP and Survivin. Both proteins are implicated in the apoptosis pathway being responsible of the resistance of tumors to conventional treatments, moreover both proteins would have an emerging role in cancer disease.

Although it has been discovered several active molecules against XIAP, based on peptidomimetics of the Smac/DIABLO protein, up to now no one of them have reached the pharmaceutical market.

Here, we will describe the methods and results obtained searching for small non-peptidic molecules inhibitors of XIAP and Survivin using different approaches of molecular modeling, such as the pharmacophore generation, the 3D database searching and the docking and scoring strategies.

2. CONTEXT

XIAP and Survivin are the most important members of the Inhibitors of Apoptosis Proteins (IAPs) [1] which are related to the progression and resistance of tumor cells to current chemotherapeutic agents. As it was indicated in the second chapter, the natural protein inhibitor of XIAP and probably Survivin is Smac/DIABLO. Several efforts have been made to discover new small molecules inhibitors of XIAP. As commented in chapter 2, the most potent inhibitors of XIAP [2-18] are based on modifications of the natural backbone of the Smac/DIABLO protein, adding additional groups or reducing the flexibility of the molecules. Unfortunately, molecules mimicking the natural peptide backbone have low activities at cell level, although new advances have been obtained in this direction. The chemical structures of the inhibitors of XIAP discovered up to now, are summarized in Figure 1.

In addition, the non sense strategy, which is based on antisense oligonucleotides that form duplexes with intracellular mRNA, has been applied to find two promising compounds to block XIAP expression. Thus, the companies Aegera (Montreal) and Avi BioPharma (Portland, Oregon) have antisense XIAP inhibitors in the clinical stage of Phase 1.

With respect to Survivin, Isis Pharmaceuticals (Texas) has also an antisense inhibitor in Phase 1 and recently, it was discovered a small molecule, named YM155, that acts by a transcriptional inhibition of

the Survivin gene promoter which is currently in Phase 2 [19]. Moreover, a novel family of dimerization inhibitors for Survivin have been reported using a HTS-NMR assay [20].

As it was stated in the second chapter, we performed a Molecular Dynamics simulation of the bioactive nine residue segment of Smac/DIABLO peptide in complex with both proteins finding a pharmacophore formed by 8 (or 7 for Survivin) interaction points.

Here, we will use this information in order to propose some molecules that would be active against these antiapoptotic targets.

Firstly, we have used the geometric and chemical information of our 8-point pharmacophore to search in 3D databases for small non peptidic compounds presenting similar interactions. For this purpose it was used the CATALYST program [21], to scan each favourable conformation of each molecule among all the available compounds and to reduce the results (if it is required) by applying some parameters of restriction, such as the molecular weight.

Secondly, all candidate molecules were docked as flexible ligands inside the protein receptor using our home-made program of docking [22] together with the geometric positions of our pharmacophore, thus the docking procedure is extremely fast and directed. Then, all accepted molecules, without steric hindrance and with a reasonable pharmacophore disposition, measured by their RMS deviation, were scored by using the XSCORE semiempirical function [23].

Using docking and scoring protocols we can order the best compounds with more possibilities to be active molecules, this rank is translated in purchasing the best compounds from the commercial sources. This strategy is named virtual screening in opposition of the traditional high throughput screening, which needs for a high number of experimental tests to obtain some positive hits.

Finally, our purchased compounds have been evaluated by an experimental group, to test the molecules at protein extract level or at cell level, where the molecules have to transfer the cell bilayer and then to reach the target protein.

By using this protocol, it has been found 4 active molecules, which act decreasing the cell division level in different tumor cell lines. Preliminary results indicate that the XIAP is the target protein of these compounds. In summary, molecular modeling tools can help in the drug design process of finding new lead compounds.

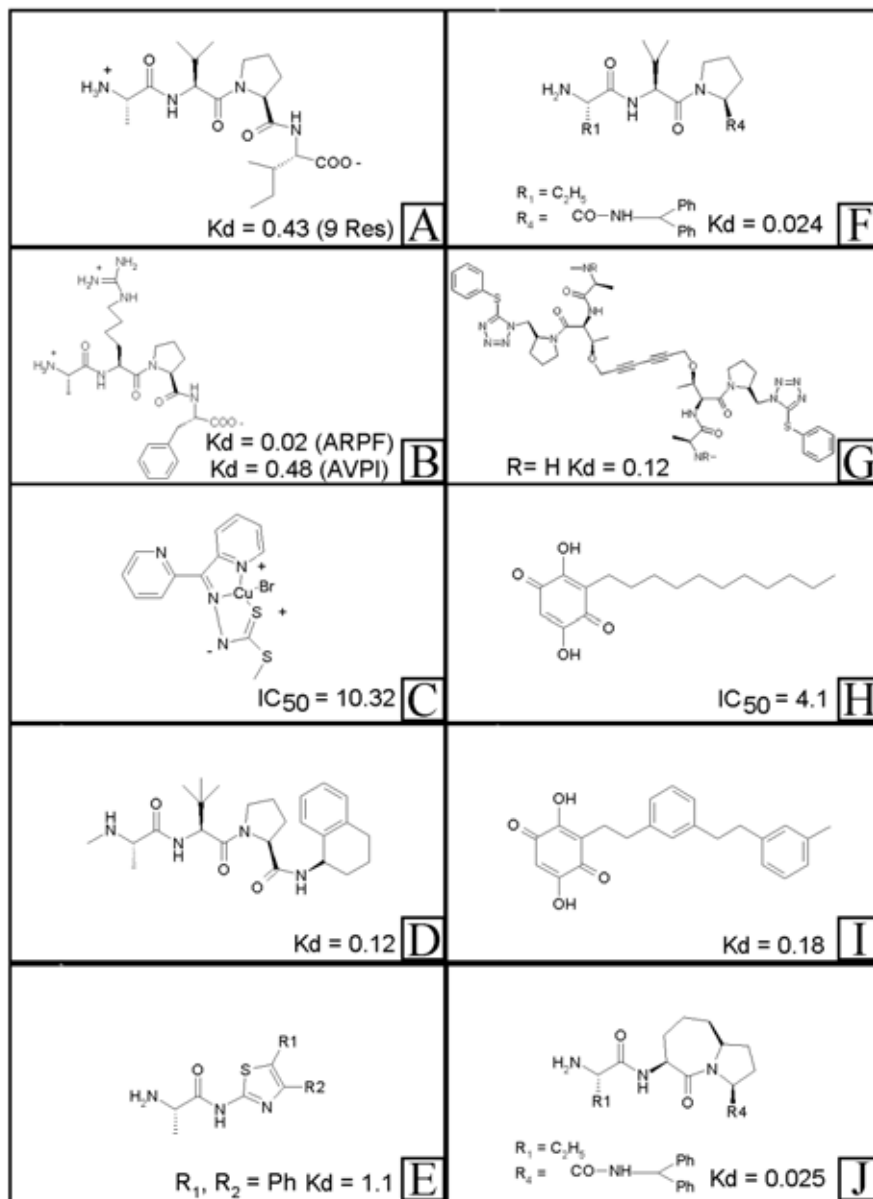


Figure 1 (I): Selected XIAP inhibitors from the literature. The dissociation constant (Kd) or half maximal inhibitory concentration (IC₅₀) are expressed in μM units. A) from reference [2-3], B) from reference [4], C) from reference [5], D) from reference [6], E) from reference [7], F) from reference [8], G) from reference [9], H) from reference [10], I) from reference [11], J) from reference [12].

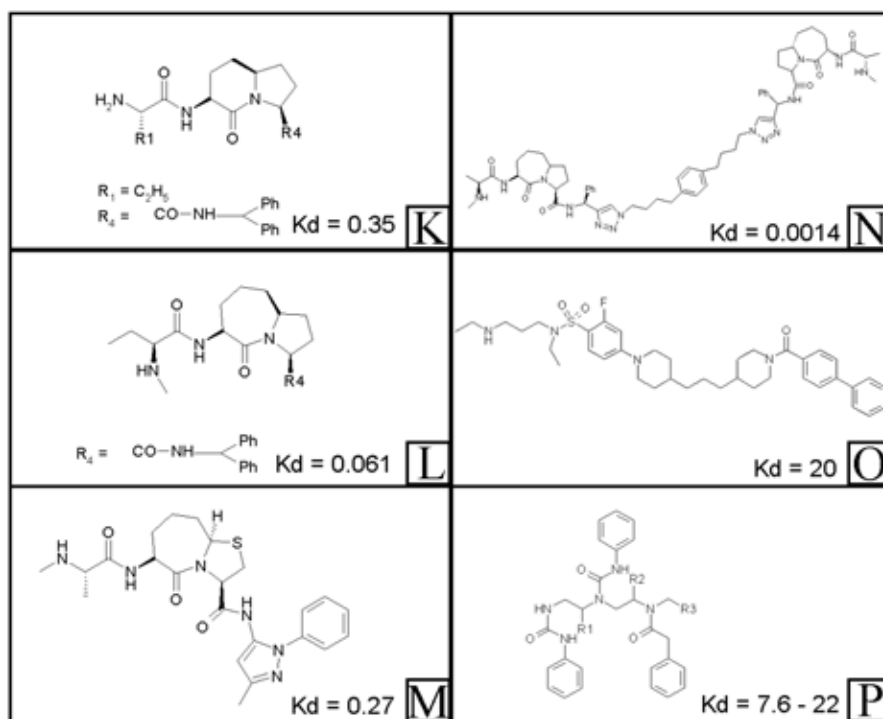


Figure 1 (II): Selected XIAP inhibitors from the literature. The dissociation constant (K_d) and half maximal inhibitory concentration (IC_{50}) are expressed in μM units. K) from reference [13], L) from reference [13], M) from reference [15], N) from reference [16], O) from reference [17], P) from reference [18].

3. RESULTS

3.1. PHARMACOPHORE AND 3D SEARCHING

Figure 2 shows our deduced pharmacophore for Smac/DIABLO when it recognizes the hydrophobic groove of the BIR3 domain of XIAP or the BIR domain of Survivin. It is formed by 8 interaction points (7 in the Survivin case) distributed among the first four residues of Smac/DIABLO (AVPI sequence). On the other hand, Table 1 shows the geometrical parameters which identify this pharmacophore and their dynamical behaviour (maximum and minimum values throughout the molecular dynamics). This information was used as the input for the pharmacophore generation module of CATALYST [20]. Molecules presenting the first 4 and 6 interaction points were found with the database search implemented in CATALYST, using the best flexible search option. Other combination of points were not tested. Finally, 132 molecules were saved from the program to a later evaluation, by means of docking and scoring protocols.

<i>Pharmacophoric Points</i>	<i>Average Distance /Å</i>	<i>Maximum Distance/Å</i>	<i>Minimum Distance/Å</i>
CONFIDENTIAL			

Table 1: Geometrical data of the Smac/DIABLO pharmacophore.

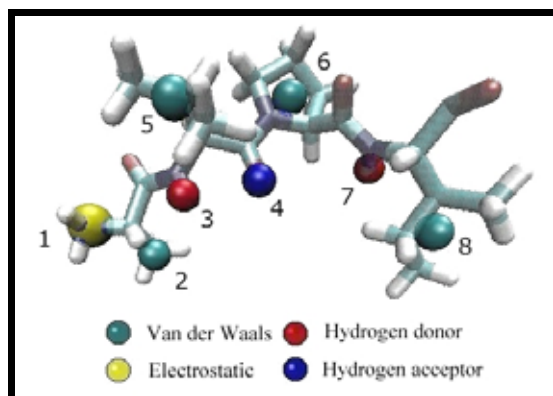


Figure 2: Pharmacophore of Smac/DIABLO.

3.2. DOCKING AND SCORING PROTOCOLS

Our home-made program of docking, Dock_Dyn [22], has been used to evaluate around 100 conformations for each of the 132 database compounds, using XIAP and Survivin as rigid receptors.

To study only the best molecules, the RMS deviation of the pharmacophores with respect to the reference one was limited to a maximum value of 3.3 Å, nevertheless, to study enough molecules and conformations, and to take into account slightly the receptor flexibility, the van der Waals radii were reduced to 60%.

Table 2 shows some general data of the docking-scoring protocol for XIAP. As can be seen, compounds with more recognition points with the receptor have a high value of XSCORE [23] (that can be translated into free energy of binding) although a high value of the RMS deviation.

	<i>Average RMS/ Å</i>	<i>Maximum RMS/ Å</i>	<i>Minimum RMS/ Å</i>	<i>Average XSCORE</i>	<i>Maximum XSCORE</i>	<i>Minimum XSCORE</i>
6 interaction points	1.4	3.3	0.3	4.8	6.0	3.9
4 interaction points	1.2	2.4	0.8	3.9	5.4	3.6

Table 2: Docking and Scoring data for the XIAP protein.

3.3. PURCHASED COMPOUNDS

As a general criterion of selection, compounds with the lowest RMS and the highest XSCORE were purchased. They were selected from the total compounds with 6 (Compounds A, D, F, G and H) and 4 pharmacophoric points (Compounds B, C and E), taking also into account a maximum chemical diversity. Results for both receptors (XIAP and Survivin) were used to select the best molecules.

The best 8 compounds (Figure 3) were acquired from Sigma, ChemDiv, Enamine and Chembridge chemical companies.

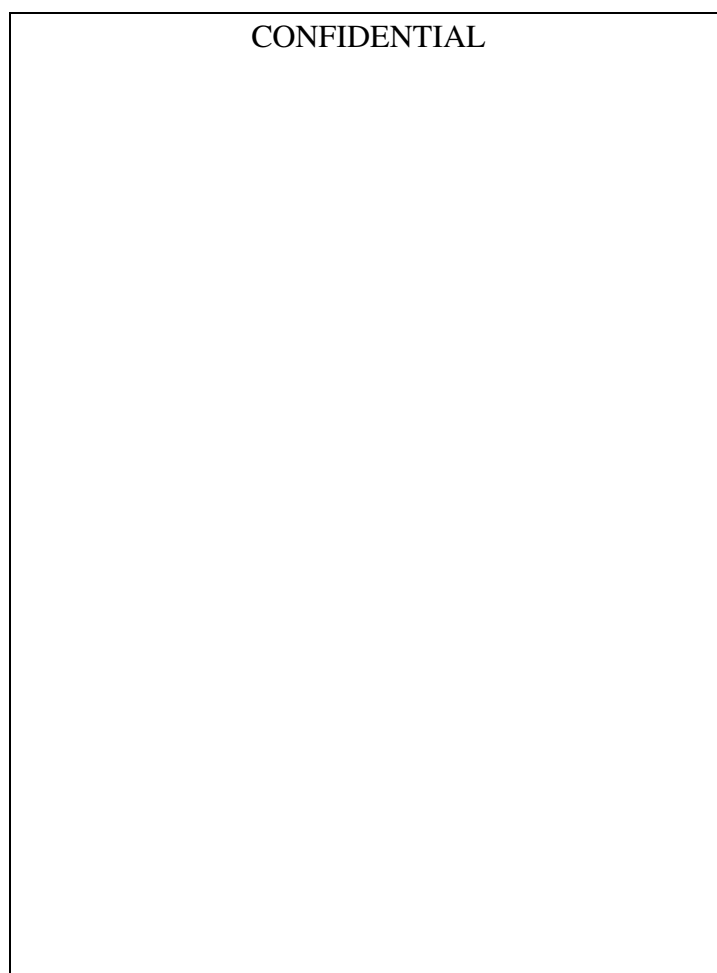


Figure 3: Purchased compounds.

Table 3 includes the best RMS and XSCORE values from the docking and the scoring protocols, for the purchased compounds and for both receptors. Figure 4 shows the best docking poses of compounds A and D, taken as an example, identifying the most important interactions. These interactions concern the residues that recognize the Smac/DIABLO protein, such as T308, D309 and E314 involved in hydrogen bonds and L307, W310, W323 and Y324, involved in van der Waals contacts. Nevertheless, as we have shown in the previous chapter, a MD simulation should be performed to study the receptor flexibility and to confirm the correct binding mode of these molecules.

<i>Compound</i>	<i>Best RMS/Å</i>	<i>Best XSCORE</i>
A	0.85 (1.48)	5.31 (5.23)
B	1.31 (1.00)	5.19 (5.22)
C	1.05 (0.89)	4.76 (4.66)
D	0.94 (0.86)	5.06 (4.68)
E	0.95 (1.45)	5.38 (5.56)
F	0.99 (1.14)	5.29 (4.95)
G	1.32 (1.00)	6.03 (5.13)
H	1.23 (0.92)	5.24 (5.05)

Table 3: Best RMS and XSCORE values from the docking and scoring protocols for the XIAP receptor and for the Survivin receptor, shown in brackets.

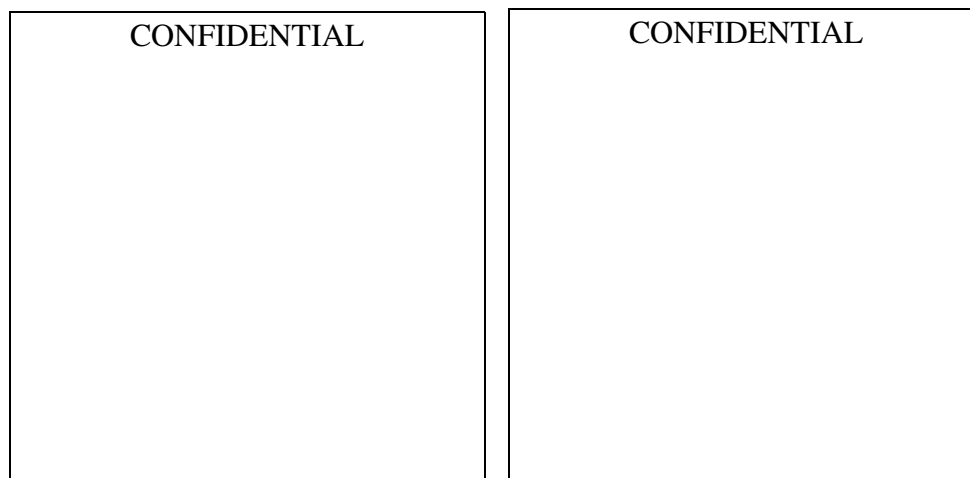


Figure 4: A) Best docking pose for compound A in complex with XIAP, B) Best docking pose of compound D in complex with XIAP. The protein is shown in transparent orange and carbon atoms of the ligand are marked in light green.

3.4. EXPERIMENTAL RESULTS

The 8 compounds (A-H) were tested by the Hematopathology group of the 'Hospital Clinic de Barcelona', coordinated by Dr. Dolors Colomer.

Dr. Roberto Alonso of the same group, performed the experimental tests on mantle lymphoma (Jeko, UPN-1, Z-138, HBL-2, Granta-519 and JVM-2 cell lines) and chronic lymphocytic leukemia cells (MEC-1, EHEB lines and cells of two patients, CLL#1 and CLL#2). Both tumors were correlated with overexpression of XIAP.

The compounds were solved in DMSO and the apoptosis induction was followed, by using an annexin V-FITC and permeability to propidium iodide assays. Results indicate that 4 of the 8 compounds (50%) induce apoptosis, in the range of micromolar activity. Compound A exhibited an average activity of 14-28 μM for all the cell lines while compounds D, F, G exhibited an average activity of 50-100 μM . Concerning the different cell lines treated, compound A induced a 90% of apoptosis for cell line CLL#1 at 14 μM , and a 90% of apoptosis for cancer cell line CLL#2 at 28 μM . Apoptosis induction for MEC-1 and EHEB were produced at 60% using 28 μM of compound A. Compound D exhibited a 40% of apoptosis induction at 56 μM for CLL#1, CLL#2 and MEC-1 lines. Compound G was active at 56 μM , producing a 80% of apoptosis using CLL#1 and CLL#2 cell lines. Compound F activity was the lowest one.

With respect to mantle lymphoma tumor, compound A induces a 90% of apoptosis in Jeko and Z-138 cell lines at 28 μM , and a 80% of apoptosis at 56 μM for cancer cell lines UPN-1 and HBL-2. Compound D exhibited a 60% of apoptosis for Jeko, Z-138 and UPN-1 cell lines at 111 μM and a 85% of induction for HBL-2 at the same concentration. Compound G induces 70% of apoptosis, using Jeko, Z-138, UPN-1 and HBL-2, at 111 μM . Compound F showed the most erratic effect. All these assays were performed during 24h. The effect of compound A, which was the most active one, was also identified using Granta-519 and JVM-2 lines at 6, 15, 24 and 48 h time intervals. The maximum of activity was found at 24 h for the first type of cells and at 15 h for the second one.

Although no tests were performed to verify the direct inhibition of XIAP and Survivin at molecular level, it was performed an additional assay concerning the TRAIL protein receptor. Recently, it was discovered that the mechanism of action of a drug targeting XIAP can potentiate the TRAIL and TNF-

mediated cell death [24-26]. Therefore, the active compounds were also studied in the presence of TRAIL, to evaluate if the apoptosis induction would be increased. All experiments were correlated with an apoptosis induction increment with the addition of TRAIL, which is an indirect fact to confirm a XIAP inhibition. TRAIL-induced cell death was measured using assays during 15 and 24h for cancer cell lines JVM-2, HBL-2, UPN-1 and Granta-519.

The 4 active compounds can be considered as our first generation of active molecules for XIAP and Survivin proteins. Further modifications will be considered, using rational drug design to improve the binding energy and the pharmacokinetic profile. One of the aims of future work will be to increase their activity from micro to submicromolar range.

Concluding, the molecular dynamics analysis of the protein-protein contacts followed by the pharmacophore searching and the docking protocols were successful, allowing us to find new non-peptidic inhibitors, with similar interactions with respect to the Smac/DIABLO protein.

4. CONCLUSIONS

Following an original drug design protocol, which includes molecular dynamics, pharmacophore searching and docking-scoring protocols, we have found 4 active molecules (50% of the total) with apoptotic activity. This is a remarkable result, in opposition with the low successful rate of high throughput screening methods, and an indication of the advantages of molecular modeling methods.

Although tests that correlate the activity of the compounds with the inhibition of XIAP or Survivin were not performed, the micromolar range of activity found at cell level together with the successful rate found, is an indication of the presumable activity of the compounds inhibiting these proteins. In addition, the compounds showed the predicted properties of a XIAP inhibitor using them in combination with TRAIL, which is a receptor protein also implicated in apoptosis.

The active molecules found will be modified in the near future, in order to achieve new ones with improved affinity and chemical properties. For this purpose, structural based methods will be applied again, once the correct binding conformation of each molecule is known. Some guidelines to improve a molecule have been described in the first chapter (2.3.2). As it was cited, one can modify directly the docked molecules into the binding groove of XIAP and Survivin and search for adding extra groups or simplifying the structure.

5. REFERENCES

1. Salvesen, G. S.; Duckett, C. S. IAP proteins: blocking the road to death's door. *Nature Rev. Mol. Cell Bio.* **2002**, *3*, 401-410.
2. Li, L.; Thomas, R. M.; Olejniczak, E.; Meadows, R.; Betz, S.; Oost, T.; Hermann, J.; Wu, J.; Fesik, S. *Nature* **2000**, *408*, 1004-1008.
3. Wu, G.; Chai, J.; Suber, T. L.; Wu, J. W.; Du, C.; Wang, S.; Shi, Y. Structural basis of IAP recognition by Smac/DIABLO. *Nature* **2000**, *408*, 1008-1012.
4. Kipp, R. A.; Case, M. A.; Wist, A. D.; Cresson, C. M.; Carrell, M.; Griner, E.; Wiita, A.; Albiniak, P. A.; Chai, J.; Shi, Y.; Semmelhack, F.; McLendon, G. L. Molecular targeting of inhibitor of apoptosis proteins based on small molecule mimics of natural binding partners. *Biochemistry* **2002**, *41*, 7344-7349.
5. Glover, C. J.; Hite, K.; DeLosh, R.; Scudiero, D. A.; Fivash, M. J.; Smith, L. R.; Fisher, R. J.; Wu, J.; Shi, Y.; Kipp, R. A.; McLendon, G. L.; Sausville, E. A.; Shoemaker, R. H. A high-throughput screen for identification of molecular mimics of Smac/DIABLO utilizing a fluorescence polarization assay. *Anal. Biochem.* **2003**, *320*, 157-169.
6. Oost, T. K.; Armstrong, R. C.; Al-Assad, A.; Betz, S. F.; Deckweth, T. L.; Ding, H.; Elmore, S. W.; Meadows, R. P.; Olejniczak, E. T.; Oleksijew, A.; Oltersdorf, T.; Rosenberg, S. H.; Shoemaker, A. R.; Tomaselli, K. J.; Zou, H.; Fesik, S. W. Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J. Med. Chem.* **2004**, *47*, 4417-4426.
7. Park, C. M.; Sun, C.; Olejniczak, E. T.; Wilson, A. E.; Meadows, R. P.; Betz, S. F.; Elmore, S. W.; Fesik, S. W. Non-peptidic small molecule inhibitors of XIAP. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 771-775.
8. Sun, H.; Nikolovska-Coleska, Z.; Chen, J.; Chao-Yie Yang, C. Y.; Tomita, Y.; Pan, H.; Yoshioka, Y.; Krajewski, K.; Rollerc P. P.; Wang, S. Structure-based design, synthesis and biochemical testing of novel and potent Smac peptido-mimetics. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 793-797.
9. Li, L.; Thomas, R. M.; Suzuki, H.; De Brabaner, J. K.; Wang, X.; Harran, P. G. A small molecule smac mimic potentiates TRAIL- and TNF α – mediated cell death. *Science* **2004**, *305*, 1471-1474.

10. Nikolovska-Coleska, Z.; Xu, L.; Hu, Z.; Tomita, Y.; Li, P.; Roller, P.; Wang R.; Fang, X.; Guo, R.; Zhang, M.; Lippman, M.; Yang, D.; Wang, S. Discovery of embelin as a cell-permeable, small-molecular weight inhibitor of XIAP through structure-based computational screening of a traditional herbal medicine three-dimensional structure database. *J. Med. Chem.* **2004**, *47*, 2430-2440.
11. Chen, J.; Nikolovska-Coleska, Z.; Wang, G.; Qiu, S.; Wang, S. Design, synthesis, and characterization of new embelin derivatives as potent inhibitors of X-linked inhibitors of apoptosis protein. *Bioorg. Med. Chem. Lett.* **2006**, *18*, 5805-5808.
12. Pan, H.; Lu, M.; Tomita, Y.; Krajewski, K.; Wang, S.; Roier, P. P.; Yoshioka, Y.; Sun, H.; Yang C.; Nikolovska-Coleska, Z. Structure-based design of potent, conformationally constrained Smac mimetics. *J. Am. Chem. Soc.* **2004**, *126*, 16686-16687.
13. Sun, H.; Nikolovska-Coleska, Z.; Yang, C. Y.; Xu, L.; Tomita, Y.; Krajewski, K.; Roller, P. P.; Wang, S. Structure-based design, synthesis, and evaluation of conformationally constrained mimetics of the second mitochondrial-derived activator of caspase that target the X-linked inhibitor of apoptosis protein/caspase 9 interaction site. *J. Med. Chem.* **2004**, *47*, 4147-4150.
14. Sun, H.; Nikolovska-Coleska, Z.; Lu, J.; Qiu, Su.; Yang, C. Y.; Gao, W.; Meagher, J.; Stuckey, J.; Wang, S. Design, synthesis, and evaluation of a potent, cell-permeable, conformationally constrained second mitochondria derived activator of caspase (Smac) mimetic. *J. Med. Chem.* **2006**, *49*, 7916-7920.
15. Zobel, K.; Wang, L.; Varfolomeev, E.; Franklin, M. C.; Elliott, L. O.; Wallweber, H. J. A.; Okawa, D. C.; Flygare, J. A.; Vucic, D.; Fairbrother, W. J.; Deshayes, K. Design, synthesis, and biological activity of a potent smac mimetic that sensitizes cancer cells to apoptosis by antagonizing IAPs. *ACS Chem. Biol.* **2006**, *1*, 525-534.
16. Sun, H.; Nikolovska-Coleska, Z.; Lu, J.; Meagher, J. L.; Yang, C. Y.; Qiu, S.; Tomita, Y.; Ueda, Y.; Jiang, S.; Krajewski, K.; Roller, P. P.; Stuckey, J. A.; Wang, S. Design, synthesis, and characterization of a potent, nonpeptide, cell-permeable, bivalent Smac mimetic that concurrently targets both the BIR2 and BIR3 Domains in XIAP. *J. Am. Chem. Soc.* **2007**, *129*, 15279-15294.
17. Tom, Y. H. Wu, T. Y. H.; Wagner, K. W.; Bursulaya, B.; Schultz, P. G.; Deveraux, Q. L. Development and characterization of nonpeptidic small molecule inhibitors of the XIAP/Caspase-3 interaction. *Chem. Biol.* **2003**, *10*, 759-767.
18. Schimmer, A.; Welsh, K.; Pinilla, C.; Wang, Z.; Krajewska, M.; Bonneau, M.; Pedersen, I.; Kitada,

S.; Scott, F.; Bailly-Maitre, B. Small-molecule antagonists of apoptosis suppressor XIAP exhibit broad antitumor activity. *Cancer Cell* **2004**, *5*, 25-35.

19. Nakahara, T.; Takeuchi, M.; Kinoyama, I.; Minematsu, T.; Shirasuna, K.; Matsushisa, A.; Kita, A.; Tominaga, F.; Yamanaka, K.; Kudoh, M.; Sasamata, M. YM155, a novel small-molecule Survivin suppressant, induces regression of established human hormone-refractory prostate tumor xenografts. *Cancer Res.* **2007**, *67*, 8014-8021.

20. Wendt, M. D.; Sun, C.; Kunzer, A.; Sauer, D.; Sarris, K.; Hoff, E.; Yu, L.; Nettesheim, D. G.; Chen, J.; Jin, S.; Comess, K. M.; Fan, Y.; Anderson, S. N.; Isaac, B.; Olejniczak, E. T.; Hajduk, P. J.; Rosenberg, S. H.; Elmore, S. W. Discovery of a novel small molecule binding site of human survivin. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3122-3129.

21. CATALYST™ (Accelrys Inc. USA).

22. Rubio-Martinez, J.; Pinto, M.; Tomas M.S.; Perez, J. J. Dock_Dyn: a program for fast molecular docking using molecular dynamics information. University of Barcelona and Technical University of Catalonia. Barcelona, **2005**.

23. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11-26.

24. Li, L.; Thomas, R.M.; Suzuki, H.; De Brabander, J.K.; Wang, X.; Harran, P.G. A Small Molecule Smac Mimic Potentiates TRAIL- and TNF-Mediated Cell Death. *Science.* **2004**, *305*, 1471-1474.

25. Chauhan, D.; Neri, P.; Velankar, M.; Podar, K.; Hideshima, T.; Fulciniti, M.; Tassone, P.; Raje, N.; Mitsiades, C.S.; Mitsiades, N.; Richardson, P.G.; Zavel, L.; Tran, M.; Munshi, N.C.; Anderson, K.C. Targeting mitochondrial factor Smac/DIABLO as therapy for multiple myeloma (MM). *Blood.* **2006**, *109*, 1220-1227.

26. Vogler, M.; Dürr, K.; Jovanovic, M.; Debatin, K.M.; Fulda, S. Regulation of TRAIL-induced apoptosis by XIAP in pancreatic carcinoma cells. *Oncogen.* **2006**, *26*, 248-257.

Published: J. Chem. Inf. Model. 2007. 47, 134-142.

CHAPTER IV: Comparative Evaluation of MMPBSA and XSCORE to
Compute Binding Free Energy in XIAP-Peptide Complexes



1. BRIEF INTRODUCTION

Evaluation of binding free energy in receptor-ligand complexes is one of the most important challenges in theoretical drug design. Free energy is directly correlated to the thermodynamic affinity constant and, as a first step in drug likeness, a lead compound must have this constant in the range of micro to nanomolar activity. Many efforts have been made to calculate it by rigorous computational approaches, such as free energy perturbation or linear response approximation. However, these methods are still computationally expensive.

We focus our work on XIAP, an antiapoptotic protein whose inhibition can lead to new drugs against cancer disease.

We report here a comparative evaluation of two completely different methodologies to estimate binding free energy, MMPBSA (a force field based function) and XSCORE (an empirical scoring function), in seven XIAP-peptide complexes using a representative set of structures generated by previous molecular dynamics simulations.

Both methods are able to predict the experimental binding free energy with acceptable errors, but if one needs to identify slight differences upon binding, MMPBSA performs better, although XSCORE is not a bad choice taking into account the low computational cost of this method.

2. CONTEXT

Protein-protein interactions[1] are crucial for many biological processes, such as signal transduction or protein inhibition. However, because these interactions are usually distributed along a big and flat surface, the design of small molecules to disrupt them is an unusual approach in drug discovery [2]. Peptides, although not having all the desirable drug-like properties, can be a good first approach in a drug design process because they are able to cover a large interaction surface area. Thus, understanding protein-peptide recognition at its atomic and energetic levels is extremely important.

With the advent of faster and cheaper computers, structure-based drug design has become an important step toward a fast and efficient drug discovery project. In this scenario, molecular docking appears as a fundamental tool that consists basically of two major tasks: the generation of all accessible conformations, and their ranking in order to identify the bioactive one. As effectiveness of molecular

docking is strongly dependent on the scoring function used, many efforts have been made to estimate binding free energy of protein-ligand complexes by computational approaches. Thus, free energy perturbation [3] or linear response approximations [4] are rigorous methods that consider the solvent explicitly. However, they are still computationally expensive to study a large and diverse set of protein-ligand or protein-peptide systems.

During the last years, different approaches have been described as cheaper alternatives to the estimation of binding free energy in a fast and more or less accurate form.

A widely used force field scoring function is MMPBSA (Molecular Mechanics Poisson Boltzmann Surface Area) [5]. Usually, this approach computes binding free energy by using a set of conformations for the complex, the ligand and the receptor taken from one molecular dynamics trajectory, together with a continuum solvent model. It has been evaluated with remarkable success in numerous and very different systems such as protein-ligand or RNA-ligand complexes [6-7]. This function can be classified as *ab initio*, in the sense that no experimental results are used for global evaluation, although they have been taken into account for the obtainment of some terms involved in its calculation.

On the other hand, XSCORE [8] is an experimental scoring function as it is calibrated with experimental information. It takes into account van der Waals interactions, hydrogen bonding, deformation penalty, and hydrophobic effects between the receptor and the ligand. This function was able to predict the binding free energy with a small deviation of 2.2 kcal/mol in a set of 30 protein-ligand complexes [8]. It has recently been evaluated in comparison with 10 different scoring functions [9] by using an exhaustive conformational sampling. Results indicate that XSCORE has an acceptable success rate for molecular docking tasks and binding free energy prediction in protein-ligand systems. This method is computationally very cheap and it has been developed to treat a large set of ligands within a rigid receptor structure approach.

Both methodologies, XSCORE and MMPBSA, are able to provide clear physical meaning of the suitable ligand features for the inhibition of a protein. The first one can provide a number of hydrogen bonds, hydrophobic zones or a number of frozen rotatable bonds during the binding process, among others. The second one can provide electrostatic, van der Waals and solvation contributions to the binding process.

We report here a comparative evaluation between both methods to estimate the binding free energy of seven protein-peptide complexes with biological relevance.

These complexes are formed by the well-known X-linked inhibitor of apoptosis protein (XIAP), related to the progress and resistance of tumors to conventional treatments [10-12], with seven inhibitor peptides (Figure 1), one of them being the 9-residue peptide derived from Smac/DIABLO protein, a natural inhibitor of the XIAP protein.

