



UNIVERSIDAD DE MURCIA
FACULTAD DE INFORMÁTICA

Evaluación de Plataformas de Alto Rendimiento
para el Descubrimiento de Fármacos

D. Ginés David Guerrero Hernández

2014



UNIVERSIDAD DE MURCIA

Facultad de Informática

EVALUATION OF HIGH PERFORMANCE COMPUTING
PLATFORMS FOR DRUG DISCOVERY

EVALUACIÓN DE PLATAFORMAS DE ALTO RENDIMIENTO
PARA EL DESCUBRIMIENTO DE FÁRMACOS

TESIS DOCTORAL

Presentada por:

D. Ginés David Guerrero Hernández

Supervisada por:

Dr. D. José María Cecilia Canales

Dr. D. José Manuel García Carrasco

Murcia, Abril de 2014

*A mi compa#era de vida,
mi mujer.*

Agradecimientos

Después de todo el trabajo realizado durante estos últimos años al fin el resultado queda materializado. Echando la mirada atrás, es cuándo se hace notorio que este trabajo jamás hubiera sido posible sin la ayuda, apoyo y participación de personas e instituciones que han permitido que este esfuerzo llegue a un feliz término. Por ello, es para mi un verdadero placer utilizar este espacio para ser justo y consecuente con ellas, expresándole mis agradecimientos.

Debo de agradecer de manera especial y sincera a Chema, uno de mis directores de tesis, su importante aporte y participación activa en el desarrollo de esta tesis. Debo de destacar sobre todo su capacidad de guiar mis ideas, disponibilidad y paciencia, sin las cuales esta tesis habría sido completamente inviable.

Quiero expresar también mi más sincero agradecimiento a mi otro director, José Manuel, por aceptarme para realizar la tesis bajo su dirección. Su apoyo y confianza en mi trabajo, no solamente por el desarrollo de esta tesis, sino también en mi formación como investigador. Le agradezco también haberme facilitado siempre los medios suficientes para llevar a cabo todas las actividades propuestas durante el desarrollo de esta tesis.

Quiero extender un sincero agradecimiento a Horacio, que no sólo ha sido un colaborador en los artículos, sino que ha contribuido y participado en cada una de las ideas que han ido surgiendo y madurando a lo largo de la tesis.

A Manolo, por trasmitirme su perseverancia, tras el duro trabajo realizado por su parte en muchos de los artículos en los que hemos colaborado. Y como no, por brindarme la gran oportunidad de un nuevo futuro que recién está comenzando.

Gracias a todos mis compañeros, y sobre todo amigos, de DITEC: Chema, Toni, Dani, Alberto, Ricardo, Ana, Dani, Antonio, Epi, José Luis, Rubén y, especialmente, a Juanma que no sólo ha sido compañero, sino también colaborador y fuente de ideas.

También quiero agradecer a mis compañeros de carrera, por todos esos momentos que hemos vivido a lo largo de esta época y los que quedan por vivir, entre otros: Eneko, Dani, Juanlu, Pablo, Espinazo, Ángel, Pedro, Nacho, Lucía, María, Conte y Sergio.

A mis amigos de Torre Pacheco, gracias a los cuales he podido desconectar del trabajo y de la tesis tantos fines de semana: Bumi, Juan Carlos, Miguel, Tomás, Trombo, Largo, Carlos, y respectivas. También a Juanfra, Frasco y respectivas, que aunque ya no están en el pueblo, siempre habrá una cercanía muy especial con ellos.

Y, por supuesto, el agradecimiento más profundo y sentido va para mi familia.

Sin su apoyo, colaboración e inspiración habría sido imposible llevar a cabo esta tesis. A mi madre, Rosi, por estar siempre ahí e inculcarme los valores correctos. A mis hermanos, Mara y Evaristo, por todos los momentos que hemos pasado, especialmente los momentos tan duros de los últimos meses, que han hecho que estemos aún más unidos. A mis abuelos, Ginés y Rosa, por ser un ejemplo de personas a seguir, donde su humildad y bondad son la mejor herencia que podrían haberme legado.

Por último y más importante, a mi compañera de vida, mi mujer, Upe. Ella ha sido mi compañera inseparable, mi pilar de apoyo en este arduo viaje. Mil gracias por acompañarme en este proceso, por sobre todo, tu amor, tu comprensión y paciencia, que permitieron que pudiese, no sólo trabajar, sino también llegar a buen puerto. Y, una vez más, gracias por emprender junto a mi la aventura de mi vida, especialmente por permanecer a mi lado, apoyarme y darme fuerzas para afrontar los grandes cambios que están aconteciendo en nuestras vidas.

Por supuesto, agradecer a las entidades que han financiado este trabajo. Sin su apoyo a la investigación, esta tesis no habría sido posible. Quiero agradecer la financiación de la Fundación Séneca (Agencia Regional de Ciencia y Tecnología, Región de Murcia), dentro de los proyectos 00001/CS/2007 y 15290/PI/2010; así como al Ministerio de Educación y Ciencia bajo los proyectos TIN2009-14475-C04 y TIN2012-31345. Además, esta investigación fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02).

Gracias.

Índice

1	Resumen	xi
1.1	Motivación y Objetivos	xi
1.2	Metodología	xiv
1.3	Resultados	xv
1.4	Conclusiones y Trabajo Futuro	xvi
2	Publicaciones que Componen esta Tesis Doctoral	1
2.1	<i>Accelerated Conformational Entropy Calculations Using Graphic Processing Units</i>	2
2.2	<i>Toward Energy Efficiency in Heterogeneous Processors: Findings on Virtual Screening Methods</i>	4
2.3	<i>A Performance/Cost Model for a CUDA Drug Discovery Application on Physical and Public Cloud Infrastructures</i>	6
3	Relevancia de la Tesis Doctoral	9
4	Bibliografía	13
4.1	Referencias	13
4.2	Publicaciones	15

Capítulo 1

Resumen

1.1 Motivación y Objetivos

La integración de los últimos avances de investigación en los campos de la biología, química, física, matemáticas, medicina e informática están permitiendo importantes avances en las áreas de la atención sanitaria, el descubrimiento de nuevos fármacos y la investigación genómica, entre otras. Estos avances están brindando nuevas estrategias terapéuticas para vivir más tiempo y mejor, ofreciendo estilos de vida más saludables que no eran imaginables hace tan solo unos años atrás. La unión de estos esfuerzos ha dado como fruto un nuevo campo de investigación multidisciplinar denominado **Bioinformática** [12], que en líneas generales se puede definir como el uso de los últimos avances en informática para resolver los retos planteados en los campos de la biología y la medicina.

Las aplicaciones bioinformáticas destacan por la necesidad de procesar ingentes bases de datos [4]. Por ejemplo, los métodos de cribado virtual (**Virtual Screening**, VS), utilizados en el proceso de descubrimiento de nuevos fármacos [22], permiten analizar de manera virtual la efectividad de la interacción de un gran número de compuestos químicos contra una proteína que, potencialmente, participa en una enfermedad concreta. De esta manera, un gran número de hipótesis pueden ser analizadas antes de los ensayos clínicos; ofreciendo, entre otros aspectos, abaratar costes y reducir el tiempo de este importante proceso.

Sin embargo, los métodos actuales de VS están limitados por la capacidad computacional necesaria para analizar grandes cantidades de datos de manera precisa y rápida, que los hagan aun más atractivos al mundo científico [14]. El proceso requiere de tiempos de respuestas inmediatos, que permitan tomar decisiones en base a las hipótesis planteadas. Este es el marco de trabajo de esta Tesis Doctoral; analizar las últimas tendencias en computación de altas prestaciones (**High Performance Computing**, HPC), evaluando sus principales ventajas y desventajas para permitir avanzar en el desarrollo de métodos de VS eficientes desde el punto de vista computacional, energético y económico.

Los últimos desarrollos en el campo de la computación de altas prestaciones vienen

marcados por una clara **consolidación del paralelismo** como alternativa para afrontar los nuevos retos computacionales, principalmente impuestos por las limitaciones físicas del silicio [7]. Además, la evolución al paralelismo llega a su exponente máximo con las arquitecturas masivamente paralelas, que incluyen miles de *cores* simples dentro del chip [5] con el fin de incrementar el rendimiento de las aplicaciones; es por ello que estas plataformas también son conocidas como arquitecturas orientadas al rendimiento (*throughput*) [9].

A la vanguardia de las arquitecturas masivamente paralelas se sitúan las Unidades de Procesamiento Gráfico (*Graphics Processing Units, GPUs*). Estas arquitecturas paralelas están proporcionando unas aceleraciones en diversos tipos de aplicaciones de aproximadamente un orden de magnitud en comparación con las versiones secuenciales de las mismas, ejecutadas en plataformas tradicionales como las CPUs, superando estas magnitudes en algunos casos [21]. Esto brinda una oportunidad única para incrementar la capacidad computacional de los ordenadores tradicionales, permitiendo así tener pequeños supercomputadores en casa a precios muy asequibles.

Este marco computacional no sólo conlleva un cambio en la arquitectura hardware, sino también en los modelos de programación subyacentes que permiten explotar estas arquitecturas. En este sentido, la comunidad científica cada vez es más consciente de la necesidad de aprender programación paralela. Los **modelos de programación paralela** clásicos, como OpenMP [6] y MPI [23], son todavía válidos en ciertos entornos. Sin embargo, otros modelos han aparecido con fuerza en los últimos años, como es el caso de CUDA [17], OpenCL [24] y OpenACC [26]. Estos modelos facilitan la programación de los dispositivos masivamente paralelos, haciendo muy asequible su programación y, por tanto, su aplicabilidad en distintos dominios computacionales. El principal problema de este paradigma de programación es la curva de aprendizaje, que en muchas ocasiones obliga a la redefinición e incluso al rediseño de los algoritmos para acceder a todo el potencial presente en estas arquitecturas.