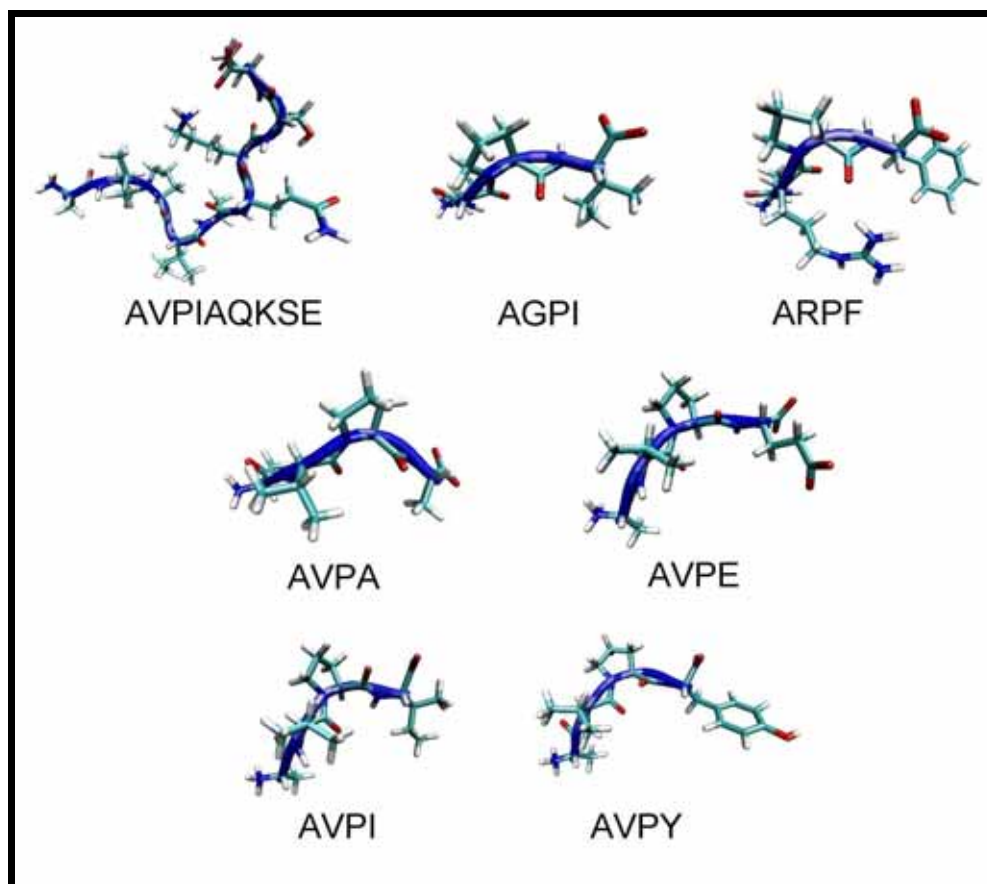


Figure 1: The 9-residue peptide and the 4-residue peptides studied.

The remaining six ones are 4-residue peptides derived from the Smac/DIABLO AVPI sequence in a wide range of experimental affinities [13], ranging from micro to nanomolar. Many recent experimental studies have been devoted to the development of potent inhibitors of XIAP based mainly on the 4-residue sequence of the Smac/DIABLO [14-16]. So it is interesting to find a fast and reliable theoretical method for the reproduction of available experimental affinities and the prediction of new ones.

3. METHODS

3.1. Construction of the XIAP-Smac/DIABLO (9 residues) complex

The initial 3D structure of the human XIAP-Smac/DIABLO(9) complex was taken from the Protein Data Bank (PDB entry 1G3F) [17]. The complex is formed by the BIR3 domain of the XIAP protein (residues 240 to 357), and nine nitrogen terminal residues of the Smac/DIABLO protein (residues 1 to 9, AVPIAQKSE sequence). The cationic dummy approach [18] was used for the treatment of the zinc atom of the XIAP protein. In this method, four dummy atoms are used to impose tetrahedral orientation required for zinc ligands (see Figure 2). This method was employed with remarkable success in farnesyltransferase [18], matrix metalloproteinase [19], phosphotriesterase [20] and beta-lactamase [21]. It solves the problem of maintaining the tetrahedral coordination of the metal throughout a molecular dynamics simulation without the loss of protein flexibility. The zinc atom is bounded covalently to the dummies and interacts with the protein only by van der Waals forces while the dummies interact with the protein only by electrostatics. A cubic box of 8,243 TIP3P waters [22] was added to the system to perform the molecular dynamics simulation in explicit solvent. No counterions were added except for a uniform plasma neutralizing the system, as implemented in AMBER package.

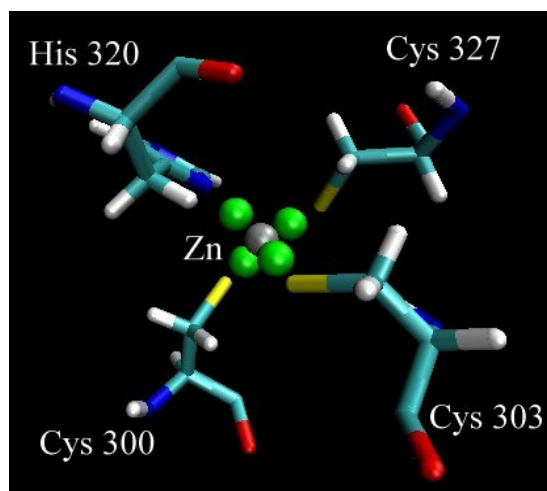


Figure 2: Tetrahedral structure of dummy atoms and zinc in XIAP.

3.2. Construction of the XIAP-peptides (4 residues) complexes

Six XIAP-peptide(4) complexes were modeled straight from the coordinates of the first complex cutting the five carbon terminal residues of Smac/DIABLO. Then, initial coordinates of main chain atoms are always the same for all designed peptides. In this form, conformational sampling begins with the expected bioactive conformation. Appropriate point mutations for lateral chains were done when required. These peptides are AVPI (simply the four first residues), ARPF, AGPI, AVPA, AVPY and AVPE. In the same way, the cationic dummy approach was used for the zinc atom present in XIAP. A cubic box of approximately 8,000 TIP3P waters [22] was added to each system to perform molecular dynamics simulations in explicit solvent. No counterions were added except for a uniform plasma neutralizing the system, as implemented in AMBER package.

3.3. Minimization and molecular dynamics

All the calculations were carried out at molecular mechanics level using the parm94 [23] force field as implemented in the AMBER-7 suite of programs [24]. The solvent was considered explicitly and the cut-off distance was kept to 9 Å to compute the nonbonded interactions. All simulations were performed under periodic boundary conditions and long-range electrostatics were treated by using the particle-mesh-Ewald method [25].

The seven complexes were energy minimized to remove possible steric stress by a multistep procedure. First, water molecules were allowed to relax while the rest of the system was kept frozen. Second, side chains of XIAP and peptides were relaxed as well as water molecules. Third, all atoms except zinc, dummies and the four coordinated residues of XIAP were relaxed, and finally all atoms were allowed to move. We used the steepest descent method followed by the conjugated gradient method to achieve energy gradients lower than 1 kcal/mol, which are reasonable gradients for local minimums, and good structures to start the molecular dynamics trajectory.

Molecular dynamics for the complexes were performed at constant temperature by coupling the systems to a thermal bath using Berendsen's algorithm [26], with a time coupling constant of 0.2 ps. The time

integration step was set to 1fs, and the list of nearest neighbor atoms was updated every 15 steps. A cut-off distance of 9 Å was used. All bond lengths were constrained with the SHAKE algorithm[27] to achieve a rapid energy convergence.

Molecular dynamics began by heating up each of the minimized systems to 300K at a constant rate of 30K/10ps constraining the protein atoms. The second step consisted of a 40ps pressure-constant period to raise the density while still keeping the protein atoms constrained. The third step was a 150ps volume-constant period with only the zinc, dummies and four coordinated residues constrained. Finally, 1ns dynamics calculations were performed for each free system in the NVT ensemble at a constant temperature of 300 K.

Once the total energy of the systems was equilibrated, one hundred time-equidistant snapshots were taken out from MD production of each XIAP-peptide complex. After removing water molecules, structures were used for the evaluation of binding free energies.

3.4. XSCORE

XSCORE [8] is an empirical scoring function that computes the binding free energy with the following terms:

$$\Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{H-bond}} + \Delta G_{\text{deformation}} + \Delta G_{\text{hydrophobic}} + \Delta G_0 \quad (1)$$

Here, ΔG_{vdw} accounts for the van der Waals interactions between the receptor and the ligand, $\Delta G_{\text{H-bond}}$ accounts for the hydrogen bonding between the receptor and the ligand, $\Delta G_{\text{deformation}}$ accounts for the deformation penalty (number of ligand rotatable bonds frozen during the binding process), $\Delta G_{\text{hydrophobic}}$ accounts for the hydrophobic effect with three different algorithms (HS, HP and HM), the HP, HM and HS term are related to optimum distances of hydrophobic contacts, to the environment of hydrophobic ligand atoms into the receptor and to the surface accessible area respectively and ΔG_0 is a regression constant. Finally, the binding affinity of a given protein-ligand complex, is expressed in pK_d units, with K_d being the dissociation constant ($\text{pK}_d = -\log \text{K}_d$). Detailed information about how these terms have

been obtained can be read in the original reference [8]. Regression constants in the present work were not altered from the original XSCORE function and version 1.2 of the program was used.

XSCORE was evaluated for each of the 100 extracted structures for each of the XIAP-peptide complexes. Thus, conformational changes of the receptor, which is used as a rigid structure in XSCORE calculations, have now been taken into account.

3.5. MMPBSA

MMPBSA (Molecular Mechanics Poisson Boltzmann Surface Area) [5] computes the binding free energy by using a thermodynamic path that includes the solvation contribution. The following expression now is used to describe the binding free energy:

$$\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^0 + \Delta G_{\text{RL}}^{0 \rightarrow \text{sol}} - \Delta G_{\text{R}}^{0 \rightarrow \text{sol}} - \Delta G_{\text{L}}^{0 \rightarrow \text{sol}} \quad (2)$$

Here, ΔG_{bind}^0 accounts for the free energy of binding *in vacuo*, and the rest of the terms are the solvation free energy of the receptor-ligand complex (RL), receptor (R) and ligand (L).

ΔG_{bind}^0 is decomposed into enthalpic plus entropic contributions, the first one being computed by the total energy of the force field and the second one computed usually by a normal mode analysis [28].

The estimation of entropic contributions is computationally intensive. For this reason, normal mode computations are carried out in the absence of water, with a distance dependent dielectric constant ($\epsilon = 4r$) and using a reduced system including only those protein atoms located within a predefined cut-off from the ligand atoms. The structures of the subsystems are minimized to a given gradient and the vibrational frequencies are computed for each of them. Moreover, for entropic estimation, it is widely accepted to work with only a few structures from the dynamics run, due the computational cost, while the rest of the terms are statistical balanced using about one hundred structures.

Each term of the solvation part of equation (2) is decomposed as follows:

$$\Delta G^{0 \rightarrow \text{sol}} = \Delta G_{\text{ele}}^{0 \rightarrow \text{sol}} + \Delta G_{\text{np}}^{0 \rightarrow \text{sol}} \quad (3)$$

Where, $\Delta G^{0 \rightarrow \text{sol}}_{\text{ele}}$ accounts for the electrostatic contribution to solvation. This term is obtained, as implemented in the MEAD program [29], by solving the linear Poisson Boltzmann equation in a continuum model of the solvent by the finite difference algorithm [30] using a 0.5 Å grid extended 20% beyond the solute. $\Delta G^{0 \rightarrow \text{sol}}_{\text{np}}$ accounts for the non-polar contribution to solvation, related linearly to the solvent accessible surface area (SASA) [31], computed in the present work through the LCPO method[32]:

$$\Delta G^{0 \rightarrow \text{sol}}_{\text{np}} = a \text{ SASA} + b \quad (4)$$

With $a=0.00542 \text{ kcal}/(\text{mol} \text{ \AA}^2)$ and $b= 0.92 \text{ kcal/mol}$. Parse radii were used for all atoms except for Zn^{2+} , which was used 2.0 Å [33]. For the calculation of ΔG^0_{bind} a parm99 Zn^{2+} was used [34].

All other constants of the MMPBSA methodology are set to standard values.

MMPBSA was performed using 100 snapshots obtained for each XIAP-peptide complex. Coordinates given by the complex structures were used to generate a separate set of structures for the XIAP and for the peptides, thus we use the one-trajectory protocol. This approximation avoids the calculation of separated trajectories for XIAP and peptides alone and supposes a little conformational change of the fragments. This approximation is always a source of error because some changes are expected upon binding especially for peptide ligands. However, this approximation is necessary in order to work in a similar way when comparing with XSCORE methodology, where conformational changes of the fragments upon binding cannot be taken into account.

4. RESULTS

The evolution of the total energy for the XIAP-Smac/DIABLO (9 residues) and XIAP-AVPY systems versus time throughout their molecular dynamics trajectories is shown in Figure 3. The first complex achieves a rapid convergence in the first 200ps while the XIAP-AVPY complex needs about 500ps to achieve energy convergence. All other XIAP-peptide complexes showed the same behavior. The different convergence rates is a reflection of the fact that initial conformation for the XIAP/Smac/DIABLO complex comes from a NMR structure while the rest were obtained by homology modeling.

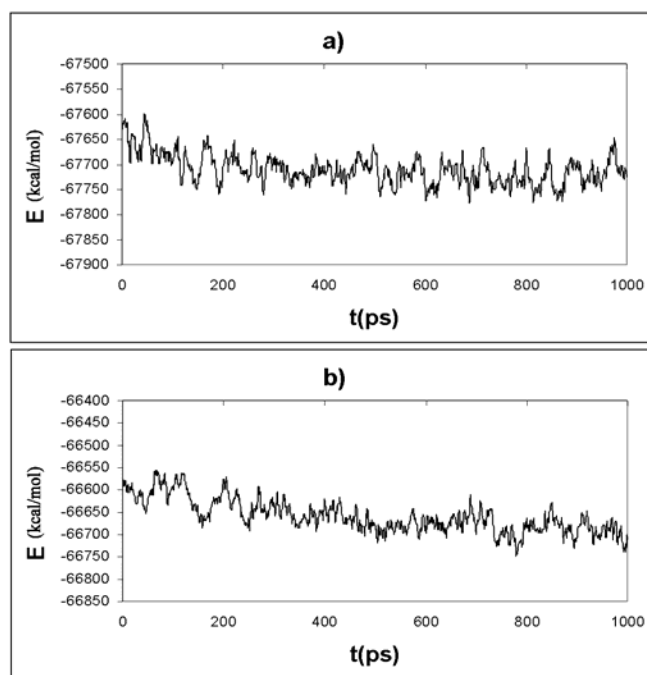


Figure 3: Evolution of total energy versus time for: a) the XIAP-Smac/DIABLO(9res) complex, b) the XIAP-AVPY complex.

Table 1 sets the experimental and averaged Zn-coordinated atom distances during the dynamics for the XIAP-Smac/DIABLO complex. It can be seen as a good agreement with respect to the experimental ones [17]. Similar results were obtained for the other XIAP-peptide complexes. Thus, the cationic dummy atom approach has proved its suitability for the molecular dynamic study of these systems.

	<i>Experimental distance/Å</i>	<i>Average distance/Å</i>
Zn-S(Cys 300)	2.10	2.12 (0.04)
Zn-S(Cys 303)	2.10	2.13 (0.04)
Zn-S(Cys 327)	2.10	2.14 (0.04)
Zn-N(His 320)	2.21	2.07 (0.05)
	<i>Experimental angle/°</i>	<i>Average angle/°</i>
S(Cys 300)-Zn-S(Cys 327)	108.5	113.9 (5.1)
S(Cys 300)-Zn-N(His 320)	107.2	105.7 (4.6)
S(Cys 300)-Zn-S(Cys 303)	107.5	106.1 (4.5)
S(Cys 303)-Zn-N(His 320)	113.1	113.8 (5.6)
S(Cys 303)-Zn-S(Cys 327)	108.6	111.4 (4.6)
S(Cys 327)-Zn-N(His 320)	111.7	101.7 (4.3)

Table 1: Experimental and average distances and angles of the zinc-coordinated atoms in XIAP-Smac/DIABLO complex, standard deviation in brackets.

Accordingly, 100 structures of receptor and ligand were extracted from the production period of the 1ns molecular dynamics trajectories, one snapshot each 8ps for the last 800ps for the experimental complex or one snapshot each 5ps for the last 500ps for the other complexes, and prepared for the evaluation of the binding free energy using XSCORE and MMPBSA methods. As described before, water molecules were removed from every snapshot.

Figure 4 shows the evolution of the binding free energy versus time (or snapshots) using both methods for the XIAP-AVPY system taken as an example (in MMPBSA the entropic estimation is not included). As can be seen, the property is well time stabilized. However, free energy fluctuation is about 0.4 kcal/mol for the XSCORE methods while it is much greater in the MMPBSA methodology (5.0 kcal/mol). Moreover, it gives negative and positive binding free energy values. This is an important fact, in the sense that for the first method, all molecular dynamics conformations show very close

energies while for the second slight changes in conformations are translated in great differences in binding, so this method seems more capable of identifying which peptides fit better in the binding pocket. The same behavior was noticed for the other systems.

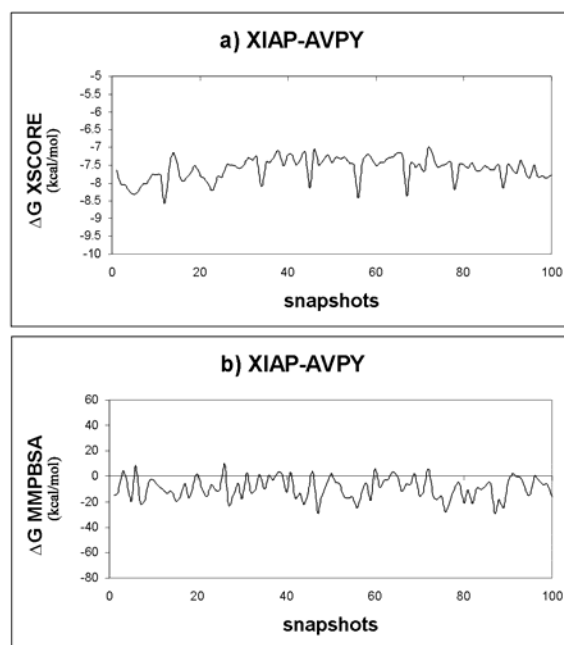


Figure 4: Evolution of binding free energy versus time for the XIAP-AVPY complex using: a) the XSCORE method, b) the MMPBSA method.

Table 2 lists the average results of the binding free energy calculations and the different energy contributions as the XSCORE function gives. The VDW term accounts for the van der Waals energy between the fragments computed with an 8-6 Lennard Jones potential. The HB term accounts for the number of optimal hydrogen bonds between the fragments. The HP, HM and HS terms account for hydrophobic effects and are related to optimum distances of hydrophobic contacts, to the environment of hydrophobic ligand atoms into the receptor and to the surface accessible area respectively. Finally, the RT term denotes the rotor or number of rotatable bonds predictably frozen during binding. The four following columns are the three scoring functions, HPScore, HMScore and HSScore depending on which hydrophobic effect has been taken into account, and the final XSCORE value as the average of the three scoring functions (in pK_a units). The last two columns show the calculated binding free energy and the experimental one in kcal/mol, using the relationship between the dissociation constant and free energy at 300K.

	VDW	HB	HP	HM	HS	RT	HP Score	HM Score	HS Score	^x SCORE	ΔG XSCORE/ kcal/mol	ΔG bind.exp/ kcal/mol
XIAP-Smac/ DIABLO (9res)	460.3	3.7	53.4	5.86	261.1	24	4.61	5.71	4.28	4.86	-6.67 (0.18)	-8.78
XIAP- AVPI	395.1	3.3	42.4	3.48	240.8	7	5.24	6.15	5.47	5.62	-7.71 (0.17)	-8.67
XIAP- ARPF	447.8	3.8	44.7	3.45	184.9	10	5.30	6.08	5.20	5.53	-7.59 (0.40)	-10.56
XIAP- AGPI	400.7	4.0	40.7	2.34	202.9	7	5.28	5.80	5.40	5.49	-7.53 (0.19)	-5.95
XIAP- AVPA	348.2	3.0	33.2	2.95	171.2	6	5.00	5.83	5.08	5.3	-7.27 (0.17)	-6.66
XIAP- AVPY	382.7	3.7	43.2	3.85	190.5	7	5.22	6.29	5.25	5.59	-7.67 (0.34)	-8.95
XIAP- AVPE	419.3	4.1	38.2	2.45	176.5	8	5.27	5.82	5.28	5.46	-7.49 (0.17)	-5.53

Table 2: Average data of XSCORE methodology. VDW accounts for van der Waals interaction, HB are the number of hydrogen bonds found, HP, HM and HS account for the hydrophobic effect, RT are the number or rotatable bonds of each ligand, HPScore, HMScore and HSScore are the three scoring functions, XSCORE is the average HPScore, HMScore and HSScore, ΔG_{XSCORE} is the binding free energy by using XSCORE method with its standard deviation in brackets and finally $\Delta G_{bind,exp}$ accounts for the experimental binding free energy.

XSCORE results are, in all cases, lower than the experimental ones with ARPF showing the highest error, 3 kcal/mol. XSCORE overscores low affinity peptides and underscores high affinity peptides, being all the calculated binding free energies in a reduced range of 0.44 kcal/mol, with the XIAP-Smac/DIABLO(9) complex as the only exception. This fact was noted previously when studying other protein-nonpeptide complexes [35]. The authors suggest that this is caused in part by a lack of enough penalty terms in XSCORE being the only one to count the number of rotatable bonds. Within this range XSCORE is not able to separate the most active ligands (XIAP-Smac/DIABLO, AVPI, ARPF and AVPY) from the least active ones (AGPI, AVPA and AVPE), which are orders of magnitude less active. So, the correlation coefficient is very small, only 0.02, which is a poor method performance.

Therefore, despite the fact that this empirical scoring function estimates the binding free energy within a 1 to 3 kcal/mol in a few minutes on a standard computer, it cannot be used to distinguish the most active ligands from the least active ones for the systems studied here, at least in its original form. Moreover, the AVPI peptide is predicted to have the highest affinity in contradiction to the experimental results.

However, it gives some insights into the ligand features, thus we see 3 to 4 intermolecular hydrogen bonds in all peptides and 6 to 24 frozen rotatable bonds during the binding process. One can analyze also what atoms are responsible for hydrogen bonds and the most important van der Waals contacts in each ligand if it is needed.

As XSCORE is composed of three functions, we can separately analyze each of them in order to see its performance when working alone. Table 3 shows the calculated binding free energy taking only into account one of the functions. Regarding the correlation coefficients, all three functions alone work better than the average final score, but HMScore gives the best results, having a correlation coefficient of 0.58 which is a good result for a scoring function [35]. Unfortunately, none of them is still able to clearly identify the most active peptides. However, if we do not take into account the XIAP-Smac/DIABLO(9) complex, HMScore suggests a value of $\Delta G_{\text{HMScore}}$ greater than 8.00 kcal/mol to separate the most active from the least active peptides. The AVPY peptide is now predicted to have the highest affinity. This behavior of XSCORE can be attributed to the inherent flexibility of the lateral chains in the peptides that are not always present in organic molecules, and the fact that XSCORE gives always close binding energies.

	$\Delta G_{\text{HPScore}}$	$\Delta G_{\text{HMScore}}$	$\Delta G_{\text{HSScore}}$	$\Delta G_{\text{bind,exp}}$
XIAP-Smac/DIABLO (9res)	-6.32 (0.21)	-7.83 (0.17)	-5.87 (0.18)	-8.78
XIAP-AVPI	-7.19 (0.19)	-8.44 (0.14)	-7.50 (0.18)	-8.67
XIAP-ARPF	-7.27 (0.45)	-8.34 (0.38)	-7.13 (0.41)	-10.56
XIAP-AGPI	-7.24 (0.19)	-7.96 (0.18)	-7.41 (0.20)	-5.95
XIAP-AVPA	-6.86 (0.19)	-8.00 (0.15)	-6.97 (0.18)	-6.66
XIAP-AVPY	-7.16 (0.36)	-8.63 (0.35)	-7.20 (0.37)	-8.95
XIAP-AVPE	-7.23 (0.17)	-7.98 (0.17)	-7.24 (0.18)	-5.53
Correlation coefficient	0.10	0.58	0.23	

Table 3: Average data of the three scoring functions included in XSCORE with its standard deviation in brackets. All values in kcal/mol units.

So far we have tested the average XSCORE data throughout the molecular dynamics, but we wanted to suggest other alternatives to this methodology. Table 4 shows the correlation coefficients taking into account the maximum and the minimum scores, both obtained along the molecular dynamic trajectory and the score obtained only from the minimization step, before the molecular dynamics simulation. Detailed values are not shown except for HMScores in Table 5, which is the best again. Here, using the maximum score, we have an improved correlation coefficient of 0.61 (Figure 5 a shows a plot of the experimental binding free energy versus this alternative HMScore function). In this case, if we do not take into account the XIAP-Smac/DIABLO(9) complex, HMScore suggest a value of $\Delta G_{\text{HMScore}}$ greater than 8.50 kcal/mol to separate most active peptides from least active ones. On the other hand, the minimized score, which is a cheap computational option, performs a third of all scores if we use the HM function. Moreover, the correlation coefficient decrease to 0.39 and it is not able to distinguish between the most and the least active peptides.

	<i>HPScore</i>	<i>HMScore</i>	<i>HSScore</i>	<i>XSCORE</i>
Maximum score	0.25	0.61	0.01	0.32
Minimum score	0.13	0.60	0.30	0.18
Minimized score	0.34	0.39	0.12	0.28

Table 4: Correlation coefficients of alternative scores.

	$\Delta G_{\text{HMScore maximum}}$	$\Delta G_{\text{HMScore minimum}}$	$\Delta G_{\text{HMScore minimized}}$	$\Delta G_{\text{bind,exp}}$
XIAP-Smac/DIABLO (9res)	-8.15	-7.34	-7.61	-8.78
XIAP-AVPI	-8.71	-8.08	-9.08	-8.67
XIAP-ARPF	-9.15	-7.67	-8.81	-10.56
XIAP-AGPI	-8.37	-6.97	-8.24	-5.95
XIAP-AVPA	-8.45	-7.60	-8.50	-6.66
XIAP-AVPY	-9.56	-7.93	-9.56	-8.95
XIAP-AVPE	-8.34	-7.32	-8.23	-5.53

Table 5: Binding free energies of the maximum, minimum and minimized HMScores. All values in kcal/mol units.

To achieve a much better XSCORE function it would be desirable to recalibrate some of the terms, particularly the rotor penalty, that is too large for the Smac/DIABLO (9 residues) peptide because the C-terminal part of this peptide is far from the binding pocket and consistently some of the bonds are free rotors. In fact, the rotor for this peptide should be about 7 instead of 24 like the rotor in the AVPI peptide. If we change the rotor penalty in the Smac/DIABLO (9residues) peptide to 7, it achieves the best score of all peptides, HPScore 5.65, HMScore 7.39, HSScore 5.85 and XSCORE of 6.30 (-8.64 kcal/mol). Now, this binding energy is close to the experimental one, but it is too high with respect to the other peptides so it does not improve the correlation.

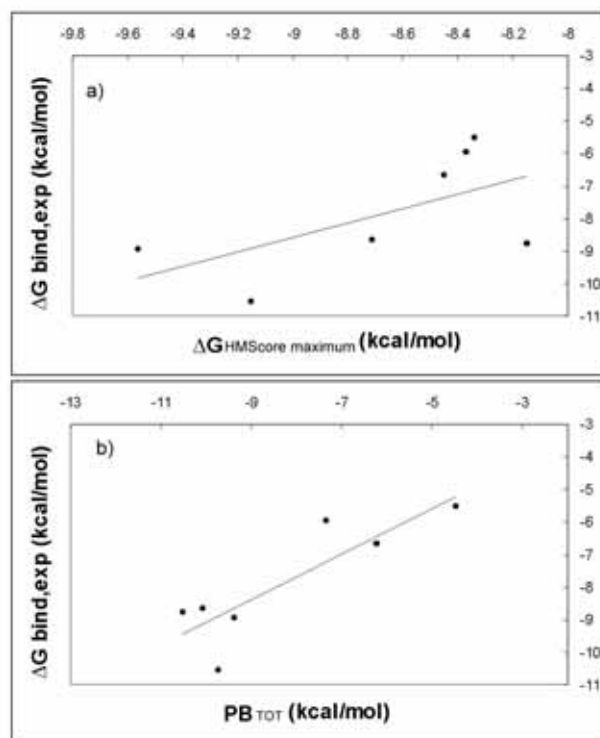


Figure 5: Plots of experimental binding free energy versus the maximum HMScore of XSCORE methodology (a) and versus the PB_{TOT} of MMPBSA methodology (b).

It is worth to note that, within the XSCORE framework, entropy variations are supposed to be included in both, the rotor and the constant terms. However, in the MMPBSA context, if absolute binding free energies are required, the entropic contribution must be determined in order to obtain meaningful results. Unfortunately, this value is extremely difficult to calculate because of the computational cost of

normal mode calculations. To get around this problem we should discuss relative binding free energies because in this case the entropic term is often assumed to cancel [5], as it would have to be in our context, being that our peptides are similar molecules in the same binding pocket.

Having this fact in mind, we will discuss our result without the inclusion of the entropic contribution and finally we will show how our results are affected by its inclusion.

Table 6 shows the average results (in kcal/mol) of the binding free energy calculations and the different energy contributions when applying the MMPBSA methodology, without the entropic contribution included. ELE accounts for the electrostatic interactions between the protein and the peptides that are responsible for large distance molecular recognition, VDW denotes van der Waals interactions between the fragments, that are related to complementary volume, and GAS accounts for the addition ELE+VDW+INT being the binding enthalpic contributions *in vacuo*. However, INT binding enthalpy, the balance of internal energy of the system, is always zero due to the one-trajectory protocol used in this work. PB_{SUR} accounts for the nonpolar contribution to solvation related to SASA, being always binding favorable, PB_{CAL} is the polar contribution of solvation, being in all cases unfavorable to binding, and PB_{SOL} denotes the PB_{SUR} + PB_{CAL} addition related to the balance of solvation total contribution. PB_{ELE} accounts for the PB_{CAL} + ELE addition, that is, the balance of favorable electrostatic interactions between the fragments and unfavorable desolvation of them. It is always a negative binding factor showing how crucial van der Waals recognition is. Finally, PB_{TOT} accounts for the total binding free energy calculated by the MMPBSA method and $\Delta G_{\text{bind,exp}}$ is the experimental binding free energy at 300K, for comparison.

MMPBSA shows a good performance, being the maximum error of 1.74 kcal/mol for the Smac/DIABLO (9residues) ligand and the correlation coefficient of 0.86 (Figure 5 b shows a plot of the experimental binding free energy versus the results obtained with the MMPBSA scoring function). Moreover, it is able to separate high affinity from bad affinity ligands (AGPI, AVPA and AVPE). A PB_{TOT} around 7.5 kcal/mol can be selected as a limiting value. However, ARPF which is the ligand with the highest electrostatic interaction with the XIAP protein and the most active peptide does not appear as the first. This peptide is a charged ligand and, as it has been noted previously, the use of a continuum electrostatic approach underestabilizes charged relative to zwitterionic ligands [36-39]. This fact is identifiable also in the AVPE peptide, with negative charge.

AVPA is the ligand with the lowest van der Waals interaction and AVPE is the ligand with the lowest

electrostatic interactions (followed by AGPI), being these two factors responsible for the poor affinity of both peptides. PB_{SUR} values are almost the same for each ligand and PB_{CAL} is the bottle neck feature of binding affinity.

	<i>ELE</i>	<i>VDW</i>	<i>GAS</i>	PB_{SUR}	PB_{CAL}	PB_{SOL}	PB_{ELE}	PB_{TOT}	$\Delta G_{bind,exp}$
XIAP-Smac/DIABLO (9res)	-179.57	-36.29	-215.86	-4.48	209.81	205.34	30.24	-10.52 (1.32)	-8.78
XIAP-AVPI	-154.41	-31.58	-185.99	-3.94	179.86	175.92	25.45	-10.07 (0.88)	-8.67
XIAP-ARPF	-302.32	-33.45	-335.76	-4.34	330.37	326.02	28.05	-9.74 (0.95)	-10.56
XIAP-AGPI	-139.27	-30.69	-169.96	-4.21	166.82	162.61	27.55	-7.35 (1.13)	-5.95
XIAP-AVPA	-146.17	-25.15	-171.31	-3.58	168.66	165.08	22.49	-6.23 (1.36)	-6.66
XIAP-AVPY	-174.42	-28.38	-202.80	-3.80	197.21	193.41	22.80	-9.39 (0.89)	-8.95
XIAP-AVPE	-124.74	-30.76	-155.50	-4.42	155.44	151.02	30.70	-4.48 (1.00)	-5.53

Table 6: Average data of MMPBSA methodology. *ELE* accounts for the electrostatic interactions, *VDW* denotes for van der Waals interactions between the fragments, *GAS* accounts for the addition $ELE+VDW+INT$ being the binding enthalpic contributions in vacuo, PB_{SUR} accounts for the non polar contribution to solvation, PB_{CAL} is the polar contribution of solvation, PB_{SOL} denotes the $PB_{SUR} + PB_{CAL}$, PB_{ELE} accounts for the $PB_{CAL} + ELE$ addition, PB_{TOT} accounts for the total binding free energy calculated by MMPBSA method with its standard error in brackets and $\Delta G_{bind,exp}$ is the experimental binding free energy. All values in kcal/mol units.

It is also interesting to study the influence of each mutation with respect to the original nine residues Smac/DIABLO sequence. By simply cutting the C-terminus, the AVPI peptide is obtained. This compound remains very active revealing the lack of interactions in this zone and confirming the point of view that AVPI is a conserved tetrapeptide motif and enough for binding. The mutation to glycine in the second position brings us the AGPI peptide. This peptide has few van der Waals and electrostatic interactions with the receptor. Thus, the valine in the second position gives hydrophobic contacts plus electrostatic ones, presumably placing correctly the hydrogen-bonding suggested experimentally [17].

The AVPA peptide allows us to notice the influence of mutations at the fourth position. Mutation to alanine implies that the van der Waals interactions will decrease around 5 kcal/mol. In fact this is the most pronounced change, being that this position is extremely important to hydrophobic recognition of both fragments.

In the same way, the glutamic acid in the AVPE peptide decreases the electrostatic interaction with the receptor being the peptide with the lowest affinity.

The AVPY peptide allows us to notice the influence of a large residue in the fourth position. Binding affinity does not change substantially with respect to the AVPI sequence, so we can see that the fourth position can allocate large and not flexible fragments like the tyrosine side-chain.

Finally ARPF has two point mutations, clearly the charged arginine at the second position increase the electrostatic contacts to -302.32 kcal/mol revealing the negative charge of the BIR3 domain of XIAP, but then the desolvation contribution is also more unfavorable, and maybe the MMPBSA notes this in extreme when using parse radii [40]. Phenylalanine at the fourth position increases the van der Waals contribution to binding more than tyrosine so the affinity increase with a large, completely apolar and not flexible residue like phenylalanine. Recognition of ARPF involves the higher electrostatic interaction of all peptides and the higher van der Waals interaction of all tetrapeptides.

ΔG_{bind} , values accounting for entropic contributions, are shown in Table 7, in kcal/mol units. As we stated before, introduction of entropic contributions is one of the more difficult aspects of the MMPBSA approximation. For this reason, we have analyzed the influence of usual approximations introduced in its calculation in order to reduce the computational complexity. The first factor considered was the size of the subsystem used to simulate the whole system. The results in table 7 have been obtained using the complete system because in our case this is computationally feasible. This is however an important factor, as can be seen from the results of the XIAP-Smac/DIABLO complex taken as a representative example. In this case, using a 12 Å cut-off value that includes 33 neighboring residues of the smac peptide, the calculated $T\Delta S$ value is -26.26 kcal/mol. Comparing with the full system result (Table 7), we obtain a difference of 6.53 kcal/mol arising basically from the vibrational entropy contribution, being $T\Delta S_{\text{vib}}$ -6.56 kcal/mol for the full system and -0.41 kcal/mol for the truncated one. Thus, selection of the appropriate cut-off is of great importance. The fact that vibrational entropy is the determinant part of the computed total entropy can be deduced easily from results reported in table 7, where contributions coming from translational and rotational entropy are mainly

constant among all the complexes. Another computational input in MMPBSA calculations is the value of the gradient root-mean-square deviation (rmsd) considered in the minimization procedure. Its value was set to 10^{-4} kcal/molÅ for all calculations in this work. However, its change to 10^{-5} kcal/molÅ did not produce any appreciable modification of the calculated entropic contribution. The last aspect studied here related to the entropy calculation is the number of snapshots used for its statistical treatment. To see how important is this number in the final entropic values, we performed three different calculations using five, ten and twenty snapshots extracted equally spaced from the production time for the XIAP-AVPA complex, which was selected as an example because of its high deviation of $T\Delta S_{\text{vib}}$. The obtained values for the vibrational contribution were 1.49, 3.02 and 1.27 kcal/mol with standard errors of 5.31, 2.77 and 1.67 respectively. As can be seen, convergence is similar to that usually observed for ΔG values, having a stable value for a reduced number of snapshots. The other two entropic contributions are approximately constant and the same general behavior was observed for all the other systems.

	$T\Delta S_{\text{TRAS}}$	$T\Delta S_{\text{ROT}}$	$T\Delta S_{\text{VIB}}$	$T\Delta S_{\text{TOT}}$	$PB_{\text{TOT}} - T\Delta S_{\text{TOT}}$	$\Delta G_{\text{bind,exp}}$
XIAP-Smac/ DIABLO (9res)	-13.85	-12.38	-6.65 (2.11)	-32.79 (2.11)	22.27	-8.78
XIAP-AVPI	-13.12	-10.91	7.80 (3.75)	-16.23 (3.77)	6.16	-8.67
XIAP-ARPF	-13.30	-11.27	-2.18 (2.36)	-26.75 (2.37)	17.01	-10.56
XIAP-AGPI	-13.02	-10.81	3.35 (1.03)	-20.48 (1.04)	13.13	-5.95
XIAP-AVPA	-13.02	-10.69	1.49 (5.31)	-22.22 (5.32)	15.99	-6.66
XIAP-AVPY	-13.22	-11.17	2.50 (4.41)	-21.89 (2.19)	12.50	-8.95
XIAP-AVPE	-13.15	-10.99	7.24 (4.55)	-16.90 (4.54)	12.42	-5.53

Table 7: Inclusion of entropy in MMPBSA methodology. Eight snapshots were used for the first complex and five for the other ones, $T\Delta S_{\text{TRAS}}$ is the translation entropy, $T\Delta S_{\text{ROT}}$ accounts for the rotational entropy, $T\Delta S_{\text{VIB}}$ denotes for vibrational entropy with its standard error in brackets, $T\Delta S_{\text{TOT}}$ accounts for the addition $T\Delta S_{\text{TRAS}} + T\Delta S_{\text{ROT}} + T\Delta S_{\text{VIB}}$, $PB_{\text{TOT}} - T\Delta S_{\text{TOT}}$ is the final free energy of binding and $\Delta G_{\text{bind,exp}}$ is the experimental binding free energy. All values in kcal/mol units.

5. CONCLUSIONS

In the present work, molecular dynamics simulations and binding free energies estimations of seven XIAP-peptide complexes with biological relevance in cancer disease have been performed. In order to determine their binding free energies, 100 snapshots from each 1ns molecular dynamics simulation were extracted and analyzed by XSCORE [8] and MMPBSA [5] methodologies.

XSCORE was originally developed to treat large ligand databases with rigid receptors structures without the need for a previous molecular dynamics run and without the need for addition of hydrogen atoms which are not present in crystal structures. However, we wanted to test this methodology in the same conditions as MMPBSA performs, that is, with a set of representative structures of each system, and using the same target with seven similar peptidic ligands.

Regarding the results, XSCORE was able to predict the binding free energy with a maximum error of 3 kcal/mol although with a very small correlation coefficient of 0.02. This value was substantially improved to 0.61 when using the HMScore function and maximum scores.

Although more work should need to be done in different systems to corroborate our results, the worst aspect of XSCORE was the ineffectiveness of identifying good and bad peptide inhibitors of XIAP in a large range of activity. This fact can be attributed to the mobility of peptide lateral chains during the molecular dynamics that is not reflected in the XSCORE values, putting all the conformations in a close binding energy range. One can overcome this fact, at least in part, by using our suggested maximum score together with the HMScore function. Nevertheless, taking into account these guidelines and the fact that this method is computationally very cheap, it can be a good choice to evaluate binding free energies as a first step in drug design. However, as indicated by the XIAP-Smac/DIABLO(9) complex results, it seems necessary to recalibrate some terms of the score function to take into account the existence of peptide residues that do not interact directly with the receptor.

MMPBSA is clearly a more robust method and in the particular case of these XIAP-peptide complexes, it predicted binding free energies in an accurate manner when no entropic contribution is added, being able to identify slight binding differences with a maximum error of 1.7 kcal/mol and a good correlation coefficient of 0.86. In this sense, MMPBSA can be used as a good scoring function, when only relative binding is important.

Unfortunately, in our case, the addition of entropic contributions moves away the absolute binding free

energies to positive values decreasing dramatically the correlation coefficient. As the total free energy is obtained with the addition of PB_{TOT} to the entropy, we could obtain negative bindings modifying the former contribution. For some authors the positive values of the total free energy of binding were mainly attributed to the use of parse radii instead of bondi radii [40] for the resolution of the Poisson-Boltzmann equation. Parse radii are smaller than bondi radii, which is translated into a higher desolvation penalty and smaller PB_{TOT} results. For other authors it was a consequence of the standard and different protocols used for MMPBSA calculations [41].

However, none of the methods is capable of answering why the ARPF ligand is so active, even more active than the original nine-residue sequence of Smac/DIABLO. It is difficult to understand this activity, and a crystal structure of ARPF in complex with XIAP would be of great help.

On the other hand, it is worth noting that we used experimental initial coordinates only for the first system, and the MMPBSA results for this one and for the other systems, modeled by homology, were correctly correlated to experimental data, indicating that active biological conformations have been located.

Finally we should point out that absolute correlation coefficients should be taken with caution given that there are only seven points in the regression.

Regarding the cationic dummy approach [18] for the treatment of the metal atom in XIAP, it was a proper choice because the four coordination of the zinc was maintained over the whole simulation with correct distances from the protein atoms. We think that the same methodology used to develop force field parameters for the zinc can be extrapolated, with necessary changes, to other metalloproteins.

6. REFERENCES

1. Stites, W. E. Protein-protein interactions: Interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **1997**, *97*, 1233-1250.
2. Jain, R.; Ernst, J.; Kutzki, O.; Park, H. S.; Hamilton, A. D. Protein recognition using synthetic surface-targeted agents. *Mol. Divers.* **2004**, *8*, 89-100.
3. Kollman, P. A. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.
4. Carlson, H. A.; Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **1995**, *99*, 10667-10673.
5. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Srivasan, J., Case, D. A.; Cheatam III, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.
6. Gohlke, H.; Kiel, C.; Case, D. A. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* **2003**, *330*, 891-913.
7. Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* **2003**, *68*, 16-34.
8. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11-26.
9. Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287-2303.
10. Shiraki, K.; Sugimoto, K.; Yamanaka, Y.; Yamaguchi, Y.; Saitou, Y.; Ito, K.; Yamamoto, N.; Yamanaka, T.; Fujikawa, K.; Murata, K.; Kano, T. Overexpression of x linked inhibitor of apoptosis protein in human hepatocellular carcinoma. *Int. J. Mol. Med.* **2003**, *12*, 705-708.
11. Tamm, I.; Kornblau, S. M.; Segall, H.; Krajewski, S.; Welsh, K.; Kitada, S.; Scudiero, D. A.; Tudor, G.; Qui, Y. H.; Monks, A.; Andreeff, M.; Reed, J. C. Expression and Prognostic Significance of IAP-Family Genes in Human Cancers and Myeloid Leukemas. *Clin. Cancer. Res.* **2000**, *6*, 1796-1803.

12. Yang, X.; Xing, H.; Gao, Q.; Chen, G.; Lu, Y.; Wang, Y.; Ma, D. Regulation of HtrA2/Omi by X-linked inhibitor of apoptosis protein in chemoresistance in human ovarian cancer cells. *Gynecol. Oncol.* **2005**, *97*, 413-421.
13. Kipp, R. A.; Case, M. A.; Wist, A. D.; Cresson, C. M.; Carrell, M.; Griner, E.; Wiita, A.; Albinak, P. A.; Chai, J.; Shi, Y.; Semmelhack, F.; McLendon, G. L. Molecular targeting of inhibitor of apoptosis proteins based on small molecule mimics of natural binding partners. *Biochemistry* **2002**, *41*, 7344-7349.
14. Glover, C. J.; Hite, K.; DeLosh, R.; Scudiero, D. A.; Fivash, M. J.; Smith, L. R.; Fisher, R. J.; Wu, J.; Shi, Y.; Kipp, R. A.; McLendon, G. L.; Sausville, E. A.; Shoemaker, R. H. A high-throughput screen for identification of molecular mimics of Smac/DIABLO utilizing a fluorescence polarization assay. *Anal. Biochem.* **2003**, *320*, 157-169.
15. Schimmer, A. D.; Welsh, K.; Pinilla, C.; Wang, Z.; Krajewska, M.; Bonneau, M.; Pedersen, I. M.; Kitada, S.; Scott, F. L.; Bailly-Maitre, B.; Glinsky, G.; Scudiero, D.; Sausville, E.; Salvesen, G.; Nefzi, A.; Ostresh, J. M.; Houghten, R. A.; Reed, J. C. Small-molecule antagonists of apoptosis suppressor XIAP exhibit broad antitumor activity. *Cancer cell* **2004**, *5*, 25-35.
16. Oost, T. K.; Sun, C.; Armstrong, R. C.; Al-Assaad, A.; Betz, S. F.; Deckwerth, T. L.; Ding, H.; Elmore, S. W.; Meadows, R. P.; Olejniczak, E. T.; Oleksijew, A.; Oltersdorf, T.; Rosenberg, S. H.; Shoemaker, A. R.; Tomaselli, K. J.; Zou, H.; Fesik, S. W. Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J. Med. Chem.* **2004**, *47*, 4417-4426.
17. Liu, Z.; Sun, C.; Olejniczak, E.; Meadows, R.; Betz, S.; Oost, T.; Herrmann, J.; Wu, J.; Fesik, S. Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* **2000**, *408*, 1004-1008.
18. Pang, Y. P.; Xu, K.; El Yazla, J.; Prendergast, F. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **2000**, *9*, 1857-1865.
19. Pang, Y. P. Novel zinc protein molecular dynamics simulations : steps towards antiangiogenesis for cancer treatment. *J. Mol. Model.* **1999**, *5*, 196-202.
20. Pang, Y. P. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins* **2001**, *45*, 183-189.
21. Oelschlaeyer, P.; Schmid, D. R.; Pleiss, J. Insight into the mechanism of the IMP-1 metallo- β -lactamase by molecular dynamics simulations. *Protein. Eng.* **2003**, *16*, 341-350.
22. Jorgensen, W. L.; Chandross, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

23. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Goud, I. R.; Mertz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P.A. A second generation of force fields for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
24. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. D.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Sligh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, Univ. California, San Francisco, 2002.
25. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald : an $W \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
26. Berendsen, H. J. C.; Postman, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. A. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
27. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *J. Comput. Chem.* **1977**, *23*, 327-341.
28. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.
29. Bashford, D.; Gerwent, K. Electrostatic calculations of the pK a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **1992**, *224*, 473-486.
30. Nicholls, A.; Honing, B. A rapid finite difference algorithm, utilizing successive overrelaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **1991**, *12*, 435-445.
31. Sitkoff, D.; Sharp, K.; Honing, B. Accurate calculations of hydration free energies using macroscopic solvents. *J. Phys. Chem.* **1994**, *98*, 1978-1988.
32. Weiser, J.; Shemkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217-230.
33. Sigfridsson, E.; Ryde, U. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.* **1997**, *19*, 377-395.
34. Hoops, S. C.; Anderson, K. W.; Merz K. M. Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262-8270.
35. Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind

refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114-2125.

36. Bonnet, P.; Bryce, R. A. Scoring binding affinity of multiple ligands using implicit solvent and a single molecular dynamics trajectory: application to influenza neuraminidase. *J. Mol. Graph. Mod.* **2005**, *24*, 147-156.

37. Bonnet, P.; Bryce, R. A. Molecular dynamics and free energy analysis of neuraminidase–ligand interactions. *Protein Sci.* **2004**, *13*, 946-957.

38. Smith, B. J.; Colman, P. M.; von Itzstein, M.; Daylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein Sci.* **2001**, *10*, 689-696.

39. Taylor, N. R.; von Itzstein, M. A structural and energetics analysis of the binding of a series of N-acetylneuraminic-acid-based inhibitors to influenza virus sialidase. *J. Comput. Aided Mol. Des.* **1996**, *10*, 233-246.

40. Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238-250.

41. Pearlman, D. A. Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a cogenic series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796-7806.

CHAPTER V: Searching for a human Transketolase Inhibitor



1. BRIEF INTRODUCTION

Transketolase, the most critical enzyme of the non-oxidative branch of the pentose phosphate pathway, has been described as a new target protein for cancer disease. Nevertheless, no crystal structure of the human Transketolase has been reported up to now and no selective inhibitors have been found. In this work, we have modeled a structure of the human Transketolase in order to find new inhibitors of this protein. As the sequence identity between the human Transketolase and the most similar crystal solved Transketolase, from yeast, is very low (27%), we have constructed the human protein only modeling the most conserved regions. This methodology has the advantage that the modeled zones will be more realistic although the entire protein will not be studied. These regions were focused on the interface between two subunits of the protein, in order to search for a new kind of inhibitors, disrupting the dimer stability of Transketolase and increasing their selectivity.

Results indicate that when the complete homology modeling is too difficult, one can study only the most conserved domains of a protein and even extract pharmacophore information. 3D database searching using our derived pharmacophores, docking and scoring protocols are also reported. Finally, experimental tests were carried out and two new Transketolase inhibitors were found. The most promising one was further optimized and tested, improving its activity.

2. CONTEXT

Transketolase catalyses the reversible transfer of two carbons, in a glycolaldehyde form, from a ketose donor substrate, to an aldose acceptor substrate. Moreover, Transketolase is the most critical enzyme of the non-oxidative branch of the pentose phosphate pathway (Figure 1). This pathway provides ribose molecules that are an essential metabolite in nucleic acid production. In addition, tumor cells require an important amount of ribose for their abnormal proliferation and it has been identified that these cells accomplish this requirement throughout the non-oxidative branch of the pentose phosphate pathway. Thus, metabolic control analysis calculated a Transketolase tumor growth control coefficient of 0.9 [1]. For these reasons, inhibition of Transketolase could lead to new drugs against cancer, decreasing the cell division level, acting at the most critical enzymatic step. Other studies, have been proposed Transketolase as a marker for Alzheimer's disease, because of a decreased activity of the protein in brain and other tissues of post mortem patients [2].

The mechanism of action of Transketolase, mediated by its cofactor thiamine pyrophosphate (Figure 2), is well known by several studies [3,4], however, drugs targeting the active centre of Transketolase, which act as cofactor analogs, have poor activity and low selectivity over other thiamine-dependent enzymes such as Pyruvate Dehydrogenase. Oxythiamine [5] and thiamine thiazolone diphosphate [6] (Figure 3) are typical examples of this kind of inhibitors, thus they do not have any pharmacological application. Recently, other studies have found interesting cofactor derivatives [7-9], with improved potency and pharmacokinetic properties as new Transketolase inhibitors.

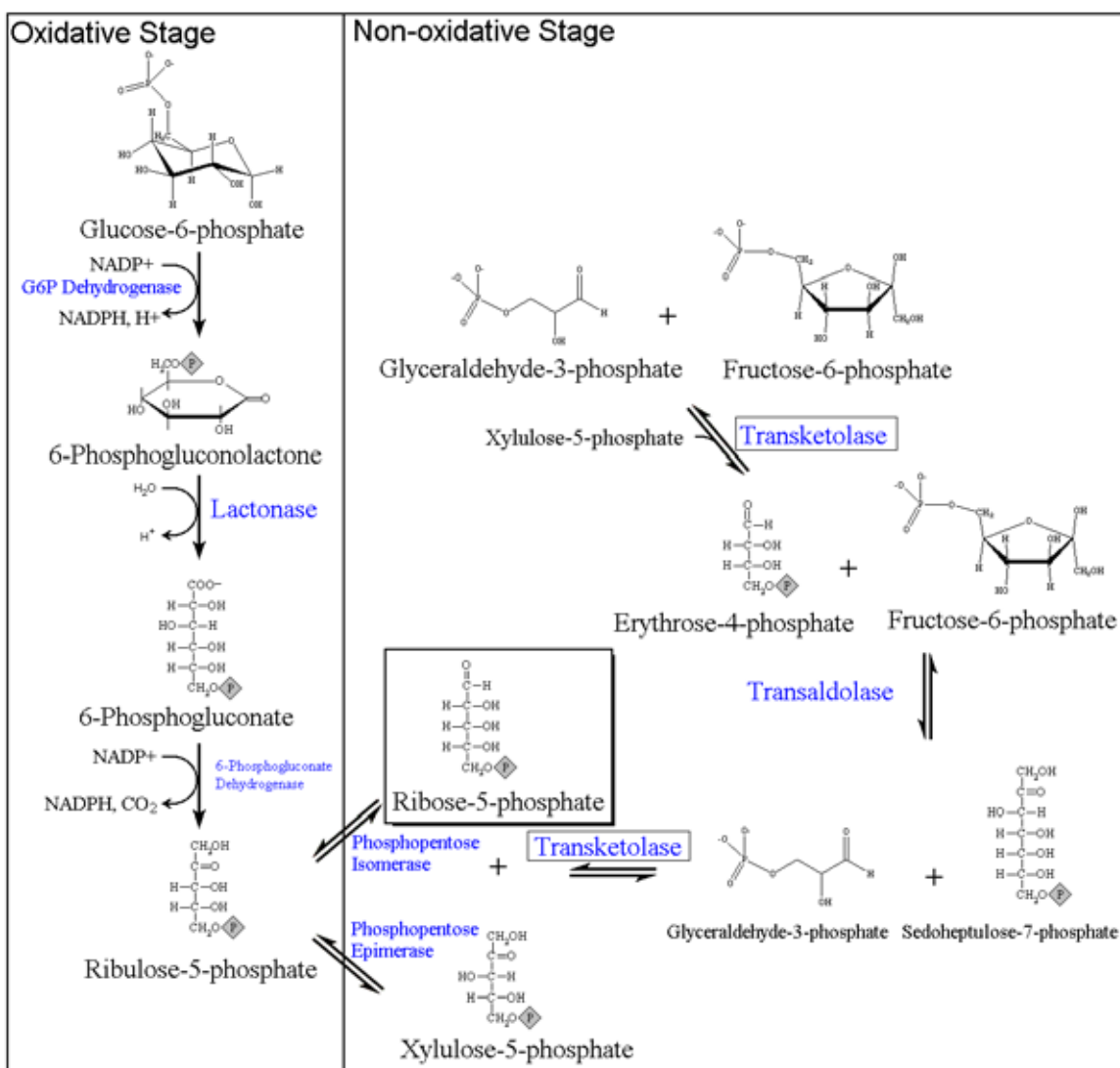


Figure 1: Non-oxidative and oxidative stages of the pentose phosphate pathway. Transketolase protein is remarked with a box. Ribose-5-phosphate, also remarked, is used for nucleic acid production.

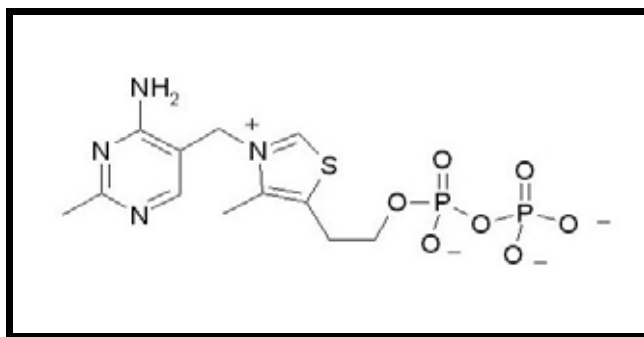


Figure 2: Transketolase cofactor, thiamine pyrophosphate.

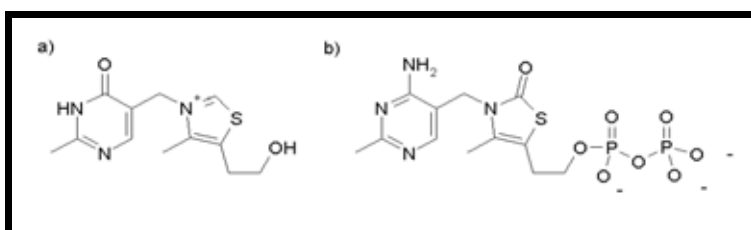


Figure 3: Two inhibitors of Transketolase. a) Oxythiamine, which active form is the pyrophosphate. b) thiamine thiazolone diphosphate.

A lot of work has been focused on Transketolase from yeast (*S. Cerevisiae*), E.Coli and maize, whose structures were solved by X-ray diffraction [10-12]. These studies revealed important aspects of the functional flexibility, metabolic profile and substrate binding of these variants.

Hence, the conserved sequence GDG(X...X)₂₅₋₃₀N, was identified as responsible for the cofactor and divalent metal binding. Two domains, called Pyr and PP domains, interact directly with the thiamine pyrophosphate cofactor, which adopts a characteristic v-shape conformation.

Nevertheless, few works are related to the human enzyme. A recent study of the human variant [13] identified the critical importance of aspartate 155, implicated in thiamine pyrophosphate binding. On the other hand, Du et al. [14] performed a high throughput screening on the human Transketolase finding two inhibitors, with an unknown mechanism of action. Interestingly, other authors discovered that some arginine residues (i.e. arginine 433) are crucial for Transketolase stability and activity, but only it was revealed for the rat variant [15]. In this scenario, a human structure of Transketolase would be desirable, in order to work in structure-based drug design. Nevertheless, the most similar Transketolase variant, from yeast, exhibits a sequence identity of only 27 %. So, the complete

homology modeling of human Transketolase is, in principle, a difficult task.

Therefore, we propose here a partial model of human Transketolase, taking the 3D structure of the yeast variant as a template, mutating only the most conserved zones and refining them by molecular dynamics simulation. This is a general strategy to obtain a protein model when no crystal structure is available, and it has been applied to several proteins with remarkable results [16-17]. Moreover, the dynamics simulation is used to identify some protein-protein hot spots at the dimer interface and to propose two pharmacophores in order to search for a human Transketolase inhibitor. Finally, experimental results are also reported.

3. METHODS

The initial 3D structure of yeast Transketolase homodimer was downloaded from the Protein Data Bank, with 1AY0 code [10]. Then, accordingly to the sequence alignment between the yeast and the human protein [18] (Figure 4), it was searched for the most conserved zones, to perform a homology modeling replacement. Obviously, one of the most conserved zones is the active centre, taking into account that both variants are thiamine dependent and catalyse the same reactions. Nevertheless, modeling of this zone and later search of inhibitors was not considered due to molecules to target this zone may be too similar to known inhibitors, such as oxythiamine or thiamine thiazolone diphosphate (Figure 3). These inhibitors have clear disadvantages, being not selective and low actives.

Two conserved zones were located at the interface between the dimer, one of them implicates the conserved R, which corresponds also to R in the rat variant that was found to be critical for rat Transketolase activity [15]. Therefore, yeast Transketolase was mutated to the human variant, performing a lateral chain substitution, only in these two interface zones identified. The first zone, formed by an alpha helix of a monomer, is recognized by a conserved loop of the other monomer (Figure 4). In addition, this loop is close to the active centre and to the cofactor, so it may be important not only for dimerization but also for the enzymatic activity. The sequence identity of the conserved loop is 57 %, while the sequence identity for the alpha helix containing R is 25 %.

The second zone is constituted by two antiparallel alpha helices of both monomers (Figure 4) with a sequence identity of 50 %. The final system can be seen in Figure 5, which shows the dimer structure of yeast Transketolase and the two modeled zones.

CONFIDENTIAL

Figure 4: Sequence alignment between yeast Transketolase (black) and human Transketolase (red) extracted from the multiple sequence alignment of [18]. Conserved and similar residues are shown in grey and light grey respectively. (***) denotes for the conserved alpha helix, (+++) denotes for the alpha helix containing R, (>>>) denotes for the conserved loop.

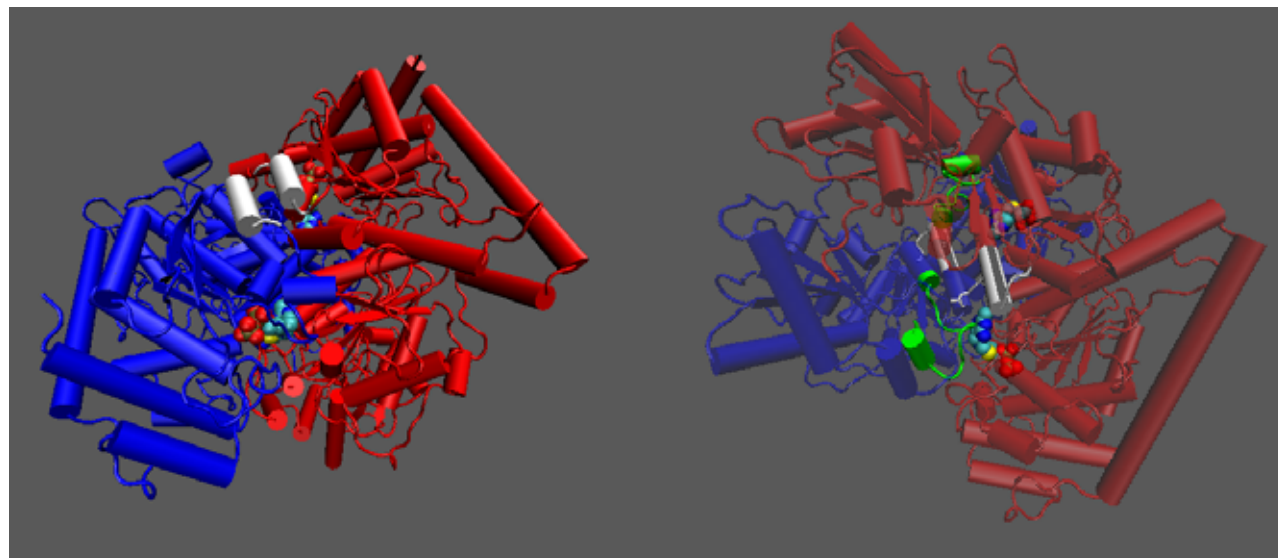


Figure 5: Dimer structure of the yeast Transketolase showing the modeled zones with the human sequence. **Left)** The antiparallel alpha helices are shown in white. **Right)** The alpha helices containing R are shown in white while the conserved loops are shown in light green. The TPP cofactor is marked in van der Waals spheres.

Once the mutations were performed in both zones, the system was prepared by using the Leap module of AMBER-7 package [19], adding counterions and TIP3P water molecules [20]. The force field parameters were extracted from parm94 [21]. The thiamine pyrophosphate cofactor parameters were treated adjusting charges to a HF/6-31+G** calculation using the GAUSSIAN package [22] while the other set of parameters were calculated by using the general amber force field (GAFF) [23]. The cutoff distance was kept to 9 Å to compute the nonbonded interactions. All simulations were performed under periodic boundary conditions, long-range electrostatics were treated by using the particle-mesh-Ewald method [24] and bonds involving hydrogens were constrained using the SHAKE algorithm [25]. The system was energy minimized in a multistep procedure during 56,000 iterations to a final structure with an energy gradient of 2.0 kcal/mol. The minimized structure was used as a starting point to the molecular dynamics simulation. Then, the system was heated up to 300 K, at a constant rate of 30 K / 10 ps coupling it to an external bath by means of the Berendsen 's algorithm [26]. In a pressure-constant step was equilibrated the density, and finally 1 ns of molecular dynamics was performed in the NVT ensemble. Last 500 ps, once the total energy was stabilized, were considered as the production time.

4. RESULTS

4.1. ANALYSIS OF INTERACTIONS

Production time dynamics was used to extract structural information of the contacts on the two human mutated interface zones. Thus, hydrogen bonds, van der Waals interactions and electrostatic interactions were followed using the Carnal and Anal programs of AMBER-7 package [19].

Most important interactions for both zones were grouped to form two pharmacophores in order to search for molecules with a similar pharmacophore and therefore, with the possibility to act disrupting the Transketolase homodimer. Table 1 shows the most stable hydrogen bonds found with respect to the contact zone formed between the alpha helix containing the critical R and the conserved loop. On the other hand, Table 2 lists the hydrogen bonds found between the antiparallel conserved helices.

<i>Alpha helix with R (monomer A)</i>	<i>Conserved loop (monomer B)</i>	<i>Average distance/ Å</i>	<i>Distance RMS/ Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>	<i>% Occupation</i>
R HH21	E OE1	1.90	0.75	131.4	45.0	78.7
R HH12	S O	1.87	0.79	138.5	49.9	86.7
R HE	E OE1	1.72	0.61	139.4	47.7	92.2
D OD1	T OH	2.29	0.77	138.9	29.3	67.7

Table 1: Hydrogen bonds between the alpha helix containing R and the conserved loop. % Occupation represents the per cent of simulation time in which the hydrogen bond is optimum (maximum N-O distance of 3.3 Å and hydrogen bond angle between $180 \pm 20^\circ$).

<i>Alpha helix (monomer A)</i>	<i>Alpha helix (monomer B)</i>	<i>Average distance/ Å</i>	<i>Distance RMS/ Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>	<i>% Occupation</i>
E OE1	K HZ	1.85	0.28	149.0	22.3	63.6
E OE2	Q H	2.02	0.24	153.0	16.6	68.0
K HZ	E OE1	2.46	0.73	111.9	53.1	54.4

Table 2: Hydrogen bonds between the antiparallel conserved alpha helices. % Occupation represents the per cent of simulation time in which the hydrogen bond is optimum (maximum N-O distance of 3.3 Å and hydrogen bond angle between $180 \pm 20^\circ$).

As can be seen, R of the first mutated zone, forms three stable hydrogen bonds that may maintain the structure of the conserved loop. The hydrogen bond between D and T (which is not located at the conserved loop) is not so important, presenting a low occupation. With respect to the interface zone formed by the antiparallel alpha helices, it was found three stable hydrogen bonds. Due to the symmetry between both chains, one must find four hydrogen bonds (two per chain), but the bond formed by Q of monomer A and E of monomer B was not so stable, with an occupation level lower than 50 %. This asymmetry may be a consequence of small differences in the initial structure of these residues, taking into account that usual molecular dynamics are not able to explore a lot of conformational space to restore the symmetry.

Moreover, Figure 6 shows the average van der Waals and electrostatic interaction energies between the alpha helix containing the conserved R, of the first monomer, and the whole second monomer (including the conserved loop). Concerning van der Waals interactions (Figure 6A), it appears F as the most important residue. R and D are also important due to their proximity to other residues to form the commented hydrogen bonds. I contributes slightly to the hydrophobic protein-protein recognition. Electrostatic energy (Figure 6B) is clearly driven by R, due to the three stable hydrogen bonds found.

Similarly, Figure 6 also shows the average van der Waals and electrostatic interaction energies between the antiparallel alpha helices of the first monomer and the complete second monomer (including its conserved antiparallel helix). Van der Waal recognition (Figure 6C) is directed almost by Q, which interacts with the same residue of the second monomer. Electrostatic interaction energies (Figure 6D) are more important for Q, K and E, as a reflex of their hydrogen bond pattern.

Residues with most favourable interactions are always pointing the same direction, towards the interaction zone of the other monomer.

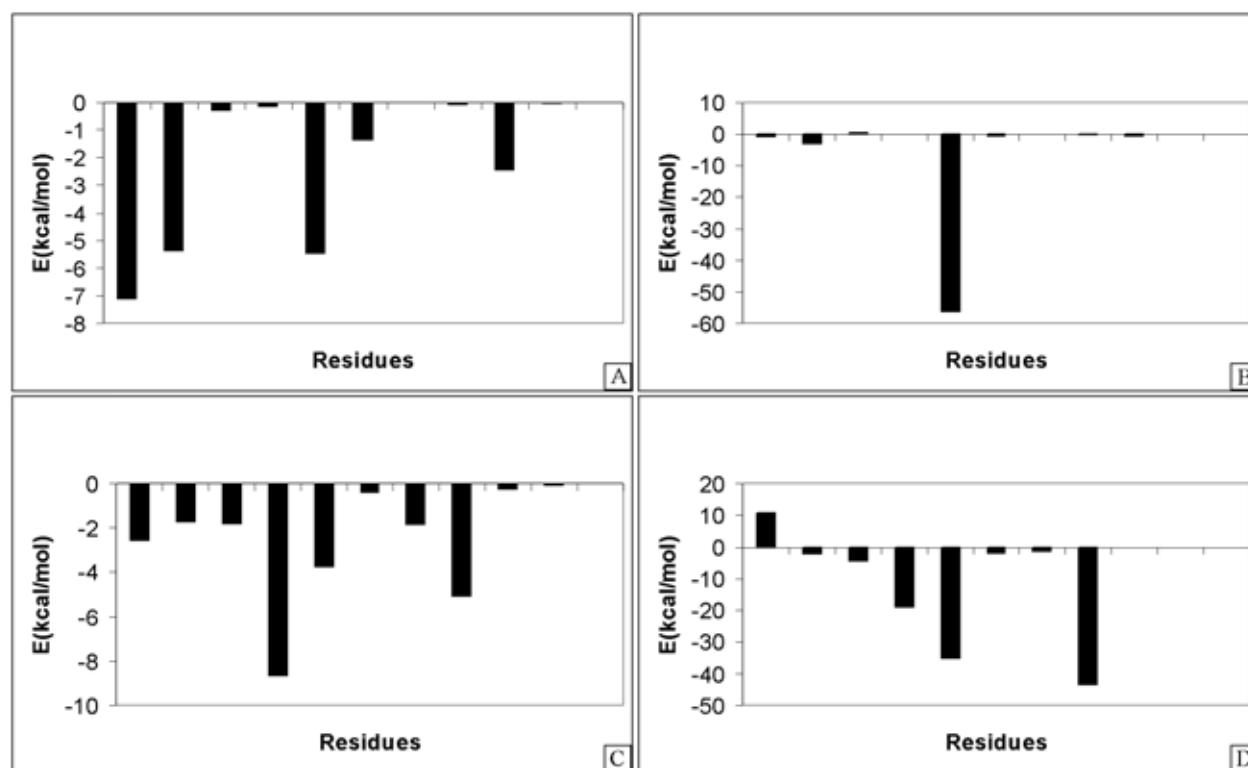


Figure 6: Average van der Waals (left) and electrostatic (right) interaction energies with respect to the alpha helix containing the critical arginine (A and B) and to the antiparallel conserved helix (C and D).

4.2. PHARMACOPHORIC HYPOTHESIS

Coupling last conclusions from hydrogen bonds, electrostatic and van der Waals analyses, it was suggested two pharmacophores for the interface recognition. The first one (Figure 7) involves the alpha helix with the conserved R and it is formed by 7 interaction points. Points 2, 3 and 7 describe the van der Waals interactions of F, D and I residues respectively. R was not included due its polar contribution is more important although its van der Waals interaction was also remarkable. Moreover, points 4, 5 and 6 describe the donor hydrogen bonds of R and point 1 identifies the acceptor hydrogen bond of D. It is a complex pharmacophore taking into account that the helix is only constituted by 11 residues.