Además, las arquitecturas masivamente paralelas se están incorporando en los ordenadores de alto rendimiento. Desde Junio del 2011, cuando el supercomputador más rápido del mundo, el Tianhe-1A, situado en el Centro Nacional de Supercomputación de Tianjin (China)¹, albergaba 7.168 GPUs Nvidia Tesla M2050, muchos centros de supercomputación han seguido su tendencia. En la actualidad el ordenador más potente del mundo, el Tianhe-2, situado en el Centro Nacional de Supercomputación de Guangzhou (China), incluye procesadores masivamente paralelos Intel Xeon Phi [13].

La inclusión de estos dispositivos aceleradores en las infraestructuras de supercómputo provoca un gran impacto en el consumo energético de un nodo en un cluster, pudiendo verse incrementado en hasta un 30 %, lo cual es una gran problemática en la actualidad. Esta es una preocupación crítica, especialmente para grandes centros de supercomputación, donde los costes relativos al suministro eléctrico representan una parte importante del Coste Total de Posesión (*Total Cost Ownership, TCO*) [8]. La comunidad científica también es consciente de esto y está haciendo grandes esfuerzos para la creación de instalaciones energéticamente más eficientes. Por este motivo se

¹<http://www.top500.org/>

elabora la lista Green500², donde se muestran los 500 superordenadores más energéticamente eficientes del mundo. De este modo, podemos ver un claro cambio desde la tendencia tradicional de la métrica FLOPS hacia los FLOPS por Vatio.

Las técnicas de virtualización permiten un ahorro considerable de energía, ya que gracias a ellas se posibilita el uso compartido de hardware entre diversos usuarios, reduciendo la cantidad de unidades requeridas de un determinado dispositivo. Como consecuencia, la virtualización se está arraigando en los centros de supercomputación. En particular, la **computación en la nube** es una técnica de virtualización que por naturaleza es energéticamente eficiente [11], en la cual los servicios se ejecutan remotamente en una nube de computación ubicua que provee recursos escalables y virtualizados. Así pues, en los picos de carga máxima los recursos acaparados por los usuarios pueden ser trasladados a otras partes de la nube, de tal modo que la agregación de los recursos de la nube puede proveer una mayor utilización hardware [3]. Además, los proveedores de las redes públicas de computación en la nube ofrecen servicios que siguen un modelo de pago según uso [18], lo cual hace que este tipo de infraestructuras sean una interesante alternativa a las infraestructuras locales.

Por otro lado, la computación voluntaria está comenzando a ser cada vez más popular para la realización de computación masiva de datos [2]. Esta nueva tendencia se presenta como una alternativa válida y gratuita para la computación de aplicaciones que demandan un gran poder computacional. Sin embargo, la **computación voluntaria** sólo es factible para aplicaciones específicas donde no haya comunicación de datos entre distintos procesos y los ficheros de salida no sean de gran tamaño, ya que estos han de ser enviados a través de Internet.

Teniendo en cuenta lo anteriormente expuesto y que el objetivo global de la tesis es **la evaluación de plataformas de alto rendimiento que permitan la ejecución eficiente de simulaciones bioinformáticas, tales como VS, en términos de rendimiento, eficiencia energética y coste**; podemos enumerar los siguientes objetivos específicos que hemos perseguido para la consecución de tal fin:

Objetivo 1 Hacer un estudio del estado del arte de la aplicación del HPC a los métodos de VS.

Objetivo 2 Analizar, rediseñar e implementar métodos de VS en CUDA.

Objetivo 3 Comparar la implementación en GPU con otras alternativas HPC.

Objetivo 4 Analizar y comparar el consumo energético en distintas plataformas de HPC.

Objetivo 5 Realizar pruebas con la plataforma mejor cualificada para la ejecución de simulaciones de VS en plataformas de computación en la nube.

Objetivo 6 Realizar pruebas con la plataforma mejor cualificada para la ejecución de simulaciones de VS en plataformas de computación voluntaria.

²<http://www.green500.org/>

1.2 Metodología

Esta Tesis Doctoral fue concebida desde el principio como una tesis basada en compendio de artículos, por lo tanto todo el trabajo realizado a lo largo de la tesis ha sido orientado en esta dirección. No obstante, fruto de la investigación realizada en esta tesis han surgido gran cantidad de colaboraciones y proyectos que han dado como resultado numerosas publicaciones que no están detalladas en este documento, pero que han enriquecido la formación de doctorando. Estas publicaciones están enumeradas en la Sección 4.2.