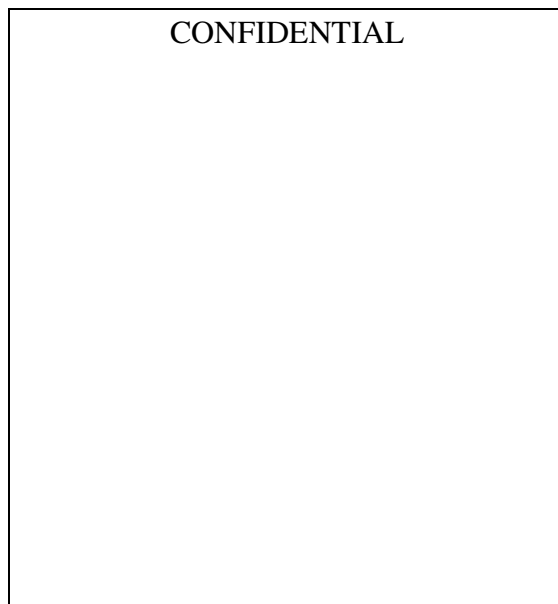


Figure 7: *Alpha helix containing R401 pharmacophore. Points 2,3 and 7 denote for van der Waals contacts, point 1 denotes for a hydrogen bond acceptor and points 4, 5 and 6 denote for hydrogen bond donors.*

In addition, Figure 8 shows the suggested pharmacophore for the antiparallel conserved helices. It is formed by 5 interaction points. A van der Waals interaction, point 2, achieved with Q; two hydrogen bond acceptors at the carboxyl group of E (points 4 and 5) and two hydrogen bond donors at the amino group of Q (point 1) and the sidechain amino group of K (point 3).

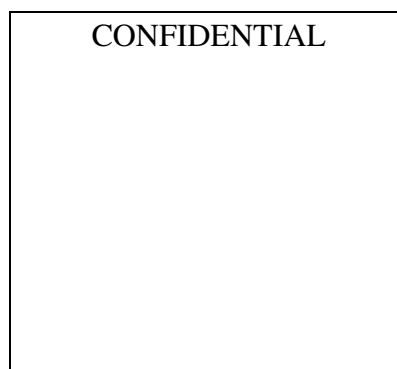


Figure 8: *Antiparallel alpha helices pharmacophore. Point 2 denotes for a van der Waals contact, points 4 and 5 denote for hydrogen bond acceptors and points 1 and 3 denote for hydrogen bond donors.*

Finally, Tables 3 and 4 list all the the distance parameters which identify the pharmacophores and their dynamical behaviour (maximum and minimum values throughout the simulation).

<i>Pharmacophoric Points</i>	<i>Average Distance/ Å</i>	<i>Maximum Distance/ Å</i>	<i>Minimum Distance/ Å</i>
CONFIDENTIAL			

Table 3: Geometrical data of the antiparallel alpha helices pharmacophore.

<i>Pharmacophoric Points</i>	<i>Average Distance/ Å</i>	<i>Maximum Distance/ Å</i>	<i>Minimum Distance/ Å</i>
CONFIDENTIAL			

Table 4: Geometrical data of the alpha helix containing R401 pharmacophore.

4.3. 3D DATABASE SEARCHING, DOCKING AND SCORING

The antiparallel alpha helix pharmacophore was used in order to find new inhibitors of Transketolase. It was used the CATALYST [27] software to search for commercial compounds containing the 5 points of the pharmacophore (Figure 8). The points 4 and 5 were grouped and searched as a carboxyl, phosphate or a nitrite group. Thus, 131 compounds were found to accomplish our pharmacophore and they were selected to evaluate them by the docking and the scoring protocols. Docking was performed by using our home-made program, Dock_Dyn [25] which directs the process adding the geometrical constrain of the interaction points. Van der Waals radii were decreased to their 60%, to achieve more docked molecules and to take into account slightly the receptor flexibility. Then, the RMS deviation between these points and the same interaction points in the Transketolase protein was used as a criterion of selection, fixing a maximum value of 3.3 Å. In addition, the XSCORE [29] semiempirical scoring function was also used to order them. Finally 8,887 conformations were docked and scored. The best 9 scored compounds (Figure 9) were purchased from the commercial sources (Sigma, Chembridge and Bachem companies). In addition, Figure 10 shows the best docking poses for compounds T1 and T2, taken as representative examples. These molecules recognize the residues involved in the interactions between the antiparallel alpha helices, Q, K and E.

Best values of RMS and XSCORE of these compounds, extracted from the docking and scoring protocols, are summarized in the Table 5.

<i>Compound</i>	<i>Best RMS/Å</i>	<i>Best XSCORE</i>	<i>Compound</i>	<i>Best RMS/Å</i>	<i>Best XSCORE</i>
T1	1.36	4.63	T6	1.59	4.78
T2	1.32	4.85	T7	2.11	4.82
T3	1.07	4.67	T8	1.28	4.64
T4	0.99	4.47	T9	1.65	4.65
T5	1.37	4.79			

Table 5: Best RMS and XSCORE values from the docking and scoring protocols, for compounds T1-T9.

On the other hand, the alpha helix containing R pharmacophore was also used to search in CATALYST [24] databases. In this case, points 2, 4, 5, 6, and 7 were selected, finding 170 molecules that accomplish the pharmacophore restrictions. Similar docking and scoring protocols were also carried out. The selection of the best compounds and the purchase of them was considered as future work.

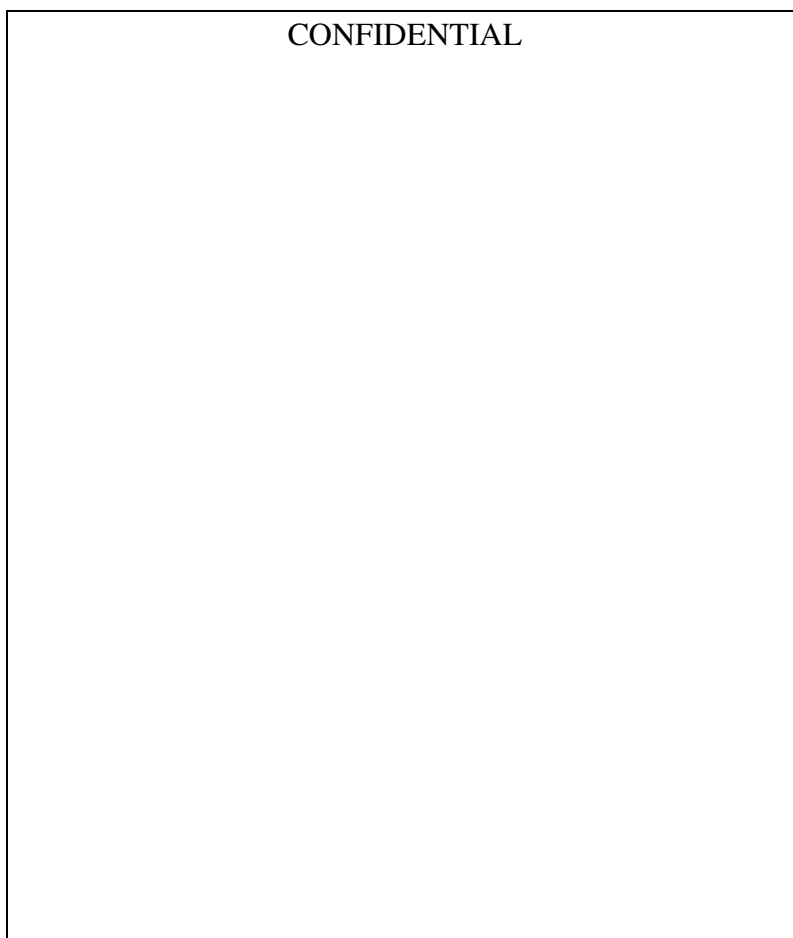


Figure 9: Purchased compounds.

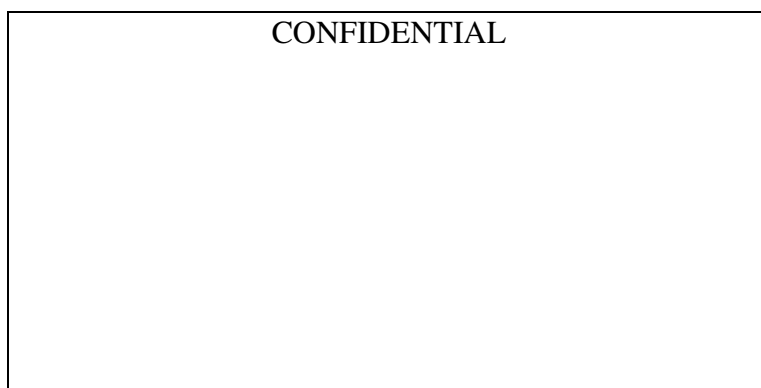


Figure 10: A) Best docking pose for compound T1 in complex with Transketolase, B) Best docking pose of compound T2 in complex with Transketolase. The protein is shown in transparent orange and carbon atoms of the ligand are marked in light green.

4.4. EXPERIMENTAL RESULTS

Prof. Cascante's group (Integrative Biochemistry and Cancer Therapy) and especially Gema Alcarraz-Vizán, at the University of Barcelona, carried out the experimental section in order to test the activity of the 9 compounds as human Transketolase inhibitors. Once they were solved in DMSO (or water for compound T1), they were firstly tested at human cell extract level, using the protein extract. Fluorescent intensity assays coupling the Transketolase inhibition on indirect changes on NADH concentration were performed. Compounds T1 and T2 were active with an IC_{50} (half maximal inhibitory concentration) around 500 μ M. Although they only exhibited moderate activity at cell extract level, they were also tested at cell level, taking into account that the inhibition of Transketolase should decrease dramatically the cell division. As we stated before, Transketolase is the most critical enzyme of the non-oxidative pentose phosphate pathway.

HT29 colon adenocarcinoma cells and HCT116 colon carcinoma cells were selected to perform the experimental assays at cell level.

Compound T1 has also a moderate activity at cell level, being its IC_{50} of 2mM. This concentration is too high to be considered T1 as a promising hit, nevertheless the known inhibitor oxythiamine (Figure 3a) inhibits Transketolase with a higher IC_{50} of 13mM. On the other hand, compound T2, has an IC_{50} of 10 μ M for both cell lines. This compound could represent an interesting new Transketolase inhibitor.

4.4.1. SECOND GENERATION OF HUMAN TRANSKETOLASE INHIBITORS

Taking the encouraging results of compound T2, different derivative molecules in order to improve the inhibition activity were designed. The docked structure of compound T2 was considered as a starting point to visualize different changes to improve the interaction pattern. Thus, the chlorine atom was considered not important and therefore substitutable. The nitrite group was maintained (or changed by a carboxylate group). The hydrophobic rings were varied adding a polar group such as a hydroxyl or a hydrophobic ethyl. Finally, the urea structure of compound T2 was identified as important to maintain hydrogen bonds with the protein. These possible modifications of T2 are summarized in Figure 11.

Finally, 16 derivative compounds were modeled, minimized at AM1 level using the GAUSSIAN [22] package and docked inside the Transketolase protein (using our Dock_Dyn software [28]). XSCORE function and RMS deviation were used again as a rank criterion.

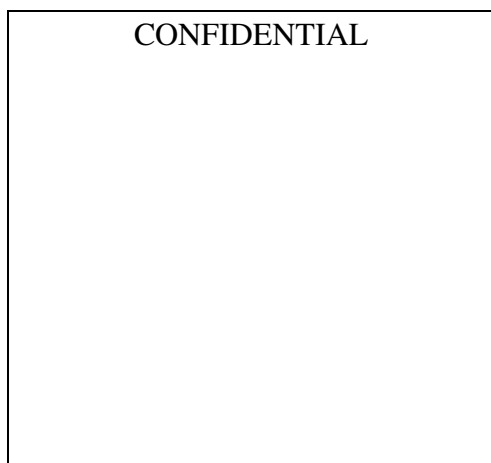


Figure 11: Possible modifications to design T2 derivatives. **Red box:** The nitrite can be replaced by similar groups. **Blue box:** The chloride can be removed. **Green box:** Hydrogen donor group can be added to interact with the carbonyl of Q. **Pink box:** The van der Waals interactions with Q can be improved adding non-polar groups.

<i>Compound</i>	<i>Best RMS/Å</i>	<i>Best XSCORE</i>	<i>Compound</i>	<i>Best RMS/Å</i>	<i>Best XSCORE</i>
1	2.22	4.67	9(T2B)	2.19	4.83
2	1.87	4.61	10 (T2E)	2.42	4.98
3	steric clash	steric clash	11(T2A)	2.08	5.23
4	2.91	4.68	12	1.70	4.96
5	2.22	4.61	13	2.23	4.95
6	2.19	4.96	14(T2C)	2.02	4.95
7	2.44	4.83	15(T2D)	2.07	4.98
8	2.22	4.71	16	2.13	4.68

Table 6: Best RMS and XSCORE values from the docking and scoring protocols, for 16 derivatives of T2.

Figure 12 shows the 16 derivative compounds selected and docked and Figure 13 shows the purchased derivatives which were ranked as the best ones. In addition Table 6 shows the best values of RMS and XSCORE for these selected derivatives.



Figure 12: Selected T2 derivative compounds.

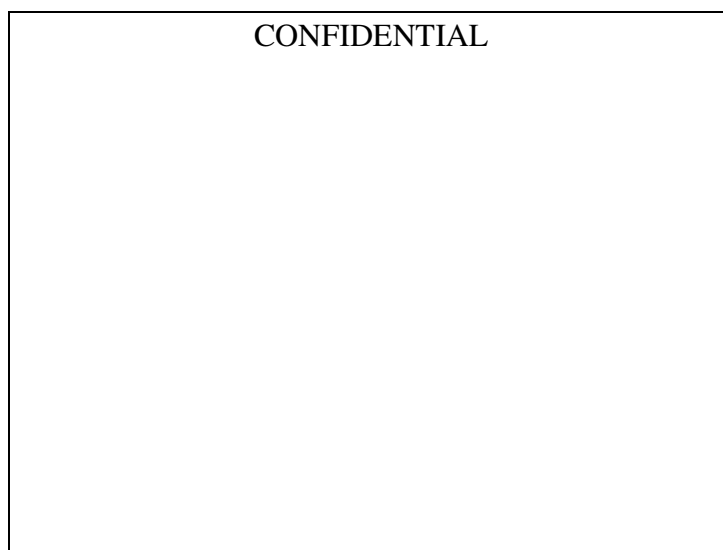


Figure 13: Purchased T2 derivative compounds (T2A-T2E).

For the five purchased compounds (T2-A to T2-E), it was performed the same Transketolase inhibition test, varying their concentration between 2 and 400 μM . Both cell lines HT29 and HCT116 were also studied. Results are summarized in the Table 7. Compounds T2-B and T2-E were poor actives while compounds T2-A and T2-D increase the activity with respect to the original molecule.

IC_{50} / μM	HT29	HCT116
Compound T2	10.0	9.0
Compound T2-A	6.0	11.5
Compound T2-B	200.0	200.0
Compound T2-C	20.0	13.0
Compound T2-D	15.0	6.5
Compound T2-E	>200.0	>200.0

Table 7: T2 derivative compounds activities. The activity of compound T2 is shown in the first row for reference.

5. CONCLUSIONS

A molecular dynamics study of the modeled human Transketolase has been presented here, finding new inhibitors of this protein. Following a partial modeling of the human protein, two conserved interface zones were mutated from the yeast variant and later refined by molecular dynamics. Hydrogen bonds, van der Waals and electrostatic contacts were followed through the simulation to configure the pharmacophores. Two pharmacophores (with 7 and 5 interaction points) were selected and 3D database searching was carried out for the 5-point pharmacophore. After, docking and scoring protocols to rank the most promising compounds, the best nine molecules were purchased. Experimental tests were then carried out, both at cell extract and cell level, indicating that two molecules (T1 and T2) were active at the micromolar range. Moreover, a second generation of T2 derivative compounds were designed, docked and purchased, finding two of them with improved potency. Best results were obtained with compounds T2-A and T2-D, whose activities were of 6.0 μM and 6.5 μM , for HT29 and HCT116 cell lines respectively.

In summary, a virtual screening protocol was carried out taking only a partial structure information of the target to obtain new human Transketolase inhibitors. More important, these molecules were designed to disrupt the dimeric structure of the protein by a novel binding mode, with the possibility to be more selective than the known Transketolase inhibitors. Future work will be based on identifying metabolic profiles and optimizing again the compounds.

Concerning the 7-point pharmacophore extracted from the dimerization zone, future work will be also focused on searching small molecules that mimic the helix alpha containing R.

6. REFERENCES

1. Comín-Anduix, B.; Boren, J.; Martinez, S.; Moro, C.; Centelles, J. J.; Trebukhina, R.; Petushok, N.; Lee, W. N.; Boros, L. G.; Cascante, M. The effect of thiamine supplementation on tumour proliferation. A metabolic control analysis study. *Eur. J. Biochem.* **2001**, *268*, 4177-4182.
2. Héroux, M.; Raghavendra Rao, V. L.; Lavoie, J.; Richardson, J. S.; Butterworth, R. F. Alterations of thiamine phosphorylation and of thiamine-dependent enzymes in Alzheimer's disease. *Metab. Brain. Dis.* **1996**, *11*, 81-88.
3. Shreve, D. S.; Holloway, M.P.; Haggerty, J. C. 3rd.; Sable, H. Z. The catalytic mechanism of Transketolase. Thiamin pyrophosphate-derived transition states for Transketolase and pyruvate dehydrogenase are not identical. *J. Biol. Chem.* **1983**, *258*, 12405-12408.
4. Schellenberger, A. Thiamin pyrophosphate binding mechanism and the function of the aminopyrimidine part. *J. Nutr. Sci. Vitaminol.* **1992**, 392-396.
5. Raïs, B.; Comin, B.; Puigjaner, J.; Brandes, J. L.; Creppy, E.; Saboureau, D.; Ennamany, R.; Lee, W. N.; Boros, L. G.; Cascante, M. Oxythiamine and dehydroepiandrosterone induce a G1 phase cycle arrest in Ehrlich's tumor cells through inhibition of the pentose cycle. *FEBS Lett.* **1999**, *456*, 113-118.
6. Nilsson, U.; Lindqvist, Y.; Kluger, R.; Schneider, G. Crystal structure of Transketolase in complex with thiamine thiazolone diphosphate, an analogue of the reaction intermediate, at 2.3 Å resolution. *FEBS Lett.* **1993**, *326*, 145-148.
7. Le Huerou, Y.; Gunawardana, I.; Thomas, A. A.; Boyd, S. A.; de Meese, J.; deWolf, W.; Gonzales, S. S.; Han, M.; Hayter, L.; Kaplan, T.; Lemieux, C.; Lee, P.; Pheneger, J.; Poch, G.; Romoff, T. T.; Sullivan, F.; Weiler, S.; Wright, S. K.; Lin, J. Prodrug thiamine analogs as inhibitors of the enzyme Transketolase. *Bioorg Med Chem Lett.* **2008**, *18*, 505-508.
8. Thomas, A. A.; Le Huerou, Y.; De Meese, J.; Gunawardana, I.; Kaplan, T.; Romoff, T. T.; Gonzales, S. S.; Condroski, K.; Boyd, S. A.; Ballard, J.; Bernat, B.; DeWolf, W.; Han, M.; Lee, P.; Lemieux, C.; Pedersen, R.; Pheneger, J.; Poch, G.; Smith, D.; Sullivan, F.; Weiler, S.; Wright, S. K.; Lin, J.; Brandhuber, B.; Vigers, G. *Bioorg Med Chem Lett.* **2008**, *18*, 2206-2210.
9. Thomas, A. A.; De Meese, J.; Le Huerou, Y.; Boyd, S. A.; Romoff, T. T.; Gonzales, S. S.;

Gunawardana, I.; Kaplan, T.; Sullivan, F.; Condroski, K.; Lyssikatos, J. P.; Aicher, T. D.; Ballard, J.; Bernat, B.; DeWolf, W.; Han, M.; Lemieux, C.; Smith, D.; Weiler, S.; Wright, SK.; Vigers, G.; Brandhuber, B. Non-Charged thiamine analogs as inhibitors of enzyme transketolase. *Bioorg Med Chem Lett.* **2008**, *18*, 509-512.

10. Wikner, C.; Nilsson, U.; Meshalkina, L.; Udekwu, C.; Lindqvist, Y.; Schneider, G. Identification of catalytically important residues in yeast Transketolase. *Biochem.* **1997**, *36*, 15643-15649.

11. Isupov, M. N.; Rupprecht, M. P.; Wilson, K. S.; Dauter, Z.; Littlechild, J. A. Crystal Structure of Escherichia coli Transketolase. To be Published.

12. Gerhardt, S.; Echt, S.; Busch, M.; Freigang, J.; Auerbach, G.; Bader, G.; Martin, W. F.; Bacher, A.; Huber, R.; Fischer, M. Structure and properties of an engineered Transketolase from maize. *Plant. Physiol.* **2003**, *132*, 1941-1949.

13. Wang, J. J.; Martin, P. R.; Singleton C. K. Aspartate 155 of human Transketolase is essential for thiamine diphosphate-magnesium binding, and cofactor binding is required for dimer formation. *Biochim. Biophys. Acta.* **1997**, *1341*, 165-172.

14. Du, M. X.; Sim, J.; Fang, L.; Yin, Z.; Koh, S.; Stratton, J.; Pons, J.; Wang, J. J.; Carte, B. Identification of novel small-molecule inhibitors for human Transketolase by high-throughput screening with fluorescent intensity (FLINT) assay. *J. Biomol. Screen.* **2004**, *9*, 427-433.

15. Soh, Y.; Song, B. J.; Jeng, J.; Kallarakal, A. T. Critical role of arg433 in rat Transketolase activity as probed by site-directed mutagenesis. *Biochem. J.* **1998**, *333*, 367-372.

16. Dalton, J.A.; Jackson, R. M. An Evaluation of Automated Homology Modelling Methods At Low Target-Template Sequence Similarity. *Bioinformatics.* **2007**, *23*, 1901-1908.

17. Patny, A.; Desai, P. V.; Avery, M. A. Ligand-supported homology modeling of the human angiotensin II type 1 (AT(1)) receptor: insights into the molecular determinants of telmisartan binding. *Proteins.* **2006**, *65*, 824-842.

18. Sundström, M.; Lindqvist, Y.; Schneider, G.; Hellman, U.; Ronne, H. Yeast TKL1 gene encodes a Transketolase that is required for efficient glycolysis and biosynthesis of aromatic amino acids. *J. Biol. Chem.* **1993**, *268*, 24346-24352.

19. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheathan III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. D.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Sligh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, Univ. California, San Francisco, 2002.
20. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
21. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Goud, I. R.; Mertz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C., Fox, T.; Caldwell, J. W.; Kollman, P.A. A second generation of force fields for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
22. Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. Gaussian, Inc., Wallingford CT, **2004**.
23. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general AMBER force field. *J. Comp. Chem.* **2004**, *25*, 1157-1174.
24. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an $W \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
25. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *J. Comput. Chem.* **1977**, *23*, 327-341.
26. Berendsen, H. J. C.; Postman, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. A. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

27. CATALYST™ (Accelrys Inc. USA).

28. Rubio-Martinez, J.; Pinto, M.; Tomas M.S.; Perez, J. J. Dock_Dyn: a program for fast molecular docking using molecular dynamics information. University of Barcelona and Technical University of Catalonia. Barcelona, 2005.

29. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11-26.

CHAPTER VI: Exploring the Dimerization Interface of Glucose-6-
Phosphate Dehydrogenase by Molecular Dynamics: Searching for
Interface Peptide Inhibitors



1. BRIEF INTRODUCTION

Glucose-6-phosphate Dehydrogenase (G6PDH) is an essential enzyme involved in the oxidative branch of pentose phosphate pathway. Similarly as Transketolase, it has been previously suggested that the inhibition of this enzyme could be a novel strategy for cancer therapy. G6PDH provides ribose molecules required for the DNA synthesis, and it is well known that cancer cells need for an increased number of nucleic acids. In addition, the production of nucleic acids in the tumour cell is poorly controlled allowing a drug targeting this pathway to be more effective.

In order to disrupt the dimer structure and to inhibit the protein, we report in this chapter the analysis by molecular dynamics of the interface contacts of the active dimer of the human G6PDH. We found some hot spots that could be covered by different short peptides, including an interesting cyclic peptide. We have performed the simulation of seven peptide-G6PDH complexes calculating the free energy of binding of each one to find peptidic candidates to be effective dimerization inhibitors of the human G6PDH. Experimental results are also reported supporting this computational design of interface peptides as an uncommon but effective strategy to disrupt the protein stability.

2. CONTEXT

Glucose-6-phosphate Dehydrogenase (G6PDH) is an NADP⁺ dependent enzyme that catalyses the transformation of D-glucose-6-phosphate to 6-phosphoglucono- δ -lactone in the first step of the pentose phosphate pathway [1] providing pentoses for nucleic acid synthesis and generating NADPH that protects the cell against oxidative stress (Figure 1). Being involved in the rate-limiting reaction of the oxidative branch of this pathway, several studies point that the inhibition of G6PDH can be considered a new strategy for the treatment of cancer disease [2], by limiting the synthesis of pentoses and therefore reducing the tumour growth. Up to now, methotrexate [3] and dehydroepiandrosterone (DHEA) [2] are the most important inhibitors of this enzyme, although the first one is not selective because it inhibits all NADP⁺ dependent enzymes and the second one is a steroid hormone, thus both have disadvantages to be considered as drugs. In this sense, it was discovered recently [4] a non-steroidal inhibitor of G6PDH, KPF-CoA, that could bind on the monomer surface, disrupting the native structure. Unfortunately, this molecule exhibited only moderate activity. In addition, the effects of few

drugs have been studied on G6PDH, thus the combined treatment with cefaperazone/sulbactam [5] and ampicillin/sulbactam [5] inhibits competitively and non-competitively with respect to the reaction substrate this protein, but the binding site of these antibiotics is unknown. Metamizol [6] is another non-competitive drug that inhibits the reaction substrate of G6PDH.

As the human crystal structure of G6PDH is available, structure-based drug design can be used to find new active molecules against this enzyme, focusing them on incrementing the activity and decreasing side-effects.

Our aim in this work is to disrupt the active homodimer of G6PDH studying the protein-protein interface at atomic level by a simulation of molecular dynamics taken the human X-ray solved structure of the enzyme. This approach of protein-protein disruption has been used to find inhibitors for enzymes such as HIV Protease, Reverse Transcriptase and Integrase [7], XIAP [8,9] and Herpesvirus protease [10] over others, and it could become a normal strategy to find new lead compounds. For a good review of protein-protein interface disruption using small molecules, see Wells and McClendon [11].

We report here a structural-based drug design to determine the important contacts that are responsible for the mutual recognition of both monomers (hot spots) and to postulate how seven interface peptides, ranging from 7 to 16 residues, could inhibit this protein dimer. Moreover, a designed cyclic peptide of 9 residues is considered as a privilege molecule that can cover the most important contacts found and inhibit G6PDH in an effective manner. The strategy to disrupt the enzyme dimerization by using short peptides derived from the protein sequence, was early identified [12] but it is not very usual.

The binding affinity of the seven peptides is predicted by the MMPB(GB)SA (molecular mechanics poisson boltzmann surface area) methodology [13], one of the most applied methods to determine free energy of binding. The results, for the cyclic peptide, as for six more interface peptides are discussed in this chapter. We argue here, that the decomposition of the interface surface of a protein dimer into several peptides, and the test of them, at theoretical and experimental stages, could be a general strategy to find small peptide inhibitors and it could be the first step into the design of non-peptidic compounds mimicking the most active interface peptides. In addition, preliminary experimental results supporting this idea are also reported.

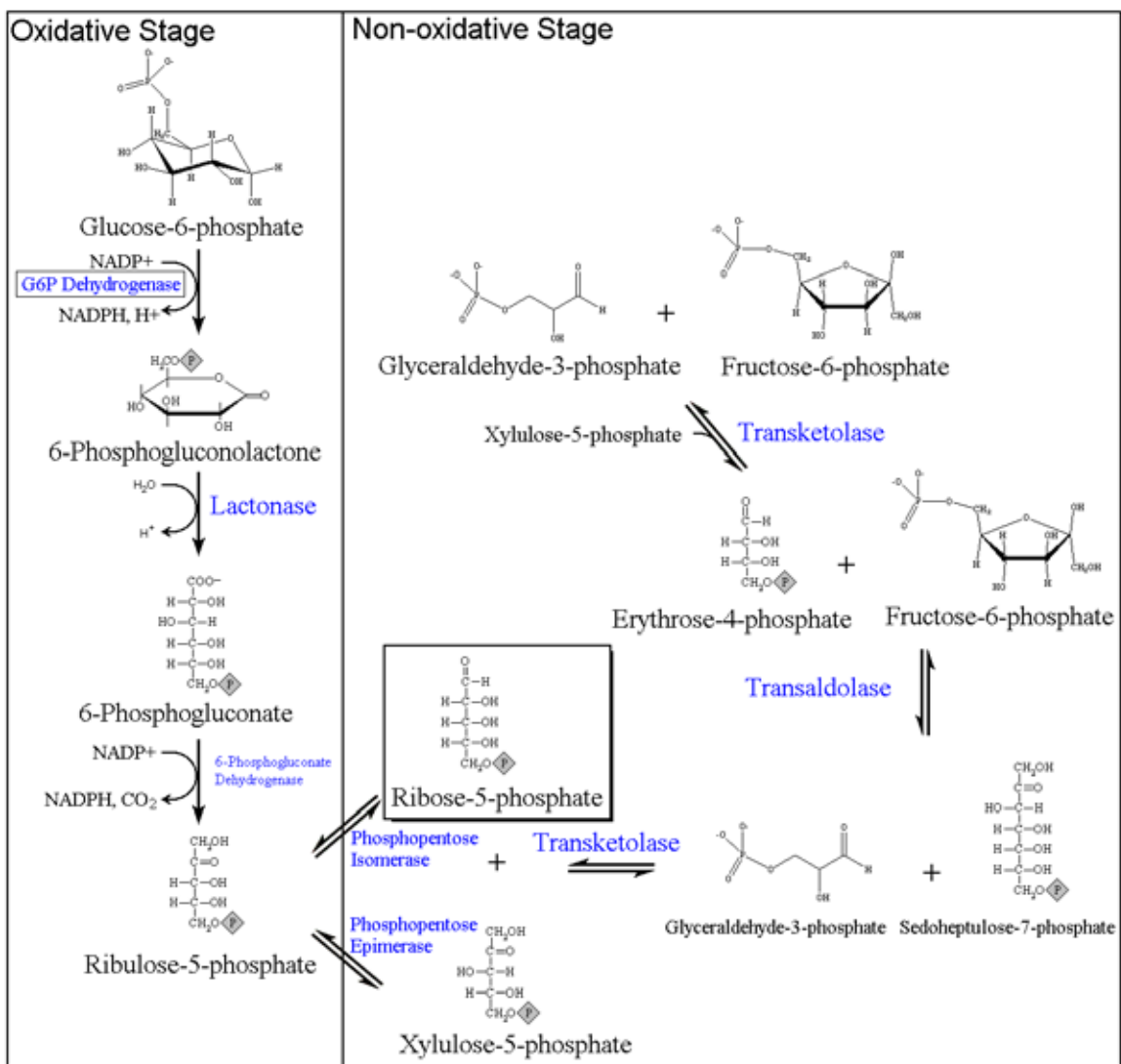


Figure 1: Non-oxidative and oxidative stages of the pentose phosphate pathway. Glucose-6-Phosphate Dehydrogenase protein is remarked with a box. Ribose-5-phosphate, also remarked, is used for nucleic acid production.

3. METHODS

All the calculations described in the present work were carried out at molecular mechanics level using parm94 force field [14] and AMBER7 [15] suite of programs.

Molecular dynamics were performed with the explicit solvent and a cutoff distance of 9 Å was selected for the non bonded terms. The systems were simulated under periodic boundary conditions and the particle-mesh-Ewald method [16] was used to treat the electrostatic interactions.

Prediction of the free energy of binding was carried out under the one-trajectory MMPB(GB)SA approximation [13] using a 0.5 Å grid extended 20% beyond the solute and computing the solvent accessible surface area through the LCPO (linear combination of pairwise overlap) method [17]. Poisson-Boltzmann equation was solved with the Solvate program of the MEAD package [18] using parse atomic radii set [19]. Generalized Born calculations were performed under the Tsui et al. parameters set [20].

All the structural figures of the present chapter were done with the VMD graphics program [21].

3.1. CONSTRUCTION OF THE G6PDH DIMER COMPLEX

The initial coordinates of the protein were taken from the Protein Data Bank, with 1QKI entry code [22]. This structure solved by X-rays is a multimeric protein, but only the active dimer (A and B subunits, Figure 2A) was selected due to the rapid dimer-tetramer equilibrium, which depends of the pH conditions. The Canton mutation of this structure was changed backwards (L459R) and the glycerol and glycolic acid molecules were removed while water molecules of the two subunits were not removed.

Residues 1 to 26 of subunit A and 1 to 27 of subunit B were not modeled because of the low resolution of the crystallographic data.

Leap program of AMBER7 [14] was used to construct the system, the two cofactors of the enzyme, NADP, were adapted to amber force field using the Ulf Ride [23] parameters considering a total charge of -3 for each one.

The final system is constituted by 2 monomers of G6PDH, 2 NADP⁺ cofactors, 11 Na⁺ counterion molecules and a cubic box of TIP3P waters [24] of 43,277 molecules.

3.2. MINIMIZATION AND MOLECULAR DYNAMICS OF THE G6PDH DIMER

The system was minimized to remove some steric stress by a multi-step procedure. First, water and Na⁺ molecules were allowed to move while the rest of the system was kept frozen. Second, side chains of the protein were relaxed as well as water and Na⁺ molecules. Third, the two NADP were also relaxed and finally all atoms were free to move. Using steepest descent following by conjugated gradient algorithms, the final minimized structure exhibited a maximum energy gradient of 1.1 kcal/mol achieved throughout 100,000 iterations.

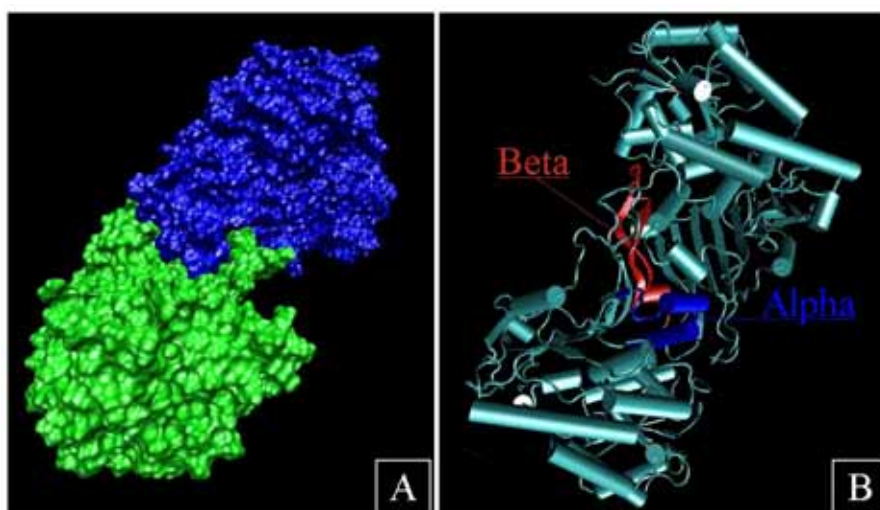


Figure 2: A) Structure of the G6PDH dimer. B) Same structure showing the alpha (blue) and beta (red) dimerization domains.

The minimized structure of G6PDH was considered as a starting point to perform the molecular dynamics at 300 K coupling the system to a thermal bath using Berendsen's algorithm, with a time coupling constant of 0.2 ps and a time step of 1 fs. A cutoff of 9 Å was used for the non bonded interactions and SHAKE algorithm [25] was used to constrain the length of bonds involving hydrogen atoms. The molecular dynamics simulation begins with a standard protocol, first the minimized structure is heat up to 300 K in 100 ps at a rate of 30 K/10 ps, keeping all atoms except water and Na⁺ molecules frozen. The second step is a 40 ps pressure-constant period to augment the density of the system with the solute atoms frozen. Finally, 1ns in the NVT ensemble was performed to extract the information about the recognition of both monomers. The last 600 ps, once the total energy was stabilized, were considered as the production time.

4. RESULTS AND DISCUSSION

4.1. HYDROGEN BOND ANALYSIS

The number of most important interface hydrogen bonds through the simulation were identified using the CARNAL program of AMBER package [15]. For this purpose, all residues of the subunit A, with respect to all residues of the subunit B, were analyzed to find the most important inter-monomer hydrogen bonds. For reasons of symmetry only the subunit A was analyzed.

In corroboration with experimental structure, the hydrogen bond pattern of the dimer interface is composed mainly of two domains (Figure 2B), the alpha domain, composed by two perpendicular alpha helices, and the beta domain composed by two antiparallel beta sheets joined by a turn.

Table 1 lists the 8 interface hydrogen bonds (3 hydrogen donors and 5 hydrogen acceptors) in the alpha domains found throughout the molecular dynamics as well as the relevant structural data of each one.

Taking into account the homodimeric structure of the enzyme, one can notice that there is not a completely symmetric hydrogen bond pattern, thus the two alpha helices are not ideal along the simulation revealing the dynamic behaviour of the interface. There are also some hydrogen bonds between both domains as defined above.

Table 2 lists, similarly, the 20 interface hydrogen bonds (10 hydrogen donors and 10 hydrogen acceptors) between the beta domains and the relevant structural data of each one.

It can also be noticed that some of the hydrogen bonds between the two domains (alpha and beta) are already described in the Table 1, with similar but not the same parameters, due to the two subunits, A and B, are not completely symmetric through the dynamics. However these small differences are not important to identify hot spots.

In corroboration with the X-ray structure [22], many charged residues are involved in highly stable hydrogen bonds along the dynamics simulation, however others also contribute to interface recognition, as the simulation confirms, and they have to be taken into account. Representative examples are shown in Table 1.

<i>H donors</i>	<i>H acceptors</i>	<i>%Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
(B) N-H (side chain)	(A) O (side chain)	91.8	1.95	0.17	154.1	12.3
(B) N-H (main chain)	(A) O (side chain)	98.5	1.89	0.12	160.4	9.6
(B) N-H (side chain)	(A) O (side chain)	63.6	2.28	0.67	123.5	50.8
(B) N-H (main chain)	(A) O (side chain)	97.2	1.89	0.13	159.7	10.0
(B) N-H (main chain)	(A) O (side chain)	66.6	2.24	0.35	148.7	13.3
(A) N-H (side chain)	(B) O (side chain)	69.8	2.17	0.25	143.9	12.8
(A) N-H (side chain)	(B) O (main chain)	97.8	1.97	0.14	160.4	9.7
(A) N-H (NH ₂ of side chain)	(B) O (main chain)	89.7	1.95	0.24	152.3	14.7

Table 1: Most important hydrogen bonds between the alpha domains, subunit A or B appears in brackets. Only subunit A was analyzed. The per cent of occupation factor is used to represent the time that the hydrogen bond is maintained optimum (maximum distance N O of 3.3 Å and hydrogen bond angle of $180 \pm 20^\circ$).

<i>H donors</i>	<i>H acceptors</i>	<i>%Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
(A) N-H (side chain)	(B) O (side chain)	94.8	1.95	0.15	157.1	11.7
(A) N-H (main chain)	(B) O (side chain)	96.8	1.94	0.14	160.5	10.7
(A) O-H	(B) O (side chain)	92.7	1.99	0.25	155.9	14.5
(A) N-H (main chain)	(B) O (side chain)	99.2	1.94	0.14	161.1	8.9
(A) N-H (main chain)	(B) O (side chain)	91.0	2.05	0.20	159.5	11.3
(A) O-H	(B) O (side chain)	78.6	2.14	0.30	150.5	15.0
(A) N-H (main chain)	1 (B) O (main chain)	70.3	2.24	0.19	160.6	7.5
(A) N-H (NH ₂ of side chain)	(B) O (side chain)	100.0	1.76	0.09	164.2	7.4
(A) N-H (NH ₂ of side chain)	(B) O (main chain)	70.6	2.04	0.25	142.6	12.6
(A) N-H (NH of side chain)	(B) O (side chain)	98.5	1.89	0.15	159.9	9.2

Table 2 (I): Most important hydrogen bonds between the beta domains, subunit A or B appears in brackets. Only subunit A was analyzed. The per cent of occupation factor is used to represent the time that the hydrogen bond is maintained optimum (maximum distance N O of 3.3 Å and hydrogen bond angle of $180 \pm 20^\circ$).

<i>H donors</i>	<i>H acceptors</i>	<i>%Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
(B) N-H (side chain)	(A) O (side chain)	58.8	2.28	0.29	142.3	13.0
(B) N-H (side chain)	(A) O (main chain)	96.5	1.98	0.15	159.1	9.9
(B) N-H (NH of side chain)	(A) O (side chain)	78.8	2.03	0.26	147.9	20.0
(B) O-H	(A) O (side chain)	78.6	2.13	0.29	148.4	14.5
(B) N-H (main chain)	(A) O (main chain)	98.8	1.93	0.12	161.1	9.6
(B) O-H	(A) O (main chain)	96.5	1.84	0.15	157.0	9.8
(B) N-H (NH ₂ of side chain)	(A) O (side chain)	100.0	1.78	0.10	164.3	7.2
(B) N-H (NH of side chain)	(A) O (side chain)	99.7	1.87	0.13	159.9	9.3
(B) O-H	(A) O (side chain)	90.2	1.94	0.48	161.5	13.8
(B) N-H (NH ₂ of side chain)	(A) O (main chain)	71.0	1.98	0.18	141.6	12.9

Table 2 (II): Most important hydrogen bonds between the beta domains, subunit A or B appears in brackets. Only subunit A was analyzed. The per cent of occupation factor is used to represent the time that the hydrogen bond is maintained optimum (maximum distance $N-O$ of 3.3 Å and hydrogen bond angle of $180 \pm 20^\circ$).

4.2. ELECTROSTATIC ANALYSIS

The ANAL program of AMBER package [15] was used to compute the electrostatic energy, without cutoff distance, using a coulombic term and a distance dependent dielectric function, along the simulation. Figure 3 shows the average energies for each residue of the alpha domain throughout the simulation. Only the sequence Glu to Ile is shown, because low interaction energies were found in the

other residues of this domain. Figure 3 also shows the electrostatic energies for the beta domain. For reasons of symmetry only the subunit A was analyzed.

This calculation is basically employed to reveal the residues involved in hydrogen bonds and especially in salt bridges. Thus, Glu and Arg are the most important residues of the alpha domain while Lys, Glu and Arg are the most important residues of the beta domain. High electrostatic energies are expected for charged residues, as this analysis confirms.

Concerning the monomer binding, one should note that these high interaction energies are always decreased by the desolvation energy penalty and therefore, the hydrophobic effect is usually a more important factor during the dimerization process.

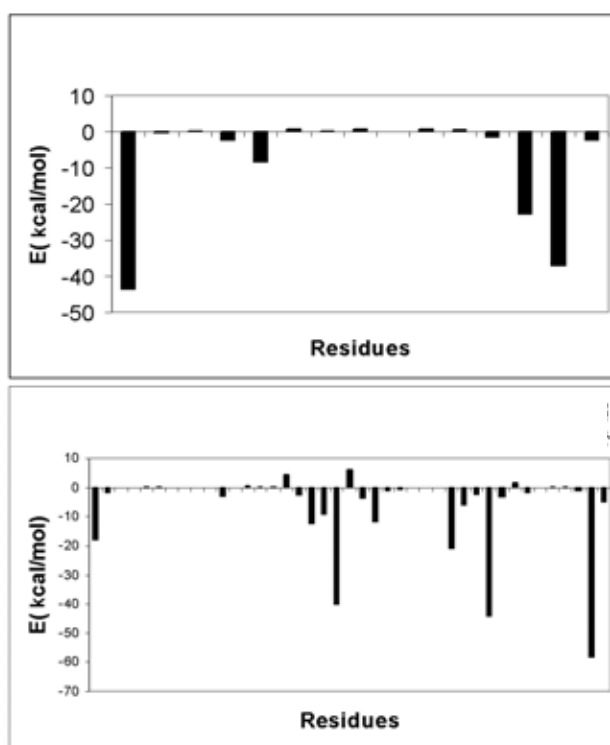


Figure 3: Average electrostatic energy for the alpha (up) and beta (down) domains of the G6PDH interface.

4.3. VAN DER WAALS INTERACTION ANALYSIS

The ANAL program of AMBER7 package [15] was used to identify the van der Waals interactions. It was analyzed the energy in the two interface subunits through the simulation without using cutoff distance. For reasons of symmetry only the subunit A was analyzed.

Average energies of each interface residue are shown in Figure 4. With respect to the alpha domain, only the sequence Glu to Ile is shown, because low interaction energies were found in the other residues of this domain. Figure 4 also shows the van der Waals profile for the beta domain.

In the first domain, Arg , Val and Ile are the most important hydrophobic residues.

In the beta domain, we have identified a much more hydrophobic core.

Hydrophobic contacts are an important factor concerning the protein dimerization (which can be seen as a process driven to reduce the solvent-accessible surface area) and therefore they can be used to identify the important zones to target by an interface peptide.

Special attention was focused on the beta turn composed by the sequence Lys to Phe, which connects the beta sheets of the whole beta domain. A cyclic peptide derived from this sequence, should be useful to disrupt the dimer, targeting this zone of the monomer surface. Moreover, coupling electrostatics and van der Waals analyses, this sequence maintains 1 salt bridge with Lys (Table 2, Figure 3), 4 hydrogen bonds of the main atoms with Lys and Gly (Table 2) and 4 hydrophobic contacts with Lys, Pro, Met and Phe (Figure 4), summing a total of 9 interaction points. Thr also interacts with the receptor by using two hydrogen bonds (Table 2), but it is not composing the beta turn.

It is worth to remark that a short beta turn of 7 residues maintains the most important interaction energies on the interface surface.

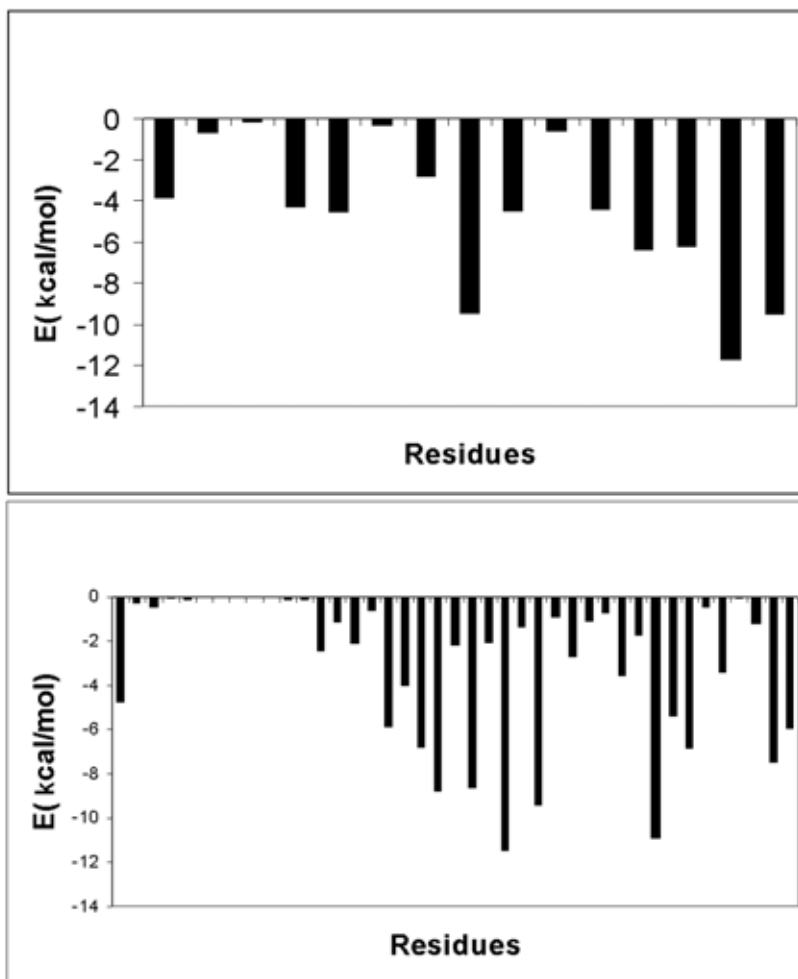


Figure 4: Average van der Waals energy for the alpha (up) and beta (down) domains of the G6PDH dimer interface.

Given the results of the analysis of the dimer interface, it was selected seven new peptide-G6PDH systems to search for an effective inhibitor of dimerization. Moreover, a cyclic peptide derived from the sequence Lys-Phe was considered as a special candidate.

The seven designed peptides were the following:

- **Peptide 1**, 16 residues.
- **Peptide 2**, 13 residues.
- **Peptide 3**, 14 residues.
- **Peptide 4**, 7 residues.

- **Cyclic peptide**, 9 residues.

(Based on peptide 4 and adding two additional Gly to close the cycle).

- **Peptide 5**, 12 residues.

- **Peptide 6**, 10 residues.

Peptides 1 and 2 were designed to cover the alpha helix domain while peptides 3, 4, 5, 6 and the cyclic peptide were designed to cover the beta domain.

4.4. CONSTRUCTION OF THE PEPTIDE-G6PDH COMPLEXES

All the peptide-G6PDH complexes were constructed from the minimized G6PDH dimer structure, cutting one of the monomers to achieve the desired peptide (Figure 5). The systems include the corresponding peptide, the monomer of G6PDH with the cofactor, Na⁺ counterion molecules and 18,000 TIP3P water molecules [24] approximately to solvate each system.

Then, a minimization step was carried out to achieve maximum energy gradients between 2.88 kcal/mol and 15.63 kcal/mol by using 50,000 iterations. Taking as the starting point the final minimized structure, 2 ns of Molecular Dynamics, within the NVT ensemble, were performed in each peptide-G6PDH complex, extracting all the structural information along the last nanosecond, considered as the production time.

4.5. CONSTRUCTION OF THE CYCLIC PEPTIDE-G6PDH COMPLEX

The cyclic peptide-G6PDH complex was constructed from the minimized G6PDH dimer structure, cutting one of the monomers to achieve the desired peptide and adding two glycines to close the sequence belonged to the beta domain. Finally, the new amide bond was created to close the peptide (Figure 6). For this purpose, it was used the Leap preparatory program of AMBER package [15].

The new system, consisting in the 9-residue cyclic peptide with 2+ charge, one monomer of G6PDH with the cofactor, 4 Na⁺ molecules and 18,600 TIP3P waters [24] was energy minimized to a structure

with a 6.8 kcal/mol final maximum gradient likewise the G6PDH dimer, and finally it was performed 2.4 ns of molecular dynamics at 300 K in the NVT ensemble. The last 500 ps were used to extract the information.

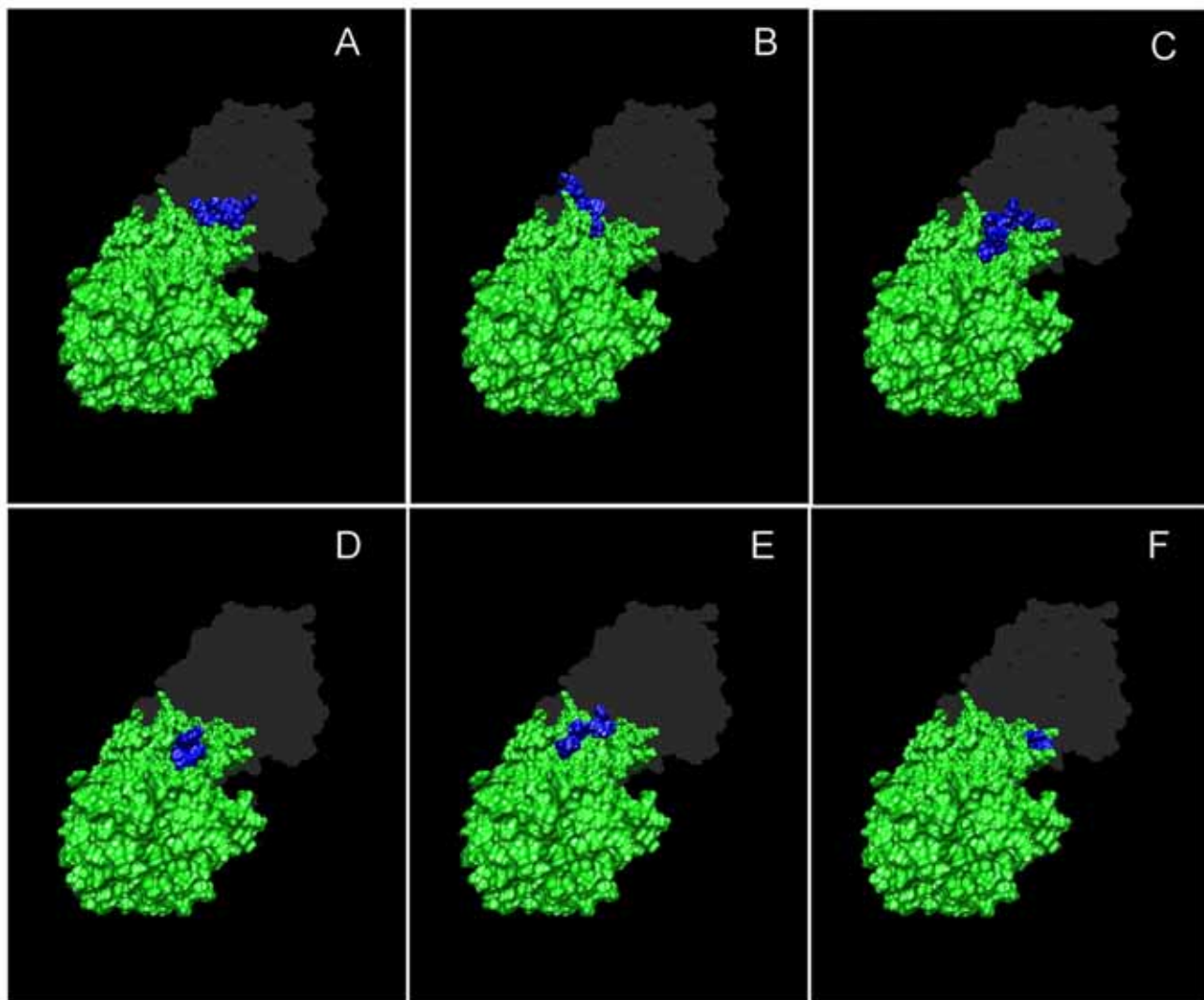


Figure 5: Peptide-G6PDH complexes, A: Peptide 1, B: Peptide 2, C: Peptide 3, D: Peptide 4, E: Peptide 5, F: Peptide 6. The whole second monomer is coloured in dark grey for reference.

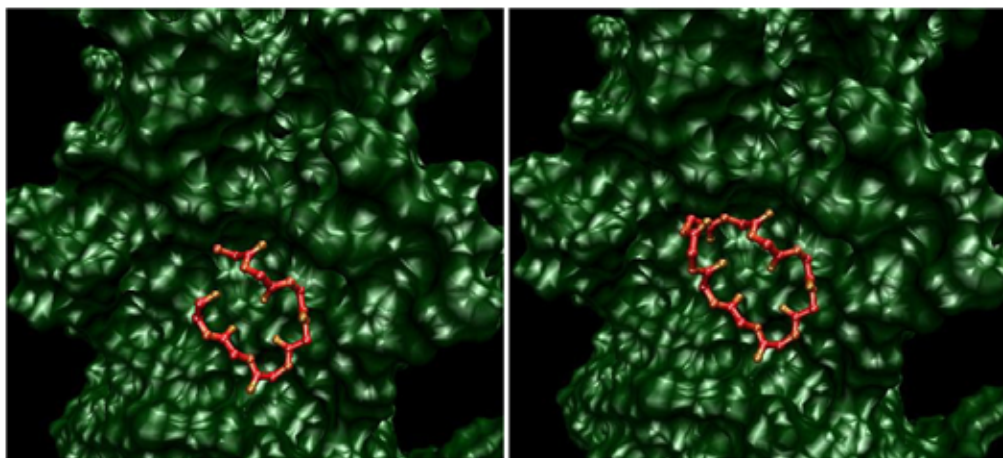


Figure 6: Construction of the Cyclic Peptide-G6PDH complex. Peptide 4 (left) and Cyclic peptide (right) adding two glycines. Only the backbone is shown.

4.6. INTERACTION ANALYSIS OF THE PEPTIDE-G6PDH COMPLEXES

Similarly to the G6PDH dimer analysis, an interaction analysis of the seven peptide-G6PDH complexes was performed. In order to find the most important contacts, hydrogen bonds, electrostatics energies and van der Waals energies were followed throughout the production time of the molecular dynamics. In this section, a discussion of the most important binding profiles of each peptide is presented.

Table 3 lists the hydrogen bond found for the peptide 1-G6PDH complex, as well as the average structural data of the simulation (peptide residues numbered from 1 to 16). All four bonds were detected in the G6PDH dimer system, so they were maintained after the peptide-G6PDH complex construction.

Similarly, Figure 7 shows the average van der Waals and electrostatic interaction of each peptide residue interacting with the G6PDH monomer. The interaction is mostly located in the C-terminal part of the peptide 1, similarly as the interaction profile found in the G6PDH dimer system.

Table 4 lists, in the same way, the hydrogen bond information extracted from the peptide 2-G6PDH complex. Peptide residues are numbered from 1 to 13. It appears three new hydrogen bonds, the first one, residue 1 with His, is composed by the N-terminal hydrogen of the first peptide residue so it cannot be identified in the G6PDH dimer system. The same can be noticed for the hydrogen bond between residue 13 and Arg, formed by the oxygen terminal atom of the peptide. In addition, a new

hydrogen bond is located between residue 2 and Asp. In addition, Figure 8 shows the average van der Waals and electrostatic energies of each peptide residue with respect to the whole G6PDH monomer. Most important contacts are distributed along the N-terminal and C-terminal zones of peptide 2.

<i>Peptide 1</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 14 ND2-HD21	O	80.2	1.97	0.17	158.6	9.6
Residue 14 OD1	ND2-HD21	73.2	2.03	0.24	156.1	13.9
Residue 15 NH1-HH11	O	72.8	1.95	0.24	154.0	12.8
Residue15 NH	OD1	64.6	1.94	0.14	153.1	13.1

Table 3: Average data of the hydrogen bonds found in the peptide 1-G6PDH complex.

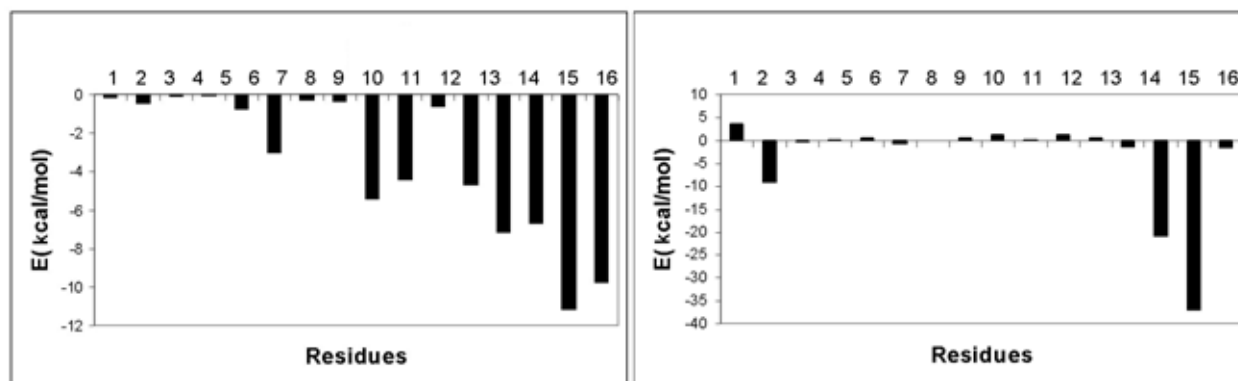


Figure 7: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Peptide 1-G6PDH complex.

<i>Peptide 2</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 1 N-H1	O	41.8	2.21	0.53	135.4	35.5
Residue 2 NH2-HH22	OD1	41.8	2.21	0.55	139.4	24.1
Residue 2 NH1-HH11	O	80.0	2.00	0.49	157.2	10.9
Residue 2 N-H	OD1	75.8	1.95	0.15	156.1	10.0
Residue 13 OXT	NH2-HH21	69.0	1.87	0.17	154.2	10.7
Residue 1 OD1	ND2-HD21	49.2	2.25	0.44	148.3	13.7

Table 4: Average data of the hydrogen bonds found in the peptide 2-G6PDH complex.

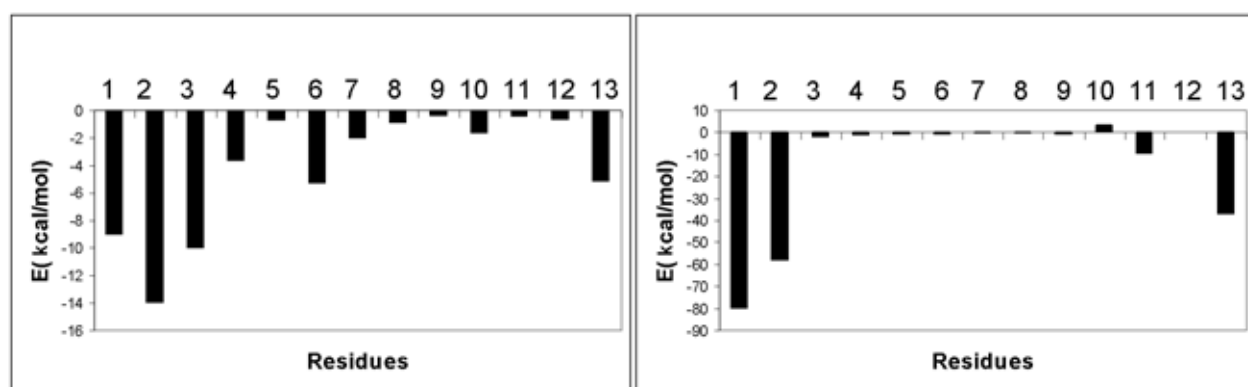


Figure 8: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Peptide 2-G6PDH complex.

Table 5 lists the hydrogen bonds found in the peptide 3-G6PDH complex (peptide residues numbered from 1 to 14). Hydrogen bonds between residue 10 and Glu (atom OE1) were not detected in the dimer

system, although residue 10 already showed high van der Waals and electrostatic energies in the protein dimer system.

Figure 9 shows the energy pattern in the complex. Concerning van der Waals interactions, they are distributed along the entire peptide while electrostatic ones are almost focused on residue 10, with an average electrostatic energy of -35 kcal/mol.

<i>Peptide 3</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 9 OG1	ND2-HD21	45.6	2.11	0.21	148.3	12.5
Residue 8 O	ND2-HD21	80.8	2.86	0.16	158.5	10.2
Residue 10 O	OH-HH	88.8	1.77	0.13	160.9	10.1
Residue 8 N-H	OD1	74.0	1.96	0.16	156.4	11.7
Residue 9 OG1-HG1	OH	79.2	1.93	0.17	157.5	12.0
Residue 10 NZ-HZ1	OE1	39.4	2.62	0.82	102.8	53.3
Residue 10 NZ-HZ2	OE1	31.2	2.72	0.71	95.4	51.5
Residue 10 N-H	OD1	81.2	1.94	0.14	157.2	10.1
Residue 13 N-H	OE2	51.8	2.44	0.68	156.3	12.5

Table 5: Average data of the hydrogen bonds found in the peptide 3-G6PDH complex.

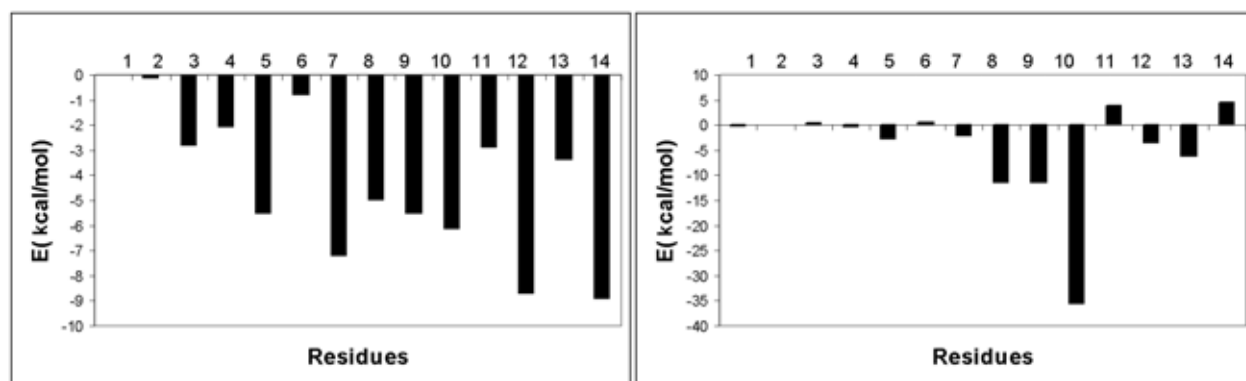


Figure 9: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Peptide 3-G6PDH complex.

Similarly, hydrogen bonds found in the peptide 4-G6PDH complex are shown in table 6 (peptide residues numbered from 1 to 7). Only one of the hydrogen bonds extracted in this zone for the G6PDH dimer was stable, it is formed between the peptide residue 4 and Glu of the monomer, with an occupation of 53.80 %. The other hydrogen bonds are also summarized in table 6, although they exhibited less stability. In addition, two hydrogen bonds were lost, maintained in the G6PDH dimer with the amino group of residue 1 and the carbonyl group of residue 4. Fluctuations of the peptide 4 structure were responsible of the lost of the described hydrogen bonds, this is the only peptide with no secondary structure even in the initial complex.

<i>Peptide 4</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 1 NZ-HZ1	OE1	33.0	2.62	0.66	101.6	48.0
Residue 1 NZ-HZ3	OE1	24.6	2.73	0.64	94.1	46.5
Residue 4 N-H	OE1	53.8	2.01	0.22	150.2	14.0
Residue 1 O	OH-HH	35.2	2.75	1.09	118.2	47.4

Table 6: Average data of the hydrogen bonds found in the peptide 4-G6PDH complex.

Figure 10 shows the average van der Waals and electrostatic energies of the system, the profile is similar as found in the G6PDH dimer, being residue 1 the most important one with respect to electrostatic interactions, and residues 1, 3, 5 the responsible residues for van der Waals recognition. Residue 7, which is an important hydrophobic residue in the dimer system, is not involved here in a high van der Waals energy, due to it is displaced from its hydrophobic pocket.

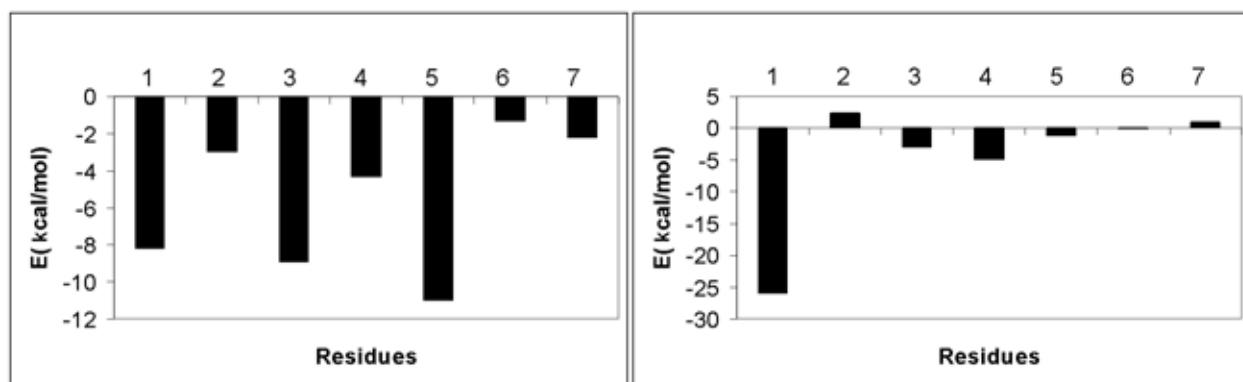


Figure 10: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Peptide 4-G6PDH complex.

Table 7 lists the hydrogen bonds found in the cyclic peptide-G6PDH complex (peptide residues numbered from 1 to 9). Only two hydrogen bonds were maintained along the simulation, 3 hydrogen bonds were lost considering them into the protein dimer system. This was also observed for the peptide 4 in complex with G6PDH. They were formed by the carbonyl group of residue 5 and residue 2 and the amino group of residue 2. This fact could be explained taking into account that the cyclic peptide is a more restricted molecule than the original beta turn fragment, together with the important directionality requirements for hydrogen bond formation. Nevertheless it is not probable that a free linear peptide could achieve this closed conformation to interact with the monomer (peptide 4), therefore the designing of this cyclic peptide is required to cover the beta turn interaction surface. In addition a cyclic peptide could be entropically favourable with respect to a linear one.

Figure 11 shows a good pattern of interactions, especially van der Waals, that are maintained by the contact of 3 residues. Electrostatic profile is almost driven by residue 2.

<i>Cyclic peptide</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 2 NZ-HZ1	OE2	55.6	2.19	0.66	130.4	46.5
Residue 5 N-H	OE2	87.9	1.95	0.18	162.5	9.2

Table 7: Average data of the hydrogen bonds found in the cyclic peptide-G6PDH complex.

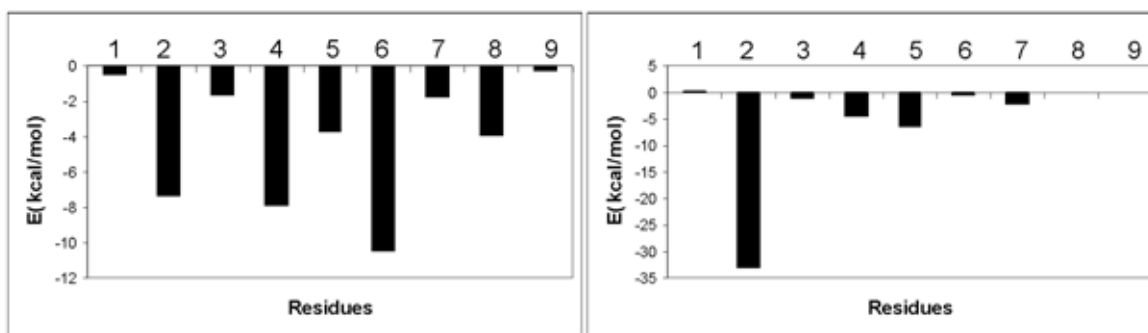


Figure 11: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Cyclic peptide -G6PDH complex.

Table 8 lists the hydrogen bonds found in the peptide 5-G6PDH complex (peptide residues numbered from 1 to 12). The hydrogen bond identified between residue 5 and Arg is new with respect to the analysis of the G6PDH dimer. In addition, Figure 12 shows the van der Waals and electrostatic energies of each peptide residue in the complex. As can be seen, residue 8 is the most important one for hydrophobic recognition in this peptide, while residues 5 and 7 are the most important ones for electrostatic recognition. Residue 7 was also observed in the protein G6PDH complex with high electrostatic energy (-45 kcal/mol).

<i>Peptide 5</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 9 O	N-H	56.6	2.26	0.28	157.4	10.4
Residue 7 OE1	NH2-HH21	41.4	2.16	0.37	147.1	13.4
Residue 5 OE1	NH2-HH22	67.8	2.16	0.71	155.4	11.8
Residue 7 OE2	NE-HE	43.4	2.50	0.82	143.3	17.6
Residue 9 N-H	O	77.0	1.98	0.18	158.3	12.5

Table 8: Average data of the hydrogen bonds found in the peptide 5-G6PDH complex.

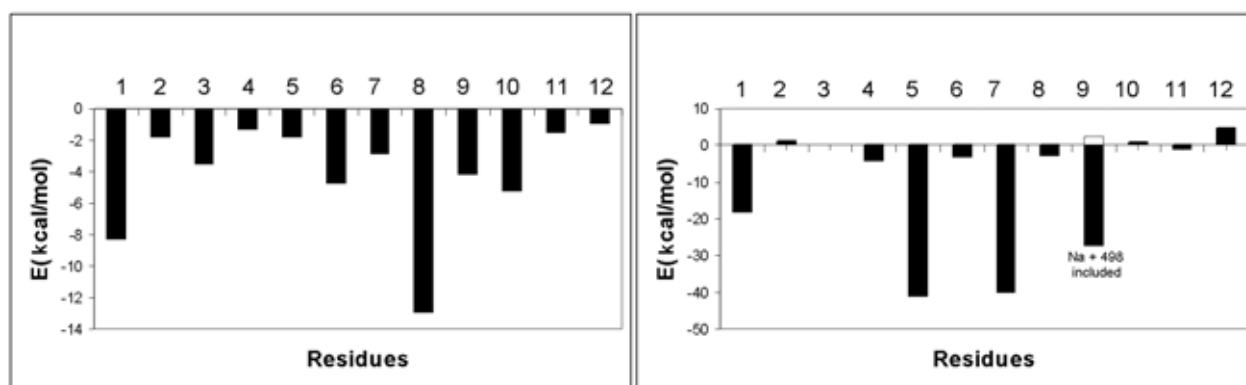


Figure 12: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Peptide 5-G6PDH complex.

Finally, Table 9 lists the hydrogen bond pattern in the last system, the peptide 6-G6PDH complex. Peptide residues are numbered from 1 to 10. All 3 hydrogen bonds were already observed in the G6PDH complex, nevertheless the hydrogen bond between residue 7 with Glu was replaced by the hydrogen bond between residue 7 and Glu in the G6PDH complex.

In addition, Figure 13 shows the energy profile for the peptide 6. Four important van der Waals contacts are described while electrostatic energy is almost directed by residue 6. In fact, this residue is the most interacting one with respect to the electrostatic profile in the dimer interface. It interacts with an electrostatic energy of -60 kcal/mol, both in the dimer and in the peptide-G6PDH complex.

<i>Peptide 6</i>	<i>G6PDH</i>	<i>% Occupation</i>	<i>Average distance/Å</i>	<i>Distance RMS/Å</i>	<i>Average angle/°</i>	<i>Angle RMS/°</i>
Residue 6 NH2-HH21	OE1	89.6	1.80	0.13	161.6	9.5
Residue 6 NE- HE	OE1	94.0	1.91	0.15	163.3	8.3
Residue 7 OH-HH	O	58.0	1.93	0.225	151.4	13.2

Table 9: Average data of the hydrogen bonds found in the peptide 6-G6PDH complex.

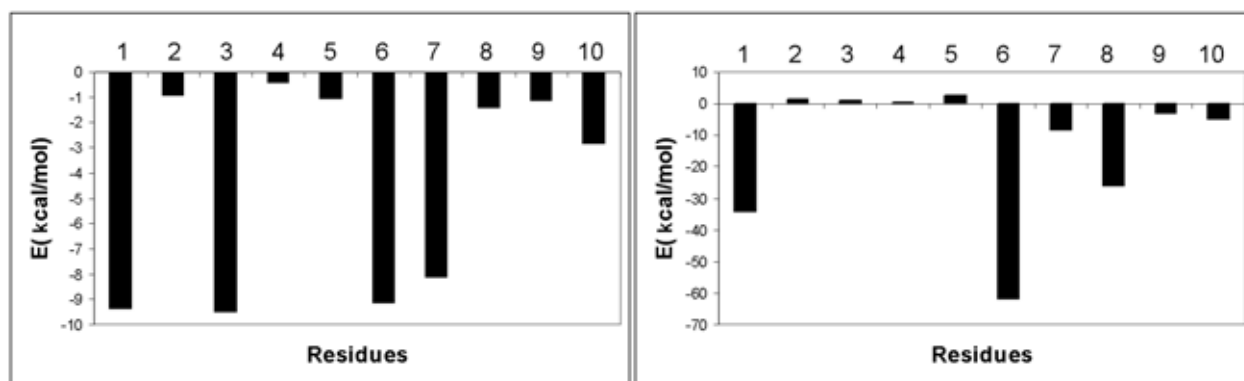


Figure 13: Average van der Waals (left) and electrostatic (right) energies of each peptide residue in the Peptide 6-G6PDH complex.

4.7. CALCULATION OF BINDING FREE ENERGY FOR THE PEPTIDE-G6PDH COMPLEXES

To complement the interaction analysis performed for the seven peptide-G6PDH complexes, it is described here the results of the binding free energy analysis applying the MMPB(GB)SA protocol [13]. For this purpose, 100 snapshots from the last 500 ps were extracted from the production time of the simulation of the cyclic peptide-G6PDH complex and 100 snapshots from the last 1000 ps for the other systems, explicit water molecules were deleted. The snapshots were used to evaluate the binding energies of the complex, the peptide and the monomer G6PDH molecule, therefore it was used the one-trajectory protocol, that reduces the computational cost of the method.

Table 10 lists the average contributions of the MMPB(GB)SA simulation, shown in kcal/mol units. ELE accounts for the electrostatic energy balance between the peptide and the protein, VDW accounts for the van der Waals interactions of both fragments, GAS accounts for the addition of the last two terms, being the binding enthalpic contribution in vacuo. For the resolution of electrostatic solvation energies both Poisson Boltzmann (PB) and Generalized Born (GB) methods were applied. Tsui et al. [20] parameters were used within the GB framework. Parse radii [19] were selected to solve the PB equation. PB_{SUR} is the non-polar contribution to solvation while PB_{CAL} is the polar contribution. PB_{SOL} is the addition of both terms, being related to total solvation energy.

PB_{ELE} is the $PB_{CAL} + ELE$ addition, related to the balance of electrostatic interactions of the fragments and the desolvation of them. Finally PB_{TOT} is the free energy of binding without entropic factors at 300 K and using the PB framework. Similarly, GB_{SUR} is the non-polar contribution to solvation using the generalized Born approach, GB_{CAL} denotes for the polar contribution. GB_{ELE} is the $GB_{CAL} + ELE$ addition, and finally GB_{TOT} is the free energy of binding without entropic factors at 300 K, using the GB approach.

In addition, to obtain absolute binding free energies, entropic factors were also calculated. Normal mode computation was carried out for this purpose using the NMODE module of AMBER7. To reduce computational cost, the receptor, the G6PDH monomer, was cut to only those residues located within 12 Å from the peptide (NADP molecule was included when this molecule belonged to the predefined cutoff). 10 snapshots were extracted from the simulation of each system and a minimization step until and energy gradient of 10^{-4} kcal/mol was performed for each one before the normal mode calculation.

	<i>Peptide 1- G6PDH</i>	<i>Peptide 2- G6PDH</i>	<i>Peptide 3- G6PDH</i>	<i>Peptide 4- G6PDH</i>	<i>Cyclic Peptide- G6PDH</i>	<i>Peptide 5- G6PDH</i>	<i>Peptide 6- G6PDH</i>
ELE	-174.77	-539.84	-164.74	-144.98	-183.38	-20.08	-406.52
VDW	-60.62	-56.79	-64.98	-40.53	-40.59	-51.49	-43.79
GAS	-235.39	-596.61	-229.72	-185.51	-223.98	-71.57	-450.31
PB _{SUR}	-7.26	-7.47	-8.50	-5.14	-5.39	-6.64	-6.49
PB _{CAL}	235.43	591.34	212.74	175.94	215.88	84.14	456.23
PB _{SOL}	228.17	583.87	204.24	170.80	210.49	77.50	449.74
PB _{ELE}	60.66	51.50	47.99	30.96	32.50	64.06	49.71
PB _{TOT}	-7.22	-12.74	-25.48	-14.71	-13.49	5.93	-0.57
GB _{SUR}	-8.42	-8.70	-10.07	-5.61	-5.94	-7.60	-7.40
GB _{CAL}	186.93	528.28	173.27	153.76	182.28	36.68	406.90
GB _{SOL}	178.51	519.58	163.21	148.16	176.34	29.08	399.50
GB _{ELE}	12.16	-11.55	8.53	8.78	-1.10	16.60	0.38
GB _{TOT}	-56.88	-77.03	-66.52	-37.35	-47.64	-42.49	-50.82
-TΔS _{TRA}	14.34	14.21	14.27	13.72	13.81	14.14	13.96
-TΔS _{ROT}	13.17	12.81	13.47	11.92	12.13	13.29	12.39
-TΔS _{VIB}	-0.31	1.33	15.50	-4.23	-1.16	12.39	1.33
-TΔS _{TOT}	27.20	28.35	43.24	21.41	24.78	39.82	27.68
ΔG _{MMPBSA}	19.98	15.61	17.76	6.70	11.29	45.75	27.11
ΔG _{MMGBSA}	-29.68	-48.68	-23.28	-15.92	-22.86	-2.67	-23.14

Table 10: MMPB(GB)SA data of the seven peptide-G6PDH complexes studied (kcal/mol units).

In Table 10, ΔS_{TRAS} , ΔS_{ROT} and ΔS_{VIB} are denoted for the translational, rotational and vibrational contributions to the entropy, while ΔS_{TOT} is the total entropic contribution. Finally, ΔG_{MMPBSA} is the absolute binding free energy using PB approach and ΔG_{MMGBSA} is the final one using the GB approach. As we can see, using PB and adding entropic factors, all binding free energies are positive. This general performance was previously observed [26, 27] and it was suggested to be a consequence of parse radii. This set of radii is related with high desolvation penalties (PB_{CAL}) and low binding enthalpic results (PB_{TOT}). However, PB_{TOT} results should show good correlation with experimental data, based on our experience with other peptide-protein systems [27]. In this sense, the most active peptides should be the peptide 3, followed by the peptide 4, the cyclic peptide and the peptide 2. Regarding only results using GB_{TOT} , peptide 2 and peptide 3 should be the most active ones. Peptide 1, 2 and 3 also showed the highest van der Waals interaction energies.

The addition of entropy does not change the general relative results, due to this contribution is almost constant among all peptides (about 21-28 kcal/mol), even in the most constrained cyclic peptide-G6PDH system. However, peptides 3 and 5 show an increased entropic penalty of 43.24 and 39.82 kcal/mol respectively, which reduces the final binding free energy of these peptides. Peptide 5 is considered as the worst one using both PB and GB methods.

It is worth to remark that the use of 1-trajectory protocol, although habitual, is based on the assumption that small changes in ligand and receptor conformations are suffered during binding. Concerning the linear peptides, one can assume that this is not completely correct, thus binding free energies of peptides 1 to 6, should be overestimated. Nevertheless, a complete study of the linear peptide conformations in solution, to improve the entropic results reported, is out of the aim of this work.

5. EXPERIMENTAL RESULTS

The experimental tests of G6PDH inhibition were performed by Gema Alcarraz Vizán, at the 'Integrative Biochemistry and Cancer Therapy (UB)' research group, coordinated by Prof. Cascante.

The enzymatic reaction was followed by adding the substrate, glucose-6-phosphate, to the enzyme and measuring the absorbance increase by NADPH formation, at $\lambda=340$ nm. Only the cyclic peptide was tested at the end of this thesis, and preliminary inhibition was observed using G6PDH from yeast. The

IC₅₀ found was estimated in 250 μ M. Although this is too high, the cyclic peptide was designed to inhibit the human variant of G6PDH. Additional tests were performed adding the peptide to human cells, but no positive results were obtained, presumably the compound is not able to penetrate the cell bilayer. Peptides 2 and 3 were also synthesized but the inhibition assay concerning these compounds is considered as future work.

6. FURTHER MODIFICATION OF THE CYCLIC PEPTIDE

The general aim of this section is to design new analogs of the cyclic peptide with improved properties. As shown before, the cyclic peptide was active in G6PDH inhibition assays but the permeability to the cell membrane was very low. We suggest that the poor permeability of the compound is based on the charged residues, residues 2 and 3. Thus, the strategy to design new cyclic peptide analogs has been focused on the replacement of charged residues and on the addition of hydrophobic residues. Moreover, new interactions with the receptor have been identified. Table 11 shows a summary of the most important changes that can be performed on the cyclic peptide sequence (residues numbered from 1 to 9) accordingly with the interaction analysis shown in section 4.6.

<i>Cyclic peptide residue</i>	<i>Modification possibilities</i>
Residue 1	Cannot be replaced, flexible residue added to close the peptide, it is far from receptor residues.
Residue 2	Cannot be replaced, it interacts with the receptor by electrostatic forces, and establishing a hydrogen bond.
Residue 3	A hydrophobic residue can be added in this position to improve intra and intermolecular interactions.
Residue 4	Cannot be replaced, the backbone is essential to maintain the cyclic structure.
Residue 5	It can be replaced by a hydrophobic residue to improve the membrane permeability and to establish van der Waals interactions with the receptor.
Residue 6	Cannot be replaced, this residue is implicated in van der Waals recognition of the peptide.
Residue 7	Cannot be replaced, it is far from receptor residues and it seems not responsible for the poor permeability.
Residue 8	Cannot be replaced, this residue is implicated in van der Waals recognition of the peptide.
Residue 9	Cannot be replaced, flexible residue added to close the peptide, it is far from receptor residues.

Table 11: Most important guidelines to modify the cyclic peptide sequence.

Therefore, as a first attempt to improve the recognition of the cyclic peptide and to increase its permeability, five new analogs were modeled. Res3F derivative was designed by replacing the third position with a hydrophobic residue (phenylalanine). Res5A, Res5L and Res5F derivatives were constructed to replace the fifth position with different hydrophobic residues, ranging from a small alanine to a large phenylalanine. Finally, a double replacement was performed in the Res3F_Res5L analog, trying to improve two substitutable positions.

The new cyclic peptide-G6PDH complexes, were constructed from the minimized coordinates of the original cyclic peptide-G6PDH complex and performing a lateral chain substitution. An accurate minimization protocol was carried out to optimize the atoms of the new residues. As explained in the methods section, a molecular dynamics of 2-2.5 ns was also carried out for each system and finally the MMPB(GB)SA methodology was also applied to evaluate the binding free energies of the five new cyclic peptide derivatives in complex with G6PDH (Table 12). Similarly, 100 snapshots were extracted from the production time (final 500 ps) to compute the enthalpic and internal contributions, and 10 snapshots were extracted to compute the entropic contribution. For the latter calculation, the receptor was cut to those residues at a maximum distance of 12 Å from the peptides.

All the five derivatives show an increased van der Waals contribution with respect to the original cyclic peptide complex, revealing that new interactions with the receptor have been established. Most important change is shown for the double replacement Res3F_Res5L, thus the van der Waals energy changes from -40.6 kcal/mol to -57.8 kcal/mol. On the other hand, the electrostatic balance is decreased in all cases because of the substitution of the charged residue 3 by phenylalanine, and the polar residue 5 by alanine, leucine or phenylalanine. Results using Poisson-Boltzmann equation (PB_{TOT}) identify the Res3F, Res5L and Res3F_Res5L as the most affinity derivatives, ranging their binding free energies between -18.9 and -20.20 kcal/mol. Taken into account implicit limitations of the MMPBSA method and limitations of the molecular dynamics simulation, one cannot discriminate among these three cyclic peptides. With respect to the results using Generalized Born equation (GB_{TOT}) the same three derivatives can be considered as the most promising compounds, ranging their binding free energies between -49.5 and -51.1 kcal/mol. Best derivatives, when adding also the entropic contribution, are the Res3F and the Res3F_Res5L. With respect to the Res5A, an unfavourable entropic contribution of 26.2 kcal/mol decreases the final binding free energy. Summarizing these results, one can identify the

peptide with the double replacement (Res3F_Res5L) as the most promising one, thus, it shows favourable VDW, PB_{TOT} and GB_{TOT} results and in addition, it is more hydrophobic than the Res3F or Res5L analogs that also present similar energies.

	<i>Cyclic Peptide-G6PDH</i>	<i>Cyclic Peptide Res3F-G6PDH</i>	<i>Cyclic Peptide Res5A-G6PDH</i>	<i>Cyclic Peptide Res5L-G6PDH</i>	<i>Cyclic Peptide Res5F-G6PDH</i>	<i>Cyclic Peptide Res3F_Res5L-G6PDH</i>
ELE	-183.38	-128.51	-172.10	-144.44	-181.16	-101.90
VDW	-40.59	-48.45	-39.47	-53.03	-41.92	-57.81
GAS	-223.98	-176.96	-211.57	-197.47	-226.08	-159.70
PB_{SUR}	-5.39	-6.14	-5.10	-6.21	-5.31	-6.52
PB_{CAL}	215.88	164.18	202.35	183.48	217.40	146.02
PB_{SOL}	210.49	158.04	197.25	177.27	212.10	139.50
PB_{ELE}	32.50	35.67	30.25	39.04	33.24	44.12
PB_{TOT}	-13.49	-18.92	-14.32	-20.20	-13.99	-20.21
GB_{SUR}	-5.94	-6.94	-5.55	-7.03	-5.83	-7.44
GB_{CAL}	182.28	133.23	175.23	155.03	188.47	116.10
GB_{SOL}	176.34	126.29	169.67	148.01	182.65	108.66
GB_{ELE}	-1.10	4.72	3.13	10.60	4.31	14.20
GB_{TOT}	-47.64	-50.67	-41.90	-49.46	-43.44	-51.05
$-\Delta S_{TRA}$	13.81	13.83	13.83	13.86	13.89	13.88
$-\Delta S_{ROT}$	12.13	12.07	12.10	12.20	12.16	12.21
$-\Delta S_{VIB}$	-1.16	-4.14	-0.13	0.09	-2.51	-0.90
$-\Delta S_{TOT}$	24.78	21.76	25.79	26.15	23.54	25.18
ΔG_{MMPBSA}	11.29	2.84	11.47	5.95	9.55	4.97
ΔG_{MMGBSA}	-22.86	-28.91	-16.11	-23.31	-19.90	-25.87

Table 12: MMPB(GB)SA data of the five new cyclic peptide-G6PDH complexes studied (kcal/mol units). The original cyclic peptide-G6PDH complex is shown in first column for reference.

A detailed view of the new interactions in the Res3F_Res5L cyclic peptide is shown in Figure 14. Most important contacts are established with hydrophobic residues of the receptor. Thus, the phenylalanine added in the third position maintains a ring-ring interaction with Y and a van der Waals contact with V. The replaced residue of the fifth position, leucine, interacts also with V.

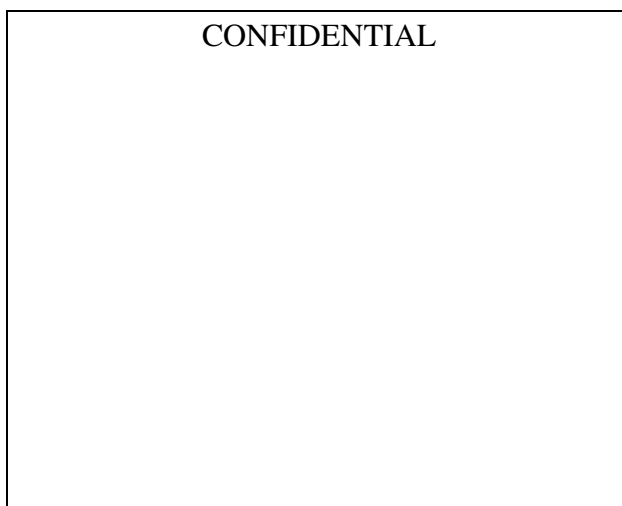


Figure 14: *New interactions appearing in the cyclic peptide Res3F_Res5L-G6PDH complex. The peptide is shown in red while the receptor is shown in green.*

The synthesis and the experimental assays for the cyclic peptide presenting the double replacement have been considered as a future work, nevertheless results indicate a better binding and a presumable improved cell permeability for this compound.

7. CONCLUSIONS

We have presented in this work, a theoretical study of the recognition of protein-protein interactions, to disrupt the dimer complex of the human G6PDH, an important enzyme that catalyses the rate-limiting step of the oxidative branch of the pentose phosphate pathway and a new target for cancer research.

Molecular dynamics in conjunction with free energy calculations have been performed. Firstly, it was studied the molecular recognition between the G6PDH monomers, classifying the interactions in hydrogen bonds, electrostatic and van der Waals contacts. In corroboration with the experimental

structure, the interface surface is composed mainly by the alpha and the beta domains, covering a total of 64 residues. It was identified a hydrophobic core on the beta domain of the surface that maintains 8 stable protein-protein contacts within a 7-residue sequence, being an important zone for the protein dimerization. Accordingly, it has been modeled six peptide-G6PDH complexes and a special cyclic peptide-G6PDH complex, based on the natural sequence of the monomer interface to find an effective dimerization inhibitor. Molecular dynamics and finally, binding free energy of each peptide-G6PDH system were analyzed, to predict the affinity of the proposed ligands against G6PDH inhibition. Taking together these analyses, peptides 2, 3 and the cyclic peptide were synthesized. The cyclic peptide was tested against G6PDH from yeast showing inhibition but it is unable to penetrate the cell membrane. In addition, 5 new cyclic peptide derivatives were designed attempting to improve the affinity and the permeability.

Concluding, this is a general work to obtain short peptide inhibitors for a crucial protein related to cancer disease, using entirely a structure-based drug design approach and focusing it on the disruption of protein-protein interactions, by using short peptides derived from the natural sequence of the protein interface. The active peptides found can be also used to derive non-peptidic molecules, with more chemical stability than the original peptides and retaining the most important contacts. This strategy will be presented in the next chapter, where the designed cyclic peptide is used as a reference to find new non-peptidic inhibitors of G6PDH.

8. REFERENCES

1. Paul, E.; Carson, M. D.; Henri Frischer, M. D. Glucose-6-phosphate dehydrogenase deficiency and related disorders of the pentose phosphate pathway. *Am. J. Med.* **1966**, *41*, 744-761.
2. Boren, J.; Ramos-Montoya, A.; De Atauri, P.; Comin-Anduix, B.; Cortes, A.; Centelles, J. J.; Frederiks, W. M.; Van Noorden, C. J. F; Cascante M. Metabolic Control Analysis Aimed at the Ribose Synthesis Pathways of Tumor Cells: A New Strategy for Antitumor Drug Development. *Mol. Biol. Rep.* **2002**, *29*, 7-12.
3. Babiak, R. M.; Campello, A.P; Carnieri, E. G.; Oliveira, M. B. Methotrexate: pentose cycle and oxidative stress. *Cell Biochem. Funct.* **1998** , *4*, 283-93.
4. Asensio, C.; Levoine, N.; Guillaume, C.; Guerquin, M. J.; Rouguieg, K.; Chretien, F.; Chapleur, Y.; Netter, P.; Minn, A.; Lapique, F. Irreversible inhibition of glucose-6-phosphate dehydrogenase by the coenzyme A conjugate of ketoprofen: a key to oxidative stress induced by non-steroidal anti-inflammatory drugs?. *Biochem. Pharmacol.* **2007**, *73*, 405-416.
5. Ciftci, M.; Buyukokuroglu, M. E.; Kufrevioglu, O. I. Effect of cefaperazone/sulbactam and ampicillin/sulbactam on the in vitro activity of human erythrocyte glucose-6-phosphate dehydrogenase. *J. Basic Clin. Physiol. Pharm.* **2001**, *12*, 305-313.
6. Ciftci, M.; Ozmen, I.; Buyukokuroglu, M. E.; Pence, S.; Kufrevioglu, O. I. Effects of metamizol and magnesium sulfate on enzyme activity of glucose 6-phosphate dehydrogenase from human erythrocyte in vitro and rat erythrocyte in vivo. *Clin. Biochem.* **2001**, *34*, 297-302.
7. Camarasa, M. J.; Velazquez, S.; San-Felix, A.; Perez-Perez. M. J.; Gago, F. Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: a single mode of inhibition for the three HIV enzymes?. *Antiviral Res.* **2006**, *71*, 260-267.
8. Reed, J. C. Apoptosis-base therapies, *Nat. Rev. Drug Discov.* **2002**, *1*, 111-121.
9. Igney, F. H.; Krammer, P. H. Death and anti-death: tumour resistance to apoptosis. *Nat. Rev. Cancer* **2002**, *2*, 277-288.
10. Shimba, N.; Nomura, A.M.; Marnett, A.B.; Craik, C.S. Herpesvirus Protease inhibition by dimer disruption. *J. Virol.* **2004**, *78*, 6657-6665.
11. Wells, J. A.; McClendon C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature.* 2007, *450*, 1001-1009
12. Zutshi, R.; Brickner, M.; Chmielewski, J. Inhibiting the assembly of protein-protein interfaces. *Curr. Opin. Chem. Biol.* **1998**, *2*, 62-66.

13. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Srivasan, J., Case, D. A.; Cheatham III, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.
14. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Goud, I. R.; Mertz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C., Fox, T.; Caldwell, J. W.; Kollman, P.A. A second generation of force fields for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
15. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. D.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Sligh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, Univ. California, San Francisco, 2002.
16. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald : an $W \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
17. Weiser, J.; Shemkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217-230.
18. Bashford, D.; Gerwent, K. Electrostatic calculations of the pK a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **1992**, *224*, 473-486.
19. Sigfridsson, E.; Ryde, U. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.* **1997**, *19*, 377-395.
20. Tsui, V.; Case, D. A. Theory and application of the generalized born solvation model in macromolecular simulations. *Nucl. Acid. Sci.* **2001**, *56*, 275-291.
21. Humphrey, W.; Dalke, A.; Schulten, K. VMD-Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33-38.
22. Au, S. W.; Gover, S.; Lam, V. M.; Adams, M. J. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Structure* **2000**, *8*, 293-303.
23. Holmberg, N.; Ryde, U.; Bulow, L. Redesign of the coenzyme specificity in L-lactate dehydrogenase from *Bacillus stearothermophilus* using site-directed mutagenesis and media engineering. *Prot. Engin.* **1999**, *12*, 851-856.
24. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

25. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *J. Comput. Chem.* **1977**, *23*, 327-341.
26. Pearlman, D. A. Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a cogenerated series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796-7806.
27. Obiol-Pardo, C.; Rubio-Martinez, J. Comparative evaluation of MMPBSA and XSCORE to compute binding free energy in XIAP-peptide complexes. *J. Chem. Inf. Model.* **2007**, *47*, 134-142.

CHAPTER VII: Pharmacophore Exploitation of Glucose-6-Phosphate
Dehydrogenase: Searching for non peptidic Inhibitors



1. BRIEF INTRODUCTION

This chapter is based on the pharmacophore exploitation of the human Glucose-6-Phosphate Dehydrogenase (G6PDH), to search for non peptidic inhibitors.

As it was commented in the previous chapter, Glucose-6-Phosphate Dehydrogenase is the first and rate-limiting enzyme of the oxidative branch of the pentose phosphate pathway and its inhibition decreases the cell division, specifically in tumor cells, by reducing the amount of ribose molecules available for the synthesis of nucleic acids.

Up to now, no drugs targeting G6PDH have reached the pharmaceutical market.

Here, it is described the methods and results obtained searching small non-peptidic compounds inhibitors of G6PDH, by means of a rational drug design protocol. This protocol includes the generation of pharmacophores, 3D database searching and use of molecular modeling tools, such as docking and scoring, in order to obtain a human G6PDH inhibitor. The best 8 candidate molecules (from a total screen of 4,298) were acquired from the commercial sources and experimental tests of G6PDH inhibition were carried out, this information is also included in this chapter.

2. CONTEXT

G6PDH is a NADP⁺ dependent enzyme that acts at the first and rate-limiting step of the oxidative branch of the pentose phosphate pathway. This metabolic pathway provides ribose molecules which are one of the basic components for nucleic acid production.

As the tumor cell needs for an extra production of nucleic acids, the target of G6PDH could be a novel strategy to discover new drugs against the cancer disease [1].

Chapter 6 was focused on the study of the dimer disruption of the human G6PDH. Two domains were identified as responsible of the most important contacts and specifically, a sequence of the beta domain, which maintains 11 stable protein-protein contacts. In addition, this sequence was used to construct the peptide 4 and to design the cyclic peptide, which was active inhibiting the protein of yeast.

Here, based on a pharmacophore description of the interactions found in this sequence, 3D database search was carried out to find new non peptidic inhibitors of G6PDH with the most important contacts

maintained in the protein dimer. The CATALYST [2] program was used for this purpose. Then, all candidate molecules, were docked with the protein monomer, acting as a rigid receptor, and using our home-made program of docking [3]. After neglecting the molecules with steric clash, accepted ones were scored using also the XSCORE semiempirical function [4].

Finally, the best 8 compounds were purchased from the commercial companies and tests of G6PDH inhibition were performed by an experimental group.

Using this protocol, it has been found 4 active molecules (50 %), with G6PDH inhibitory activity. The experimental results are also reported in this chapter.

3. RESULTS

3.1. PHARMACOPHORE AND 3D-SEARCHING

Figure 1 shows the selected pharmacophore of the sequence found, that maintains the most important interface contacts. It presents 11 interaction points, classified into hydrogen bonds (points 1, 2, 3, 6, 8, and 9), electrostatic (point 5) and van der Waals (points 4, 7, 10 and 11). In addition, Table 1 summarizes the geometric parameters which identify this pharmacophore, as well as their dynamical behaviour throughout the Molecular Dynamics simulation. This information was used to search in the 3D database of CATALYST [2], with the best flexible search option.

Taking into account that this pharmacophore is composed by a high number of interactions, distributed along a turn structure, it was selected the subset of points 7, 8, 9 and 10 to search in the database.

This subset is composed by an acceptor hydrogen bond (point 9), a donor hydrogen bond (point 8), and two of the most important hydrophobic contacts (points 7 and 10). More important, these points are distributed along a linear structure, which should be easy to mimic by a small molecule.

Search concluded with 4,298 molecules that were saved from the program to perform a second evaluation, by means of docking and scoring protocols.

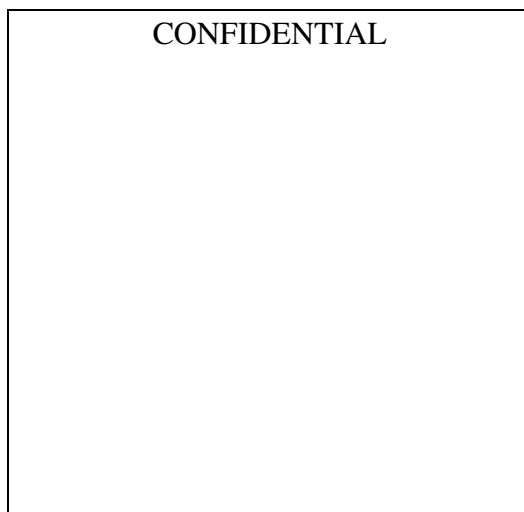


Figure 1: G6PDH pharmacophore, selected in this work. Points 4, 7, 10 and 11 account for van der Waals contacts, points 2, 3 and 8 denote for hydrogen bond donors, point 5 denotes for an electrostatic contact and points 1, 6 and 9 denote for hydrogen bond acceptors.

<i>Pharmacophoric Points</i>	<i>Average Distance/ Å</i>	<i>Maximum Distance/ Å</i>	<i>Minimum Distance/ Å</i>
CONFIDENTIAL			

Table 1 (I): Geometrical data of the G6PDH pharmacophore.

<i>Pharmacophoric Points</i>	<i>Average Distance/ Å</i>	<i>Maximum Distance/ Å</i>	<i>Minimum Distance/ Å</i>
CONFIDENTIAL			

Table 1 (II): Geometrical data of the G6PDH pharmacophore.

<i>Pharmacophoric Points</i>	<i>Average Distance/ Å</i>	<i>Maximum Distance/ Å</i>	<i>Minimum Distance/ Å</i>
CONFIDENTIAL			

Table 1 (III): Geometrical data of the G6PDH pharmacophore.

3.2. DOCKING AND SCORING PROTOCOLS. PURCHASED COMPOUNDS

The docking methodology was carried out by using our home-made program, Dock_Dyn [3], whose basic issues were explained at the first chapter.

It has been evaluated 100 conformations for each molecule, to an amount of 429,800 poses approximately into the surface of the G6PDH monomer.

To study only the best molecules and their conformations, the maximum RMS deviation of the 4-point

pharmacophore was set to 1.0 Å. Nevertheless, to study enough molecules and to take into account the receptor flexibility, the van der Waals radii were reduced to 80%.

Favorable docked poses were then scored by means of the XSCORE [4] semiempirical function.

As a criterion of selection, molecules presenting a low RMS deviation from the reference pharmacophore together with a high XSCORE value were purchased. A maximum chemical diversity was also taken into account. The best 8 compounds (Figure 2) were acquired from the ChemDiv and SPECS chemical companies. Table 2 shows the RMS and XSCORE values of them extracted from the docking and scoring protocols.

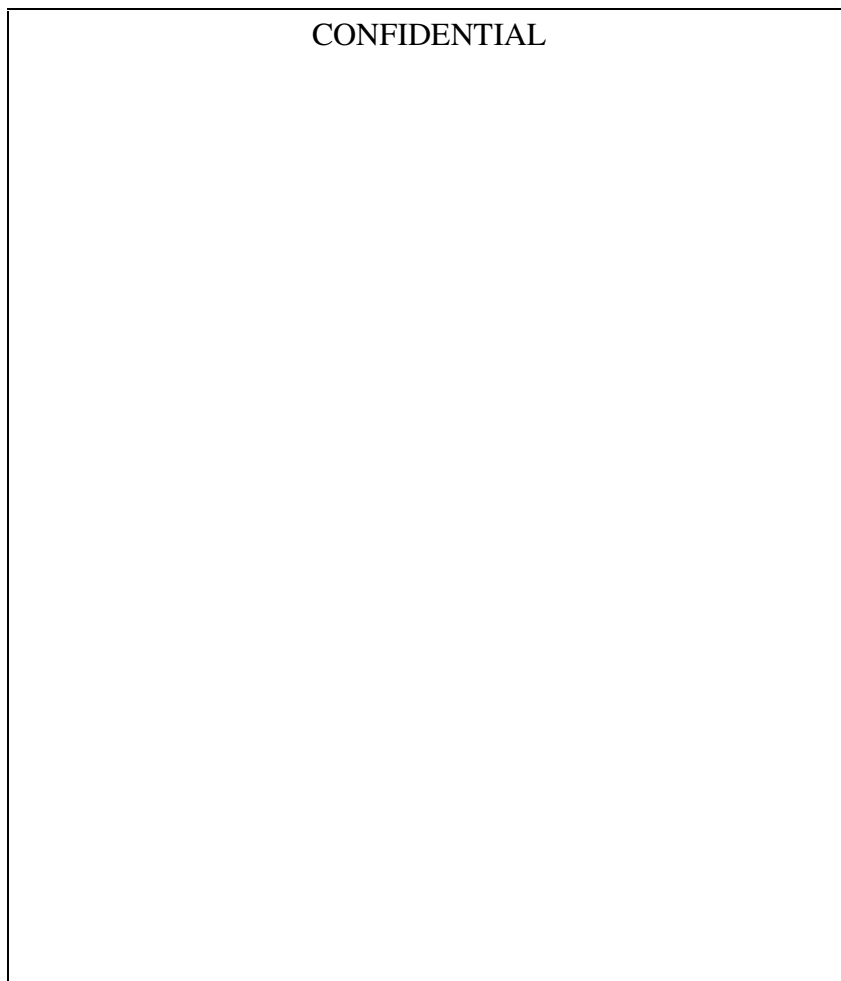


Figure 2: Purchased compounds to target G6PDH.

<i>Compound</i>	<i>Best RMS/Å</i>	<i>Best XSCORE</i>	<i>Predicted logP</i>
G1	0.16	5.49	4.78
G2	0.83	6.23	5.28
G3	0.22	5.45	4.88
G4	0.16	5.70	5.38
G5	0.19	5.54	3.06
G6	0.16	5.45	4.47
G7	0.46	6.17	5.65
G8	0.25	5.07	1.77

Table 2: Best RMS and XSCORE values from the docking and scoring protocols. It is also shown the predicted logP concerning the solubility of the compounds.

In addition, Figure 3 depicts the best docking poses for compounds G5 and G6 in complex with the G6PDH monomer. These molecules recognize the same residues that recognize the sequence found in the complete dimer. Thus, the compounds form two hydrogen bonds with polar residues, and the hydrophobic rings interact tightly with several non-polar residues.

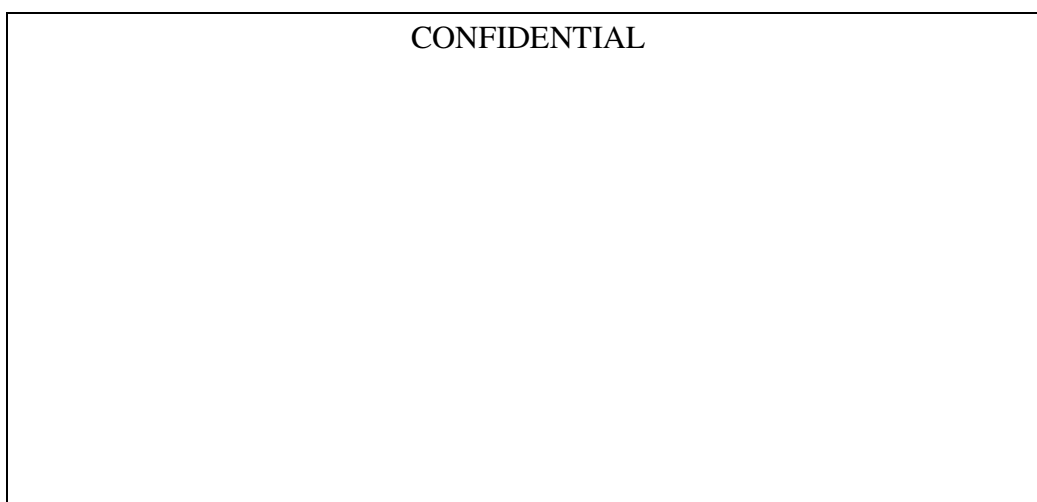


Figure 3: Best docking poses for compounds G5 (A) and G6 (B) in complex with the monomer of G6PDH. The protein is shown in transparent orange and carbon atoms of the ligand are marked in light green.

4. EXPERIMENTAL RESULTS

The experimental tests of G6PDH inhibition were performed by Gema Alcarraz Vizán, at the 'Integrative Biochemistry and Cancer Therapy (UB)' research group, coordinated by Prof. Cascante.

The enzymatic reaction was followed by adding the substrate, glucose-6-phosphate, to the human enzyme and measuring the absorbance increase by NADPH formation, at $\lambda=340$ nm.

Four of the eight compounds, G2, G4, G5 and G6, were identified as actives, presenting G6PDH inhibition in the micromolar range of concentration.

Table 3 summarizes the experimental results, expressed as per cent of inhibition, testing the active compounds at concentrations between 50 and 500 μM .

% G6PDH Inhibition	G2	G4	G5	G6
50 μM	28.7	32.0	38.0	5.2
100 μM	45.9	42.2	73.2	-
250 μM	88.7	95.0	100.0	53.3
500 μM	-	-	100.0	100.0

Table 3: Experimental activities of compounds G2, G4, G5 and G6.

The IC_{50} , or concentration required to reduce a 50 % of enzymatic activity, was estimated in 100 μM for compounds G2 and G4, 250 μM for compound G6 and 67 μM for compound G5 approximately. Unfortunately, compounds G1 and G3 were insoluble even in DMSO and therefore they were not tested. Concerning the solubility, Table 2 also shows the predicted logP values of these compounds using the XLOGP3 algorithm [4]. Results do not predict the low solubility of G1 and G3, moreover there are other molecules that present a higher logP, such as G2 and G4, and these molecules were solved.

In addition, it was performed an inhibitor assay at cell level, by using the HT29 colon adenocarcinoma cells; IC_{50} of compound G2 is > 500 μM having low cell bilayer permeability. Finally, IC_{50} of compound G4 is 10-20 μM and IC_{50} of compound G5 is 50 μM .

For reference, the well-known inhibitor of the protein, DHEA, presents an $\text{IC}_{50} < 50$ μM (in vitro) and

an IC_{50} of 20 μM in HT29 cells.

These active molecules can be considered as new lead compounds to target G6PDH, acting as dimerization inhibitors. It is worth to remark the success rate found by using molecular modeling together with pharmacophore-based drug design. Future work will be devoted to improve the potency of the active compounds found and especially for the molecules G4 and G5.

5. CONCLUSIONS

By using an original drug design protocol, which includes molecular dynamics, pharmacophore generation, database searching and docking-scoring methods, it has been found 4 active molecules (50% of the total tested) with G6PDH inhibitory activity.

They were designed to disrupt the native dimer of G6PDH covering 4 of the 11-point pharmacophore found in the beta domain of the G6PDH interface. Moreover, they were designed to mimic the most important contacts covered by the peptide 4 or the cyclic peptide commented in the previous chapter.

Although the activity of the compounds is only moderate, ranging their IC_{50} from 67 to 250 μM , they can be used as starting point to improve the activity of a second generation of G6PDH inhibitors. Interestingly, compound G4 showed an IC_{50} of 10-20 μM at cell level, this is a good result to search new G4 derivatives.

To our knowledge they are the first non-peptidic and non-steroid molecules targeting this protein by disrupting the dimerization structure.

Further experiments will be devoted to characterize their profile in other tumor cells and their activity in conjunction with the Transketolase inhibitors described in chapter 5. Hence, a combined treatment using a G6PDH inhibitor and a Transketolase inhibitor, that targets both the oxidative and non-oxidative branches of the pentose phosphate pathway, should be of great interest.

6. REFERENCES

1. Boren, J.; Ramos-Montoya, A.; De Atauri, P.; Comin-Anduix, B.; Cortes, A.; Centelles, J. J.; Frederiks, W. M.; Van Noorden, C. J. F; Cascante M. Metabolic Control Analysis Aimed at the Ribose Synthesis Pathways of Tumor Cells: A New Strategy for Antitumor Drug Development. *Mol. Biol. Rep.* **2002**, *29*, 7-12.
2. CATALYST™ (Accelrys Inc. USA).
3. Rubio-Martinez, J.; Pinto, M.; Tomas M.S.; Perez, J. J. Dock_Dyn: a program for fast molecular docking using molecular dynamics information. University of Barcelona and Technical University of Catalonia. Barcelona, **2005**.
4. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11-26.

CHAPTER VIII: Final Conclusions



The following general conclusions, covering some topics in Molecular Modeling and Drug Design, can be derived from this work:

◆ **With respect to Protein-Protein Recognition:**

Although protein-protein recognition is established by a high number of interaction points distributed along a complex and large surface, most important contacts can be identified in small clusters (also named hot spots) presenting the residues with the most hydrophobic interactions and a small number of charged residues involved in hydrogen bonds. Our results show that the protein-protein recognition along the dimerization interface, can be disrupted using short peptides derived from the natural sequence and using also small molecules designed with similar interaction points.

◆ **With respect to Protein-Protein Pharmacophores:**

It is demonstrated in this work that a pharmacophore derived from a natural protein-protein or peptide-protein complex is useful to search for small molecules acting as protein mimetics and competing for the same binding site. In addition, a pharmacophore derived from a molecular dynamics simulation can raise the successful rate found in virtual screening protocols, including the flexibility of contacts not taken into account in a pharmacophore derived directly from the crystal structure.

◆ **With respect to Binding Free Energy Prediction:**

The MMPB(GB)SA methodology has been applied in order to predict the affinity for 32 complexes, especially in peptide-protein complexes. Although more work has to be done, this protocol is not able to reproduce the experimental affinities with fine accuracy (2 kcal/mol). Nevertheless, it is still useful to characterize the most important forces implicated in the binding process. The addition of entropy to the final binding free energy by means of a normal mode analysis lacks of consistency, with respect to the number of structures to evaluate and to the extension of the required reduced system. Moreover, this term has the largest variation among the other contributions to the binding free energy.

Finally, more development in terms of theory and algorithms is required in order to improve the results of this method.

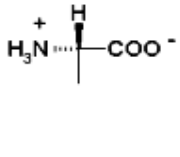
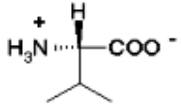
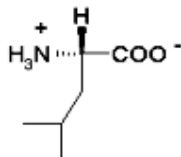
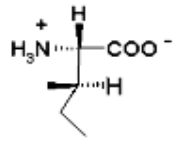
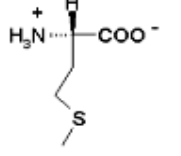
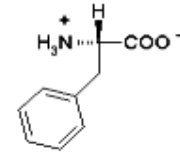
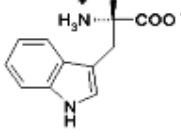
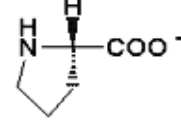
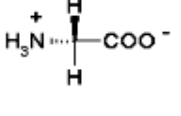
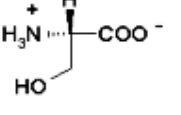
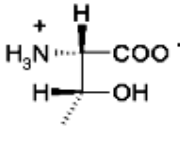
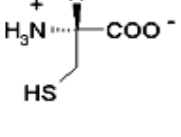
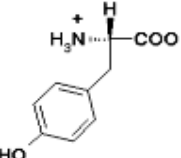
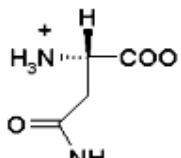
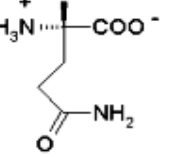
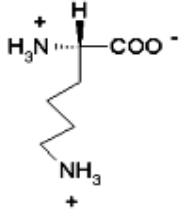
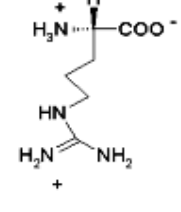
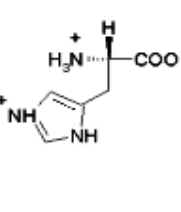
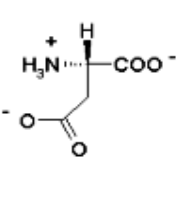
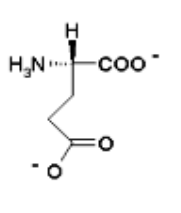
◆ **With respect to Inhibitor Design by means of Molecular Modeling:**

Molecular Modeling methods can be very useful to search new active molecules inhibitors of target proteins. It will be even more important due to the fast growth of solved crystallographic data and the continuous improvement of theoretical methods.

By disrupting different protein complexes, formed by the proteins XIAP, Survivin, Glucose-6-Phosphate Dehydrogenase and Transketolase, new promising molecules have been found. Biological activities at extract and cell level have been confirmed and a successful rate close to 50 % has been identified by means of our virtual screening protocol.

APPENDIX

Amino Acids Structures and Nomenclature

				
Alanine, Ala, A	Valine, Val, V	Leucine, Leu, L	Isoleucine, Ile, I	Methionine, Met, M
				
Phenylalanine, Phe, F	Tryptophan, Trp, W	Proline, Pro, P	Glycine, Gly, G	Serine, Ser, S
				
Threonine, Thr, T	Cysteine, Cys, C	Tyrosine, Tyr, Y	Asparagine, Asn, N	Glutamine, Gln, Q
				
Lysine, Lys, K	Arginine, Arg, R	Histidine, His, H	Aspartate, Asp, D	Glutamate, Glu, E

**Ruptura del Reconocimiento
Proteína-Proteína en Rutas Tumoraes
mediante Modelización Molecular**

1. INTERÉS DEL PROYECTO Y OBJETIVOS

El cáncer es la segunda causa de muerte por enfermedad en los países industrializados. A pesar de la existencia de métodos eficaces de detección precoz y tratamientos cada vez más efectivos responsables de la reducción de mortalidad, algunos tipos de tumores presentan todavía tratamientos difíciles y bajos índices de supervivencia. Los fármacos convencionales sólo exhiben un índice terapéutico moderado, entre células sanas y tumorales, por ello los avances recientes se centran en encontrar tratamientos menos tóxicos para esta enfermedad. Así pues, los fármacos del futuro deberán incidir en rutas biológicas específicas, involucrando el crecimiento celular y la proliferación descontrolada. Siguiendo este planteamiento, en este trabajo se han seleccionado dos mecanismos biológicos involucrados en el cáncer, llamados **Apoptosis** (o muerte celular programada) y **Ruta de las Pentosas Fosfato**, con el objetivo de encontrar nuevos inhibidores de las proteínas más sensibles de ambas rutas.

La sobreexpresión de genes antiapoptóticos se ha correlacionado con el crecimiento tumoral y la resistencia a los tratamientos habituales. Así, se está trabajando en entender el funcionamiento de dos proteínas importantes de esta ruta, el **XIAP** y el **Survivin**, las cuales se han seleccionado en este trabajo, debido a que todavía no existen fármacos en el mercado que actúen sobre estas dos proteínas y debido a que su interés terapéutico se ha demostrado claramente.

Por otro lado, en este trabajo también se han estudiado las dos proteínas más activas detectadas en la rama oxidativa y no oxidativa de la Ruta de las Pentosas Fosfato, la **Glucosa-6-Fosfato Deshidrogenasa** y la **Transketolasa**.

El objetivo principal ha consistido en aplicar métodos de la Modelización Molecular, que cubren tópicos recientes, como el reconocimiento de péptidos y proteínas, la búsqueda en bases de datos, el anclaje y evaluación del cribado virtual de compuestos y la predicción de energías libres de unión, para encontrar nuevos inhibidores de las proteínas XIAP, Survivin, Glucosa-6-Fosfato Deshidrogenasa y Transketolasa.

En esta sección se presenta un breve resumen de los métodos aplicados y los objetivos alcanzados en la presente Tesis.

2. MÉTODOS

Los métodos teóricos empleados en este proyecto pertenecen principalmente a tres ramas de la ciencia, la Mecánica Cuántica, la Mecánica Molecular y el Diseño de Fármacos mediante Modelización Molecular.

La Mecánica Cuántica [1] describe con rigor el comportamiento a nivel atómico-molecular pero actualmente no se puede aplicar a grandes sistemas como las biomoléculas. Aún así, se ha utilizado esta metodología para estudiar pequeñas moléculas involucradas en nuestros sistemas biológicos, básicamente cofactores de las proteínas, metales estructurales de las mismas, o inhibidores de su actividad. Principalmente se ha usado la aproximación Hartree-Fock (HF) y el método semiempírico AM1 para encontrar las geometrías de mínimo energético de estas moléculas y encontrar sus cargas (AM1-BCC). Y poder incluir esta información en nuestros sistemas biológicos de interés.

Respecto a la Mecánica Molecular, ésta se basa en aplicar un potencial sencillo a los sistemas y requiere de un menor coste computacional, y por esta razón ha sido la principal metodología aplicada en este trabajo. La energía potencial de un sistema tratado con Mecánica Molecular se conoce como campo de fuerzas, en concreto, nuestros trabajos utilizan el campo de fuerzas implementado en el paquete de programas AMBER7 [2] con la parametrización parm94 [3]:

$$E_{\text{tot}} = \sum k (R - R_0)^2 + \sum k (\Theta - \Theta_0)^2 + \sum \frac{V_n}{2} [1 + \cos(n\omega - \phi)] + \sum \sum [(A_{ij}/R_{ij})^{12} - (B_{ij}/R_{ij})^6] + \sum \sum \frac{q_i q_j}{R_{ij}} \quad (1)$$

Donde el primer sumatorio expresa el potencial debido al movimiento de los átomos enlazados, el segundo el debido al movimiento angular de tres átomos enlazados, el tercero el debido a la torsión que involucra cuatro átomos, el cuarto expresa la interacción de van der Waals mediante el potencial de Lennard-Jones 12-6, y el último término describe las interacciones electrostáticas representadas por cargas puntuales.

Además la evolución temporal del sistema se ha tratado mediante la llamada Dinámica Molecular (DM), que utiliza las ecuaciones clásicas del movimiento más el potencial descrito mediante Mecánica Molecular para estudiar la evolución del sistema a lo largo del tiempo a cierta temperatura y extraer mediante tratamientos estadísticos las propiedades de interés.

Actualmente existen diferentes implementaciones en los programas de Dinámica Molecular para reducir los costes computacionales y también para tratar de forma más realista los sistemas. Así todas las simulaciones de Dinámica Molecular se han llevado a cabo aplicando una solvatación a la proteína mediante la inclusión de una caja cúbica de aguas TIP3P [4], además de utilizar condiciones periódicas de contorno para minimizar el efecto de la dimensión finita del sistema. Por otro lado el número de interacciones de van der Waals se reduce a aquellas que se dan solo dentro de la llamada distancia de corte (cut-off) mientras que las interacciones electrostáticas se obtienen mediante el método de la suma de Ewald (PME) [5].

La inclusión de la temperatura se lleva a cabo mediante el acoplamiento del sistema a un baño térmico mediante el algoritmo de Berendsen [6], y normalmente se trabaja en el colectivo canónico, a número de partículas, volumen y temperatura constante de 300 K (NVT).

Por último, se han utilizado diferentes herramientas teóricas aplicadas al diseño de fármacos, y en particular a caracterizar las interacciones ligando-proteína, péptido-proteína o proteína-proteína.

Básicamente nuestro objetivo ha sido diseñar inhibidores de tipo competitivo, ya que actúan de la misma manera que el inhibidor natural o de la misma manera que la proteína que interactúa con nuestra proteína de interés biomédico. En este contexto se ha empleado el concepto de farmacóforo (Figura 1), como aquel conjunto de propiedades electrónicas y estéricas necesarias para asegurar una interacción óptima entre el ligando y el receptor, para inhibir o activar su acción biológica. El farmacóforo incluye también la información geométrica de estas propiedades.

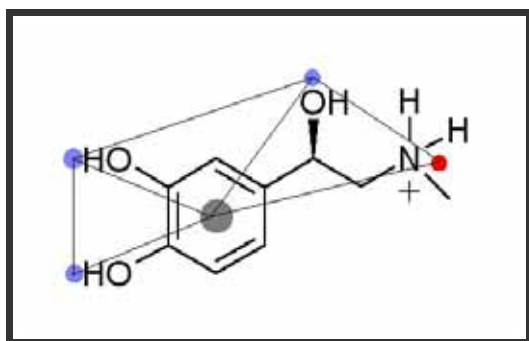


Figura 1: Farmacóforo de cinco puntos de la molécula de adrenalina, mostrado como ejemplo.

Una vez determinado el farmacóforo responsable de la interacción proteína-proteína, se llevan a cabo búsquedas en bases de datos 3D, para encontrar moléculas pequeñas con un farmacóforo similar y por tanto con la potencial capacidad de actuar imitando sus interacciones biológicas. En este proceso de búsqueda se utiliza el programa CATALYST [7], que dispone de un gran número de compuestos comerciales, que sintetizan las diferentes empresas. Normalmente se pueden encontrar entre 100 y 5000 moléculas con un farmacóforo similar, dependiendo de lo complejo que éste sea, por tanto el siguiente paso es discriminar cuales de ellas pueden tener una actividad mayor frente a la proteína. Para ello se ha utilizado la técnica de docking, o anclaje de los compuestos al receptor, mediante un programa diseñado en nuestro grupo (Dock_Dyn [8]). Éste está especialmente pensado para encontrar moléculas peptidomiméticas, es decir, que mantengan interacciones similares a las que presentan los péptidos o las proteínas y se basa en añadir las restricciones geométricas que impone un farmacóforo. Por último la capacidad de unión de los compuestos se evalúa mediante la desviación respecto dicho farmacóforo de referencia (empleando el RMS) y mediante la función de score Xscore [9], desarrollada para tratar un gran número de compuestos sobre un receptor rígido.

Dentro del diseño y evaluación de nuevos compuestos y del conocimiento del reconocimiento de proteínas, es muy importante disponer de un método teórico de evaluación de la energía libre de unión receptor-ligando, ya que un valor elevado de esta propiedad es fundamental para obtener un fármaco selectivo y nos puede servir para caracterizar las uniones entre proteínas. En este trabajo se ha utilizado la metodología MMPB(GB)SA (Molecular Mechanics Poisson Boltzmann Solvation Area) [10] para evaluar esta propiedad.

Este método requiere realizar al menos una Dinámica Molecular y extraer un número adecuado de estructuras de un complejo ligando-receptor para evaluar su energía libre de unión, incluyendo la solvatación que se trata mediante un modelo continuo basado en la resolución de la ecuación de Poisson-Boltzman (PB) o bien mediante la aproximación Generalizada de Born (GB). Las ecuaciones básicas que definen este método son:

$$\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^0 + \Delta G_{\text{RL}}^{0 \rightarrow \text{sol}} - \Delta G_{\text{R}}^{0 \rightarrow \text{sol}} - \Delta G_{\text{L}}^{0 \rightarrow \text{sol}} \quad (2)$$

$$\Delta G_{\text{bind}}^0 = \Delta H_{\text{bind}}^0 - T \Delta S_{\text{bind}}^0 \quad (3)$$

$$\Delta G^{0 \rightarrow \text{sol}} = \Delta G_{\text{elec}}^{0 \rightarrow \text{sol}} + \Delta G_{\text{np}}^{0 \rightarrow \text{sol}} \quad (4)$$

$$\nabla [\epsilon(r) \nabla \phi(r)] - K' [\phi(r)] = -4\pi \rho(r) \quad (5)$$

$$\Delta G^{0 \rightarrow \text{sol}}_{\text{np}} = \gamma \text{SASA} - b \quad (6)$$

Donde la ecuación 2 describe el marco general del método que calcula la energía libre de unión como evaluación de la energía libre de unión en el vacío (ΔG^0_{bind}) más los términos derivados de la solvatación del sistema. La primera contribución se calcula mediante la expresión 3, donde la componente entrópica se puede estimar mediante un análisis de modos normales, y la entálpica se extrae del campo de fuerzas. La solvatación en medio continuo se describe diferenciando su contribución polar y no-polar (ecuación 4). Por último, la expresión 5 es la ecuación de Poisson-Boltzmann linealizada que se utiliza para encontrar la contribución polar a la solvatación, y la ecuación 6 se utiliza para evaluar la contribución no polar a la solvatación, mediante una relación empírica que depende del área superficial accesible al solvente (SASA).

A continuación se mostrarán los resultados más importantes concernientes al estudio de las proteínas XIAP, Survivin, Transketolasa y Glucosa-6-Fosfato Deshidrogenasa.

3. ESTUDIO DE LAS PROTEÍNAS ANTIAPOPTÓTICAS XIAP Y SURVIVIN:

Descripción de su interacción con Smac/DIABLO y búsqueda de compuestos inhibidores

Las proteínas XIAP y Survivin están involucradas en la apoptosis [11], que es el mecanismo de muerte celular programada y por el cual el organismo mantiene balanceado el número de células.

Este mecanismo se lleva a cabo mediante dos vías complementarias, la llamada vía extrínseca se produce con la activación de determinados receptores de membrana (CD95, TNF y TRAIL), que a su vez, activan una cascada enzimática de Caspasas (Caspasas 8, 10, 3 y 7). Estas proteínas son las encargadas de producir los cambios morfológicos de una célula en apoptosis. Existe también la vía intrínseca que funciona mediante la activación de algunas proteínas mitocondriales, como el Citocromo-c o el Smac/DIABLO. Se han descubierto una familia de proteínas, llamadas inhibidoras de la apoptosis o IAPs [12], que regulan tanto la vía extrínseca como la intrínseca. Entre ellas destacan el XIAP y el Survivin. A su vez, el Smac/DIABLO se une a las IAPs inhibiéndolas y por tanto produciendo apoptosis. El complejo formado por Smac/DIABLO y XIAP se ha resuelto tanto por

técnicas de Rayos X [13] como por RMN [14], encontrando que esta proteína se une al XIAP mediante sus cuatro primeros residuos N-terminales, AVPI. Por otro lado se ha sugerido que el Smac/DIABLO también se une al Survivin de la misma manera e inhibe su actividad [15].

Estos mecanismos celulares se ven alterados en una célula tumoral, así diferentes tumores sobreexpresan XIAP y otros componentes que permiten que la célula tumoral resista a la apoptosis. La proteína Survivin se expresa específicamente en tejido tumoral, siendo una diana muy atractiva para desarrollar fármacos selectivos solo al tejido dañado. Debido a que se conoce la estructura del complejo de XIAP con Smac/DIABLO, se han propuesto muchas moléculas que imitan la secuencia AVPI del Smac/DIABLO, inhiben a esta IAP y por tanto inducen apoptosis especialmente en las células tumorales [16-23]. Por otra parte, por el momento ninguna de ellas ha alcanzado las fases clínicas finales, y muchas de ellas tienen poca permeabilidad a la membrana celular debido a su carácter peptídico.

En este trabajo se llevaron a cabo simulaciones de Dinámica Molecular tanto del complejo Smac/DIABLO-XIAP extraído del Protein Data Bank (PDB, 1G3F), con 9 residuos en el ligando, como del complejo Smac/DIABLO-Survivin el cual se contruyó mediante modelado por homología, superponiendo el dominio BIR3 del XIAP al dominio BIR del Survivin y adaptando así las coordenadas del Smac/DIABLO al dominio de unión del Survivin. La Figura 2 muestra visualmente el fragmento AVPI del Smac/DIABLO interaccionando con estos dos receptores.

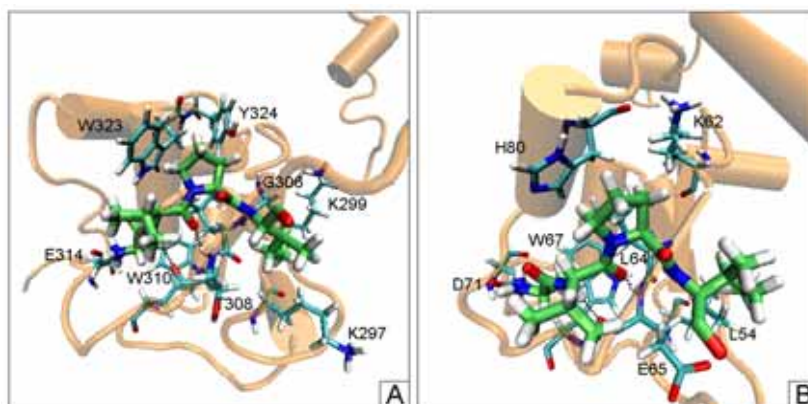


Figura 2: Interacciones más importantes de los complejos Smac/DIABLO-XIAP (A) y Smac/DIABLO-Survivin (B). Únicamente se representa la secuencia AVPI del ligando. Los receptores se representan en color naranja mientras que los átomos de carbono de los ligandos se muestran en verde claro.

Una particularidad de la simulación es el tratamiento que se utilizó para el metal (zinc) estructural que presentan tanto el XIAP como el Survivin. Para ello se escogió la parametrización que desarrolló Pang [24] que impone la coordinación mediante átomos tipo dummy, y combina el modelo enlazante (en la interacción zinc-dummy) con el modelo no enlazante (en la interacción zinc-proteína). Se muestra a continuación la estructura que se utilizó para tratar esta zona.

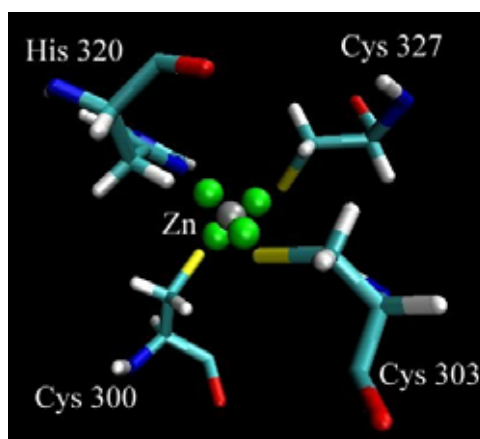


Figura 3: Estructura tetraédrica del sistema zinc-dummies y coordinación del XIAP.

Se utilizó la simulación de DM para estudiar las interacciones más estables que se producían entre Smac/DIABLO y ambos receptores. Se localizaron tres puentes de hidrógeno estables en el complejo con XIAP formados por el residuo T308 del XIAP con V2 del Smac/DIABLO y G306 del XIAP con I4 del Smac/DIABLO. Con respecto al Survivin se localizaron sólo dos puentes, formados por E65 del receptor y V2 del ligando. Además se calcularon las interacciones de van der Waals y electrostáticas de los nueve primeros residuos del Smac/DIABLO con ambos receptores (Figura 4) revelando la fuerte interacción hidrofóbica de la secuencia AVPI y la alta energía electrostática del primer residuo, entorno a -60 kcal/mol reconociendo al XIAP y cercana a -100 kcal/mol reconociendo al Survivin.

De hecho, se calculó un mapa de potencial electrostático (Figura 5) para ambos receptores encontrado que el dominio BIR del Survivin tiene un potencial mucho más negativo, debido a la mayor presencia de residuos ácidos. Esto dificulta el diseño de fármacos, puesto que el dominio de unión al Smac/DIABLO del Survivin presentaría una energía de desolvatación muy desfavorable influyendo en

la energía libre de unión del complejo. Este hecho se confirmó aplicando el protocolo MMGBSA [10] a ambos sistemas.

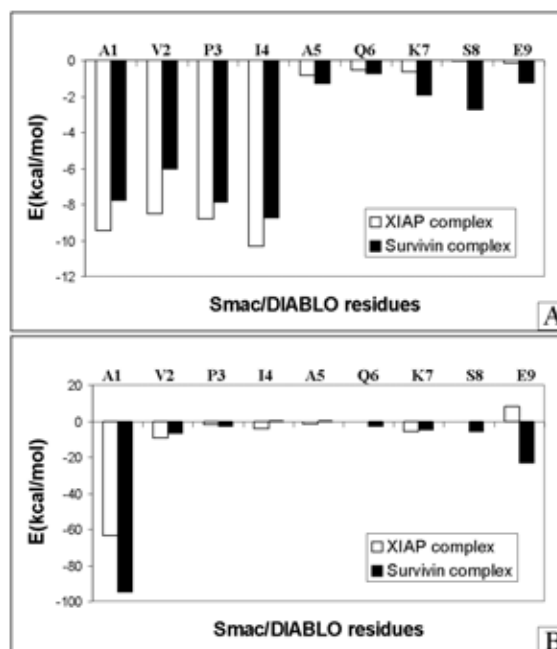


Figura 4: Interacciones medias de van der Waals (A) y electrostáticas (B) de los residuos de Smac/DIABLO en ambos receptores, XIAP en blanco y Survivin en negro.

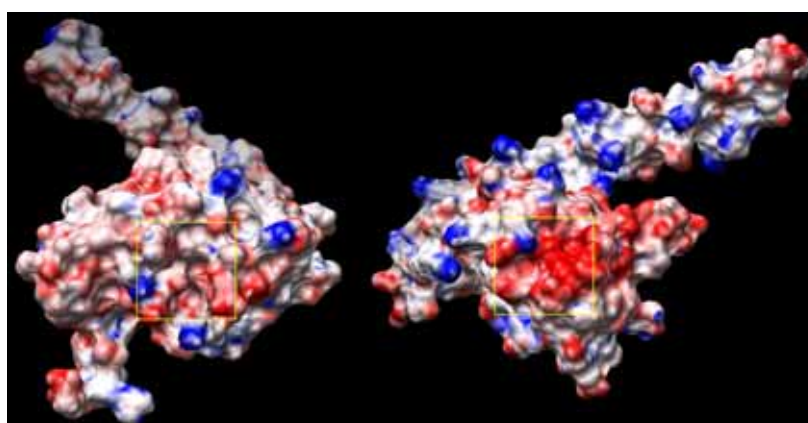


Figura 5: Potencial electrostático coloreado para el XIAP (izquierda) y para el Survivin (derecha). El potencial electronegativo se muestra en rojo, el electropositivo en azul y el neutro en blanco. Se remarcan los dominios BIR en un cuadro amarillo.

Mediante este mismo protocolo se evaluó la energía libre de unión entre los dos complejos encontrando una diferencia entre ellos de 2.29 kcal/mol, en buen acuerdo con el valor experimental de 2.39 kcal/mol y revelando el buen funcionamiento de este método en este sistema.

Adicionalmente se evaluó la metodología MMPBSA [10] correlacionando la energía libre de unión experimental de 6 tetrapéptidos basados en la secuencia N-terminal de la proteína Smac/DIABLO (AVPI, AVPY, AVPA, AVPE, AGPI y ARPF) y el propio péptido de 9 residuos, encontrando un coeficiente de correlación de 0.86 cuando no se incluía la componente entrópica.

Todos estos análisis nos sirvieron para describir la superficie de contacto entre Smac/DIABLO y los receptores Survivin y XIAP, y formular un farmacóforo del ligando formado por 8 puntos en el caso de que interaccione con el XIAP, y por los 7 primeros puntos en caso de interacción con la proteína Survivin (Figura 6). Adicionalmente, incluir las desviaciones de los puntos que configuran el farmacóforo a lo largo de la DM nos permitió incluir la flexibilidad de los contactos en el proceso de búsqueda de compuestos y de anclaje al receptor.

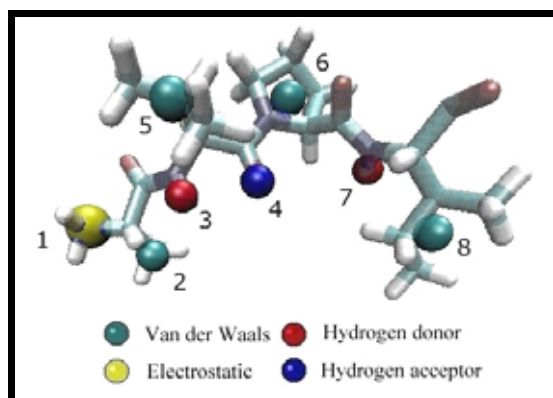


Figura 6: Farmacóforo de Smac/DIABLO.

Así pues, se realizó una búsqueda en bases de datos de moléculas pequeñas que presentaran un farmacóforo similar al encontrado, pero reducido a los 6 y 4 primeros puntos. Se utilizó el programa CATALYST [7] para encontrar 132 moléculas orgánicas con la capacidad de actuar como miméticos de la proteína Smac/DIABLO. Después de realizar los estudios de anclaje, utilizando nuestro programa Dock_Dyn [8], y evaluar la energía de unión mediante el método Xscore [9] seleccionamos los mejores 8 compuestos y fueron comprados de las diferentes empresas.

En colaboración con el grupo experimental de Hematopatología del Hospital Clínic de Barcelona, coordinado por la Dra. Dolors Colomer, el Dr. Roberto Alonso del mismo grupo nos realizó las pruebas de actividad de los 8 compuestos con diferentes líneas tumorales. Las líneas celulares escogidas fueron de linfoma de manto y de leucemia linfocítica. Los resultados fueron exitosos, encontrando que 4 de nuestros compuestos (50%) inducían apoptosis, en rangos medios de entre 14-28 μM para el compuesto A y 50-100 μM para los compuestos D, F y G. Además los compuestos mostraban una actividad mejorada cuando se administraban conjuntamente a la proteína TRAIL lo cual es una prueba indirecta de que los compuestos inhiben la proteína XIAP.

En el futuro se espera caracterizar la actividad de los compuestos a nivel molecular, confirmando su actividad frente a XIAP y Survivin, poder cuantificar su selectividad y proceder a mejorar los compuestos utilizando herramientas del diseño de fármacos.

4. ESTUDIO DE LA PROTEÍNA TRANSKETOLASA:

Modelización parcial de la Transketolasa humana, reconocimiento de los monómeros en zonas conservadas, deducción del farmacóforo y búsqueda de compuestos inhibidores.

La proteína Transketolasa cataliza la transferencia reversible de dos carbonos entre substratos de tipo cetosa hacia substratos aceptores de tipo aldosa. Esta proteína es la enzima más sensible de la Ruta no oxidativa de las Pentosas Fosfato [25], que se encarga de generar moléculas de ribosa que son a su vez un metabolito esencial requerido en la síntesis de ADN. Se sabe que las células tumorales necesitan de una síntesis elevada de ADN para su proliferación y esta ruta les proporciona uno de los componentes básicos para ello. Así, estudios de control metabólico encontraron que el coeficiente de control tumoral de la Transketolasa era de 0.9 [25], el más alto de toda la ruta de pentosas. Por estas razones seleccionamos esta proteína para su estudio, más aún, cuando es una proteína poco estudiada y las moléculas inhibidoras descubiertas hasta el momento no son adecuadas, debido a que son poco activas y poco selectivas ya que se basan en mimetizar su cofactor, el pirofosfato de tiamina.

Los trabajos más exhaustivos realizados con esta proteína se han centrado en las variantes de levadura (*S.Cerevisiae*), *E.Coli*, cuyas estructuras se determinaron por difracción de Rayos X [28-30], pero desafortunadamente todavía no existe la estructura resuelta de la Transketolasa humana.

Inicialmente nuestro trabajo se centró en modelar parcialmente la Transketolasa humana basándonos en la estructura experimental de la variante de levadura. No se llevó a cabo el modelado completo por homología porque la similitud de secuencia entre estas dos variantes es muy baja, entorno al 27 %.

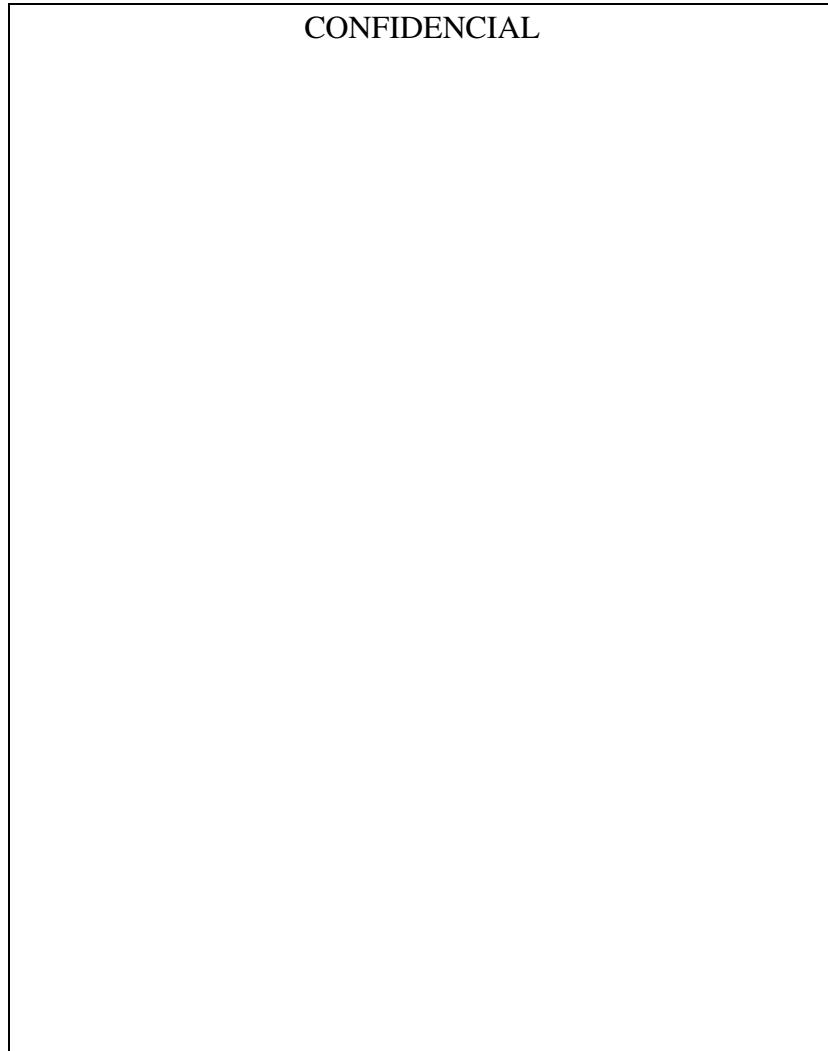


Figura 7: Alineamiento de las secuencias de Transketolasa de levadura (negro) y humana (rojo) extraídas del alineamiento múltiple de la referencia [26]. Los residuos conservados o altamente similares se marcan en gris y gris claro respectivamente. ***) Marca la secuencia de la hélice alfa conservada. +++) Marca la secuencia de la hélice alfa que contiene R401 y >>>) Marca la secuencia del bucle conservado.

La Figura 7 muestra el alineamiento entre la variante humana y de levadura que hemos utilizado, extraído de la referencia [26]. Adicionalmente esta figura muestra las tres secuencias que seleccionamos para su modelización. Éstas requerían cumplir dos condiciones fundamentales; ser zonas

pertenecientes a la superficie de dimerización de la enzima y tener una conservación alta respecto a la variante de levadura. Así, se encontraron dos zonas importantes de contacto entre dímeros, la primera formada por una hélice alfa que contenía un residuo arginina muy crítico para la actividad de la Transketolasa de rata [27] y el cual se reconocía por el llamado bucle conservado (conserved loop) del otro monómero, Figura 7. La segunda zona modelada, está formada por dos hélices alfa antiparalelas (Figura 7). Una vez realizadas las mutaciones para tener estas dos zonas con su secuencia humana, pero con una estructura global de levadura, se preparó el sistema para realizar una simulación por Dinámica Molecular. Esta proteína presenta un cofactor, el pirofosfato de tiamina, para el cual se optimizaron sus parámetros de campo de fuerzas y especialmente sus cargas. La dinámica molecular nos permitió asegurar la estructura de las zonas modeladas y examinar los contactos más estables entre ellas. Se encontraron 4 puentes de hidrógeno entre la hélice alfa que contiene R y el bucle conservado, mientras que se encontraron 3 puentes en la zona de dimerización formada entre las dos hélices alfa conservadas. Posteriormente se analizaron las interacciones intermoleculares medias de la simulación para cada residuo de la hélice alfa con R y de la hélice alfa conservada, esta información se muestra en la Figura 8.

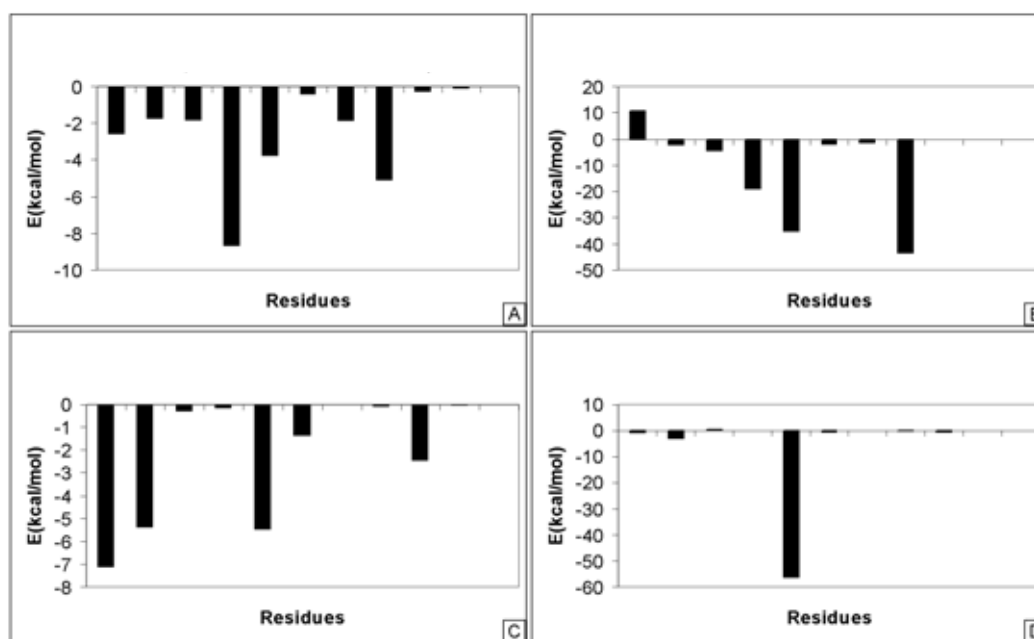


Figura 8: Energías media de van der Waals (izquierda) y electrostáticas (derecha) de la secuencia de la hélice alfa conservada (A y B) y de la hélice alfa que contiene R (C y D).

Este análisis nos permitió deducir los contactos más estables que se establecían entre estas dos zonas de la proteína, tal como podrían presentarse en la variante humana, y configurar dos farmacóforos para describir la interacción en estas zonas. El farmacóforo perteneciente a la hélice alfa conservada con R (Figura 9) consta de 7 puntos de interacción que involucran 4 residuos, mientras que el farmacóforo identificado en la hélice alfa conservada es más simple, formado por 5 puntos de interacción (Figura 10) que involucran 3 residuos.

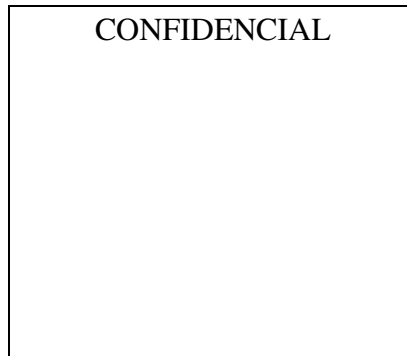


Figura 9: *Farmacóforo de la hélice alfa que contiene R401. Los puntos 2, 3 y 7 marcan los contactos de van der Waals, el punto 1 marca un puente de hidrógeno aceptor, los puntos 4, 5,6 marcan tres puentes de hidrógeno dadores.*

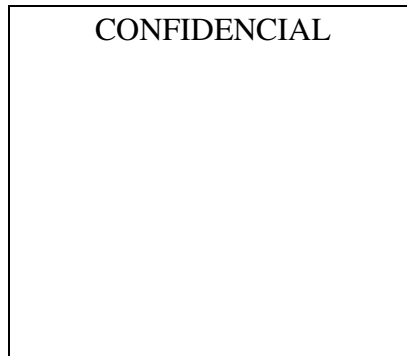


Figura 10: *Farmacóforo de la helice alfa conservada. El punto 2 marca un contacto de van der Waals, los puntos 4 y 5 marcan dos puentes de hidrógeno aceptores y los puntos 1 y 3 marcan dos puentes de hidrógeno dadores.*

Siguiendo el mismo protocolo que el utilizado para las proteínas XIAP y Survivin, se llevó a cabo una búsqueda en las bases de datos 3D del programa CATALYST [7] de moléculas con un farmacóforo similar a los encontrados. En concreto se diseñó un farmacóforo con los 5 puntos de contacto de la hélice alfa conservada y se encontraron 131 moléculas de bajo peso molecular, con la capacidad de actuar mimetizando esta zona de la proteína y por tanto con la presumible capacidad de competir en la formación de un complejo con un monómero de Transketolasa. Posteriormente, utilizando el programa de docking Dock_Dyn [8] y la función de scoring Xscore [9] se seleccionaron las 9 mejores moléculas (llamadas T1-T9), que debían mantener los contactos de la hélice alfa conservada.

En colaboración con el grupo experimental de Bioquímica Integrativa y Terapia contra el Cáncer UB de la Prof. Marta Cascante, y en particular Gema Alcarraz-Vizán, se realizaron los test de actividad de estos 9 compuestos tanto a nivel de extracto celular (fracción de proteínas) como a nivel celular seleccionando dos líneas tumorales de carcinoma de colon. Se encontraron dos compuestos activos, T1 y T2, presentando valores de actividad IC_{50} de 500 μ M a nivel de extracto celular. A nivel celular el compuesto T2 resultó muy activo con un IC_{50} de 10 μ M, sugiriendo que dicho compuesto actuaba en más vías además de en la ruta de las pentosas fosfato.

Adicionalmente se diseñaron 16 derivados del compuesto T2 para aumentar su actividad y se probaron experimentalmente los 5 mejores candidatos, dos de ellos (T2-A y T2-B) presentaron una actividad mejorada respecto al compuesto de partida T2.

5. ESTUDIO DE LA PROTEÍNA GLUCOSA-6-FOSFATO DESHIDROGENASA:

Reconocimiento de los monómeros, diseño de péptidos inhibidores, diseño de compuestos inhibidores no peptídicos.

La Glucosa-6-Fosfato Deshidrogenasa (G6PDH) es una enzima dependiente de NADP⁺ involucrada en la vía oxidativa de la Ruta de las Pentosas Fosfato [25]. En concreto, cataliza la transformación de glucosa-6-fosfato en la primera etapa de esta vía, que en última instancia se utiliza para sintetizar moléculas de ribosa. Se ha sugerido la G6PDH como una nueva diana para el tratamiento de cáncer, debido que actúa en la etapa limitante de velocidad de la vía oxidativa [31] que proporciona estas moléculas de ribosa para sintetizar ácidos nucleicos. Como se ha comentado, las células tumorales

necesitan una cantidad mayor de ácidos nucleicos para su crecimiento descontrolado, y esta enzima les proporciona uno de los componentes esenciales. Así, al inhibir G6PDH, al igual que ocurre con la proteína Transketolasa, se limita el crecimiento tumoral. Actualmente no existen buenos inhibidores de esta proteína, el metotrexato [32] y la dehidroepiandrosterona [31] son dos ejemplos, pero el primero es un compuesto poco selectivo ya que inhibe otras enzimas dependientes de NADP+, y el segundo compuesto pertenece a la familia de las hormonas, y su administración provocaría muchos efectos secundarios. Se conocen otros compuestos inhibidores de la G6PDH, pero presentan una actividad baja y además se desconoce su modo de unión a esta enzima.

En este trabajo, se partió de la estructura dimérica de la G6PDH humana (Figura 11), resuelta por difracción de rayos X (con código PDB 1QKI), y se estudió a fondo su superficie de dimerización, compuesta por los dominios alfa y beta, que cubren un total de 64 residuos.

El cofactor de esta proteína, NADP+, se adaptó al campo de fuerzas utilizando una parametrización ya descrita [33].

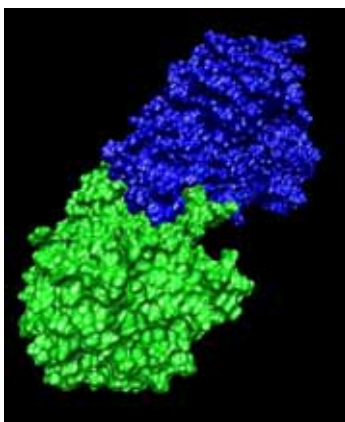


Figura 11: Estructura dimérica de la G6PDH humana.

El sistema se minimizó adecuadamente y se realizó una Dinámica Molecular de 1 ns. Se utilizaron los resultados del tiempo de producción para estudiar la superficie de dimerización de la enzima, para posteriormente poder diseñar compuestos que pudieran romper la estabilidad de este dímero, y actuar como nuevos inhibidores de la G6PDH humana.

Así, se encontraron en el dominio alfa de dimerización un total de 8 puentes de hidrógeno (3 dadores y 5 aceptores) y un total de 20 puentes de hidrógeno (10 dadores y 10 aceptores) en el dominio de dimerización beta.

Por otro lado, este análisis se completó, calculando las energías intermoleculares medias de interacción de cada residuo de los dos dominios, con todos los residuos del monómero complementario. La Figura 12 muestra el patrón de energías electrostáticas para los dos dominios, mientras que la Figura 13, muestra el mismo análisis pero concerniente a las energías de van der Waals.

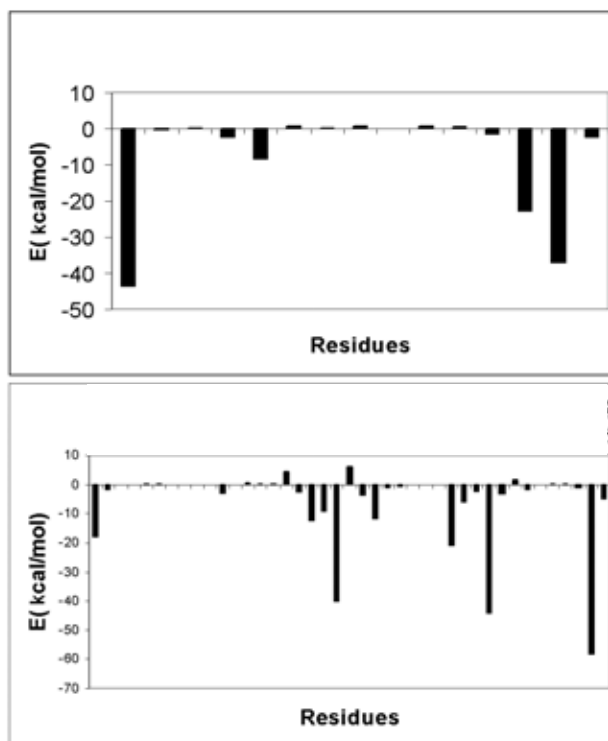


Figura 12: Energías medias de interacción electrostática para los residuos de dimerización del dominio alfa (arriba) y beta (abajo) de G6PDH.

Estos resultados nos permitieron encontrar donde se localizaban las interacciones más importantes y poder diseñar un total de 7 péptidos cortos, con fragmentos de la secuencia de la proteína, que pudieran competir por la formación del dímero desplazando uno de los monómeros de G6PDH. Esta aproximación, aunque ya descrita hace 10 años [34], no ha sido adecuadamente explotada.

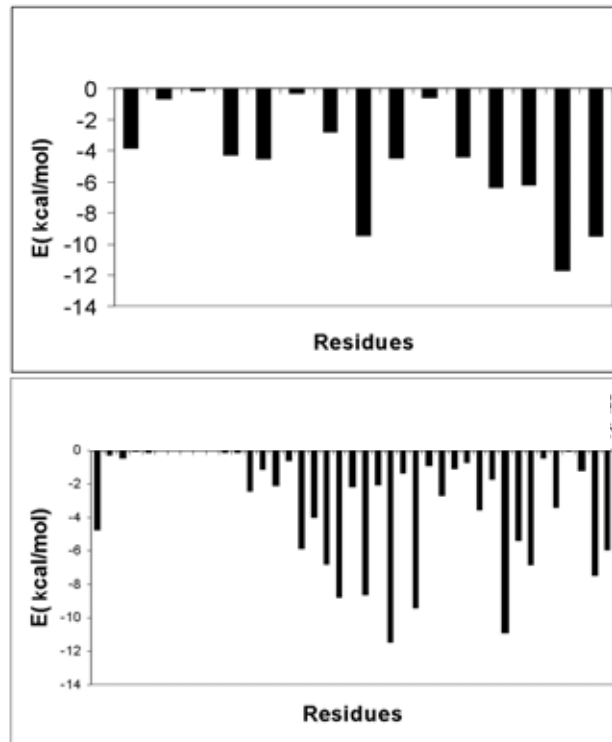


Figura 13: Energías medias de interacción de van der Waals para los residuos de dimerización del dominio alfa (arriba) y beta (abajo) de G6PDH.

Así, se construyeron 6 péptidos lineales más un péptido de secuencia cíclica basado en el péptido 4 (Figura 14), que debía mantener las interacciones más importantes detectadas en el dímero de esta proteína:

- **Péptido 1**, 16 residuos.
- **Péptido 2**, 13 residuos.
- **Péptido 3**, 14 residuos.
- **Péptido 4**, 7 residuos.
- **Péptido Cíclico**, 9 residuos.
- **Péptido 5**, 12 residuos.
- **Péptido 6**, 10 residuos.

Los péptidos 1 y 2 cubren el dominio de dimerización alfa, mientras que los demás cubren las zonas

que encontramos como más importantes del dominio de dimerización beta.

Cada uno de los complejos péptido-G6PDH, se construyó directamente a partir del dímero de la G6PDH, manteniendo una estructura inicial con las mismas interacciones. Los sistemas se minimizaron y se llevó a cabo una simulación de Dinámica Molecular de 2 ns para cada sistema. Posteriormente se analizaron sus interacciones, tanto las interacciones por puente de hidrógeno, como los contactos de van der Waals y electrostáticos. En general, los contactos iniciales se mantuvieron durante las simulaciones, aunque el patrón de puentes de hidrógeno se modificó ligeramente en algunos casos, pero este comportamiento es previsible, teniendo en cuenta que la flexibilidad de los péptidos es mucho mayor que la flexibilidad de la misma secuencia en la proteína. Por último para complementar este análisis, se evaluó la energía libre de unión de los diferentes sistemas péptido-G6PDH mediante la metodología MMPB(GB)SA.

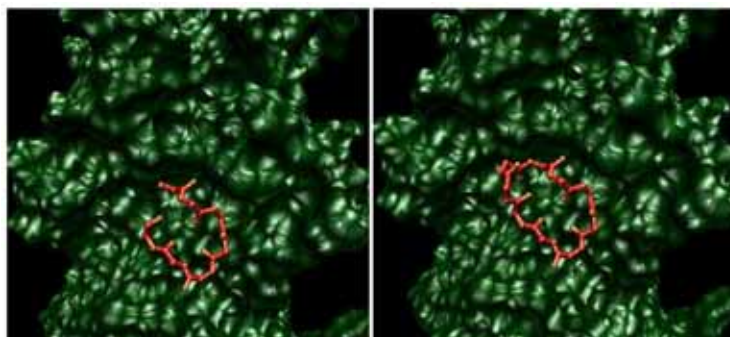


Figura 14: Construcción del Péptido Cíclico (derecha) añadiendo dos residuos glicina al Péptido 4 (izquierda). El monómero de G6PDH se muestra en verde, y los esqueletos de ambos péptidos en rojo.

Este análisis nos permitió sugerir que los Péptidos 2, 3 y el Péptido Cíclico presentaban una energía libre de unión más favorable, y así estos tres péptidos se sintetizaron en el servicio de síntesis de péptidos del Parc Científic de Barcelona. En colaboración con el grupo experimental de Bioquímica Integrativa y Terapia contra el Cáncer de la Prof. Marta Cascante, y en particular Gema Alcarraz-Vizán, se realizaron pruebas preliminares para identificar la actividad del Péptido Cíclico como inhibidor de G6PDH. Aunque se utilizó una G6PDH de levadura, y el péptido está diseñado para inhibir la variante humana, este compuesto presentó una inhibición clara, aunque baja, de G6PDH. Posteriormente, se realizaron ensayos sobre células humanas con resultado negativo, esto nos indujo a pensar que el

Péptido Cíclico no es permeable a la membrana celular. Para mejorar el Péptido Cíclico, tanto a nivel del reconocimiento por el monómero de la G6PDH, como para intentar aumentar la permeabilidad celular, se diseñaron 4 derivados sustituyendo el tercer o quinto residuo de sus secuencia (Res3F, Res5A, Res5L y Res5F), y un derivado sustituyendo dos residuos (Res3F_Res5L). Se llevaron a cabo las simulaciones mediante Dinámica Molecular de estos nuevos sistemas Péptido Cíclico-G6PDH y se estimó la energía libre de unión de cada uno. Los resultados indicaron que el péptido con dos sustituciones debía unirse con mayor afinidad a la proteína. Su síntesis y evaluación se considerará en un futuro.

Por otro lado y de manera complementaria a este estudio, se buscaron también inhibidores de la dimerización de la G6PDH humana de carácter no peptídico, utilizando un protocolo similar al ya descrito para las proteínas XIAP, Survivin y Transketolasa. En este caso, se seleccionó un farmacóforo que cubría las interacciones más importantes del dímero de la G6PDH, centrado en el dominio de dimerización beta (Figura 15). Este corto fragmento de la proteína mantenía 11 puntos de interacción estables a lo largo de la DM del sistema dímero de G6PDH.

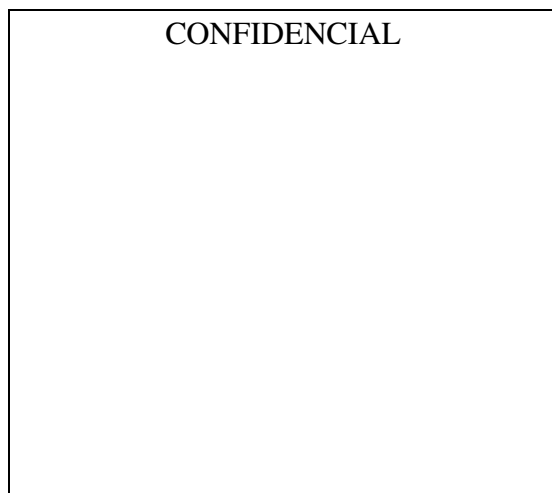


Figura 15: Farmacóforo seleccionado de G6PDH. Los puntos 4, 7, 10 y 11 marcan los contactos de van der Waals, el punto 5 marca una interacción electrostática, los puntos 2, 3 y 8 marcan los puentes de hidrógeno dadores, por último los puntos 1, 6 y 9 marcan los puentes de hidrógeno aceptores.

Se escogió el subconjunto de puntos 7, 8, 9 y 10, para maximizar la búsqueda de compuestos con el programa CATALYST [7]. Así, se encontraron 4298 moléculas con un farmacóforo de 4 puntos similar al que presenta la proteína y en los márgenes que presentan estos puntos durante la simulación dinámica. Estos compuestos fueron anclados al monómero de G6PDH, tratado como receptor rígido, mediante el programa Dock_Dyn [8], y se realizó un proceso de evaluación energética empleando la función Xscore [9]. Los mejores 8 compuestos, llamados G1-G8, se compraron de las diferentes compañías comerciales y el grupo de la Prof. Marta Cascante (Bioquímica Integrativa y Terapia del Cáncer, UB), en especial Gema Alcarraz-Vizán, nos realizó la pruebas de inhibición frente a la G6PDH humana. Los resultados fueron positivos, encontrando que el 50% de los compuestos presentaban actividad, con un IC_{50} entre 250 μM y 67 μM . Además los compuestos G4 y G5 también presentaron actividad frente a células de carcinoma de colon. Estos compuestos pueden ser un buen punto de partida para el diseño de un fármaco con actividad inhibidora de G6PDH.

Por último, hay que decir, que un tratamiento combinado con inhibidores de Transketolasa y de Glucosa-6-Fosfato Deshidrogenasa, ambos encontrados en este proyecto, debería tener un efecto muy importante en la disminución del crecimiento tumoral. Se espera llevar a cabo estas pruebas, así como etapas de mejora para los compuestos activos G4 y G5.

6. CONCLUSIONES GENERALES

Las siguientes conclusiones generales, que cubren diferentes aspectos de la Modelización Molecular y el diseño de fármacos, pueden extraerse de este trabajo:

♦ Con respecto al Reconocimiento Proteína-Proteína:

Aunque el reconocimiento proteína-proteína se establece mediante un número alto de interacciones distribuidas a lo largo de una superficie compleja, los contactos más importantes se pueden identificar en pequeños clústers. Estas zonas, presentan los residuos con una interacción hidrofóbica mayor y un número pequeño de residuos cargados, involucrados en puentes de hidrógeno. El reconocimiento proteína-proteína se puede romper usando pequeños péptidos derivados de la secuencia natural de la proteína y mediante pequeñas moléculas diseñadas conteniendo puntos de interacción similares.

◆ Con respecto a los Farmacóforos Proteína-Proteína:

Se ha demostrado que un farmacóforo derivado de un complejo proteína-proteína o péptido-proteína se puede utilizar para buscar pequeñas moléculas que actúen mimetizando la proteína y compitiendo por el mismo sitio de unión. Además, un farmacóforo de estas características aumenta el éxito en la búsqueda de compuestos activos, ya que incluye la flexibilidad de los contactos que no se pueden tener en cuenta con un farmacóforo derivado sólo de la estructura cristalina.

◆ Con respecto a la predicción de Energías Libres de Unión:

Se ha aplicado la metodología MMPB(GB)SA para predecir la energía libre de unión en 32 complejos, especialmente en complejos péptidos-proteína. Parece que este protocolo no es capaz de reproducir la energía libre de unión experimental con la exactitud suficiente (2 kcal/mol). A pesar de eso, nos ha sido muy útil para caracterizar las fuerzas más importantes que dirigen el proceso de unión. La adición de la entropía mediante el análisis de modos normales al valor final de la energía libre tiene poca consistencia, principalmente respecto al número de estructuras que se deben evaluar y a la extensión del sistema cortado que es necesario para reducir el coste computacional. Además es la contribución con una dispersión mayor respecto los demás términos que afectan a la energía libre de unión.

Finalmente, creemos que es necesario un mayor desarrollo teórico y de algoritmos para mejorar cuantitativamente los resultados de este método.

◆ Con respecto al Diseño de Inhibidores mediante Modelización Molecular:

Los métodos de Modelización Molecular se presentan como una herramienta muy útil en la búsqueda de nuevos inhibidores para las proteínas de interés biológico. Es de esperar que debido al rápido aumento del número de estructuras cristalinas y a la constante mejora de los métodos teóricos, su importancia irá en aumento. Mediante la ruptura de diferentes complejos, formados por las proteínas XIAP, Survivin, Transketolasa y Glucosa-6-Fosfato Deshidrogenasa, se han encontrado nuevas moléculas activas prometedoras. Se han confirmado sus actividades biológicas con un acierto cercano al 50% mediante la aplicación de nuestro protocolo de cribado virtual.

7. BIBLIOGRAFÍA

1. Modern Quantum Chemistry, Szabo A., Ostlund N.S., New York : Mc Graw-Hill. **1989**.
2. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. D.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Sligh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, Univ. California, San Francisco, **2002**.
3. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D.M.; Spellmeyer, D. C.; Fox, T.; Cadwell, J.W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
4. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
5. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
6. Berendsen, H. J. C.; Postman, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
7. CATALYST™ (Accelrys Inc. USA).
8. Rubio-Martinez, J.; Pinto, M.; Tomas M.S.; Perez, J. J. Dock_Dyn: a program for fast molecular docking using molecular dynamics information. University of Barcelona and Technical University of Catalonia. Barcelona, **2005**.
9. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11-26.
10. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srivasan, J.; Case, D. A.; Cheatham III, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.
11. Salgado, J.; Garcia-Saez, A.; Malet, G.; Mingarro, I.; Perez-Paya, E. Peptides in apoptosis research. *J. Pept. Sci.* **2002**, *8*, 543-560.
12. Salvesen, G. S.; Duckett, C. S. IAP proteins: blocking the road to death's door. *Mol. Cell Biol.* **2002**, *3*, 401-410.
13. Liu, Z.; Sun, C.; Olejniczak, E.; Meadows, R.; Betz, S.; Oost, T.; Herrmann, J.; Wu, J.; Fesik, S.

Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* **2000**, *408*, 1004-1008.

14. Wu, G.; Chai, J.; Suber, T.; Wu, J.; Du, C.; Wang, X.; Shi, Y. Structural basis of IAP recognition by Smac/DIABLO. *Nature* **2000**, *408*, 1008-1012.

15. Sun, C.; Nettlesheim, D.; Liu, Z.; Olejniczak, E. T. Solution structure of human Survivin and its binding interface with Smac/Diablo. *Biochemistry* **2005**, *44*, 11-17.

16. Li, L.; Thomas, R. M.; Olejniczak, E.; Meadows, R.; Betz, S.; Oost, T.; Hermann, J.; Wu, J.; Fesik, S. *Nature* **2000**, *408*, 1004-1008.

17. Wu, G.; Chai, J.; Suber, T. L.; Wu, J. W.; Du, C.; Wang, S.; Shi, Y. Structural basis of IAP recognition by Smac/DIABLO. *Nature* **2000**, *408*, 1008-1012.

18. Kipp, R. A.; Case, M. A.; Wist, A. D.; Cresson, C. M.; Carrell, M.; Griner, E.; Wiita, A.; Albiniak, P. A.; Chai, J.; Shi, Y.; Semmelhack, F.; McLendon, G. L. Molecular targeting of inhibitor of apoptosis proteins based on small molecule mimics of natural binding partners. *Biochemistry* **2002**, *41*, 7344-7349.

19. Glover, C. J.; Hite, K.; DeLosh, R.; Scudiero, D. A.; Fivash, M. J.; Smith, L. R.; Fisher, R. J.; Wu, J.; Shi, Y.; Kipp, R. A.; McLendon, G. L.; Sausville, E. A.; Shoemaker, R. H. A high-throughput screen for identification of molecular mimics of Smac/DIABLO utilizing a fluorescence polarization assay. *Anal. Biochem.* **2003**, *320*, 157-169.

20. Oost, T. K.; Armstrong, R. C.; Al-Assad, A.; Betz, S. F.; Deckweth, T. L.; Ding, H.; Elmore, S. W.; Meadows, R. P.; Olejniczak, E. T.; Oleksijew, A.; Oltersdorf, T.; Rosenberg, S. H.; Shoemaker, A. R.; Tomaselli, K. J.; Zou, H.; Fesik, S. W. Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J. Med. Chem.* **2004**, *47*, 4417-4426.

21. Park, C. M.; Sun, C.; Olejniczak, E. T.; Wilson, A. E.; Meadows, R. P.; Betz, S. F.; Elmore, S. W.; Fesik, S. W. Non-peptidic small molecule inhibitors of XIAP. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 771-775.

22. Sun, H.; Nikolovska-Coleska, Z.; Chen, J.; Chao-Yie Yang, C. Y.; Tomita, Y.; Pan, H.; Yoshioka, Y.; Krajewski, K.; Rollerc P. P.; Wang, S. Structure-based design, synthesis and biochemical testing of novel and potent Smac peptido-mimetics. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 793-797.

23. Li, L.; Thomas, R. M.; Suzuki, H.; De Brabaner, J. K.; Wang, X.; Harran, P. G. A small molecule smac mimic potentiates TRAIL- and TNF α – mediated cell death. *Science* **2004**, *305*, 1471-1474.
24. Pang, Y.; Xu, K.; El Yazla, J.; Prendergast, F. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **2000**, *9*, 1857-1865.
25. Comín-Anduix, B.; Boren, J.; Martinez, S.; Moro, C.; Centelles, J. J.; Trebukhina, R.; Petushok, N.; Lee, W. N.; Boros, L. G.; Cascante, M. The effect of thiamine supplementation on tumour proliferation. A metabolic control analysis study. *Eur. J. Biochem.* **2001**, *268*, 4177-4182.
26. Sundström, M.; Lindqvist, Y.; Schneider, G.; Hellman, U.; Ronne, H. Yeast TKL1 gene encodes a Transketolase that is required for efficient glycolysis and biosynthesis of aromatic amino acids. *J. Biol. Chem.* **1993**, *268*, 24346-24352.
27. Soh, Y.; Song, B. J.; Jeng, J.; Kallarakal, A. T. Critical role of arg433 in rat Transketolase activity as probed by site-directed mutagenesis. *Biochem. J.* **1998**, *333*, 367-372.
28. Wikner, C.; Nilsson, U.; Meshalkina, L.; Udekwu, C.; Lindqvist, Y.; Schneider, G. Identification of catalytically important residues in yeast Transketolase. *Biochem.* **1997**, *36*, 15643-15649.
29. Isupov, M. N.; Rupprecht, M. P.; Wilson, K. S.; Dauter, Z.; Littlechild, J. A. Crystal Structure of Escherichia coli Transketolase. To be Published.
30. Gerhardt, S.; Echt, S.; Busch, M.; Freigang, J.; Auerbach, G.; Bader, G.; Martin, W. F.; Bacher, A.; Huber, R.; Fischer, M. Structure and properties of an engineered Transketolase from maize. *Plant. Physiol.* **2003**, *132*, 1941-1949.
31. Boren, J.; Ramos-Montoya, A.; De Atauri, P.; Comin-Anduix, B.; Cortes, A.; Centelles, J. J.; Frederiks, W. M.; Van Noorden, C. J. F.; Cascante M. Metabolic Control Analysis Aimed at the Ribose Synthesis Pathways of Tumor Cells: A New Strategy for Antitumor Drug Development. *Mol. Biol. Rep.* **2002**, *29*, 7-12.
32. Babiak, R. M.; Campello, A.P.; Carnieri, E. G.; Oliveira, M. B. Methotrexate: pentose cycle and oxidative stress. *Cell Biochem. Funct.* **1998**, *4*, 283-93.

33. Holmberg, N.; Ryde, U.; Bulow, L. Redesign of the coenzyme specificity in L-lactate dehydrogenase from *Bacillus stearothermophilus* using site-directed mutagenesis and media engineering. *Prot. Engin.* **1999**, *12*, 851-856.

34. Zutshi, R.; Brickner, M.; Chmielewski, J. Inhibiting the assembly of protein-protein interfaces. *Curr. Opin. Chem. Biol.* **1998**, *2*, 62-66.