Tras realizar un profundo y exhaustivo estudio del estado del arte planteado en el **Objetivo 1**, se observó la aceleración obtenida por la paralelización de uno de los núcleos computacionales de la simulación de VS, y en concreto de los métodos de *docking* [10], encargado del **cálculo del potencial electrostático**. Este algoritmo [19] conseguía un ratio de aceleración cercano a las 150 veces más rápida que su versión secuencial (150x de *speed-up*), utilizando para ello la arquitectura Cell de IBM [20]. Aunque queda en entredicho que las plataformas masivamente paralelas sean capaces de conseguir estas diferencias con las plataformas modestamente paralelas [16], es cierto que este trabajo identifica a los métodos de *docking* como aplicaciones con unas grandes necesidades computacionales, y que además cumplen con un patrón de cómputo altamente paralelizable. Por tanto, éste fue nuestro punto de partida.

Después de analizar el patrón de cómputo presente en el cálculo del potencial electrostático, pasamos a su diseño e implementación en la GPU, ejecutando así el **Objetivo 2**. Además del cálculo del potencial electrostático, y viendo el claro beneficio que los cálculos *docking* podían adquirir en este tipo de plataformas, decidimos abordar la paralelización de otro de los núcleos computacionales presentes en estas simulaciones; esto es, el cálculo del área de la superficie accesible (*Solvent-Accessible Surface Area*, **SASA**) [15], el cual se usa para calcular el área de la superficie de una biomolécula que es accesible a un disolvente. El patrón cómputo en SASA era, a priori, inherentemente secuencial, y por tanto tuvimos que crear un diseño masivamente paralelo replanteando el problema desde su definición.

Una vez planteados los patrones de cómputo objeto de estudio de esta tesis, decidimos analizar el actual panorama de computación de altas prestaciones, para así poder cubrir el **Objetivo 3**. Para ello evaluamos el patrón de cómputo definido en el cálculo del potencial electrostático en distintas plataformas de HPC: *i*) una máquina de memoria compartida y *ii*) una máquina de memoria distribuida. Con tal propósito se realizaron varias implementaciones del cálculo del potencial electrostático en OpenMP y MPI, culminando con una implementación híbrida de MPI-OpenMP con vectorización y comunicación asíncrona que permite el solapamiento de las transferencias y el cómputo.

Una de las mayores preocupaciones en la actualidad de la comunidad de HPC es el consumo energético de las plataformas. Por tanto, desde nuestro punto de vista era fundamental incluir, además de un análisis de rendimiento, un estudio del comportamiento energético de las distintas plataformas y lenguajes de programación al ejecutar aplicaciones tan exigentes como las abordadas en este Tesis Doctoral. Para realizar los

experimentos usamos como banco de pruebas cinco GPUs de Nvidia, una GPU de ATI, un procesador de bajo consumo de Intel, dos procesadores embebidos de Samsung y un procesador de alto rendimiento de AMD. Con el análisis de estos resultados se dio por cumplido el **Objetivo 4**. Cabe destacar que los núcleos computacionales tratados en esta tesis fueron recodificados en OpenCL, para así poder hacer un estudio comparativo del comportamiento de éste frente a su versión OpenMP vectorizada en las plataformas *multicores* y su versión CUDA en las plataformas *many-cores*.

Una vez comprobado los sobresalientes resultados de la implementación CUDA para las GPUs de Nvidia en cuanto a rendimiento y a eficiencia energética, se optó por usar esta plataforma para codificar **BINDSURF** [25], una herramienta de VS que integra, entre otros, los núcleos computacionales estudiados en esta tesis. BINDSURF es un método de *docking* completo que, desde su génesis, tiene el paralelismo como su principal estandarte para hacer un barrido completo de la simulación de la integración entre un proteína y un ligando. Para acercarnos a simulaciones válidas en el entorno científico, BINDSURF necesita ser ejecutado en plataformas de gran escala que faciliten el análisis de grandes bases de datos de compuestos en un tiempo factible. En la actualidad existen diversos tipos de organizaciones, modelos y esquemas para la unión de las capacidades computacionales de diversos ordenadores; entre ellas destacamos los supercomputadores, la computación en la nube y la computación voluntaria. En esta tesis estudiamos las ventajas y desventajas de estas alternativas en términos de rendimiento, eficiencia energética y costes económicos. Este estudio representa los **Objetivos 5 y 6** de la tesis.

1.3 Resultados

En esta sección se muestran los resultados obtenidos durante la elaboración de esta Tesis Doctoral que están directamente relacionados con su propósito, omitiéndose aquellos han sido elaborados de forma transversal.

Así pues, en el trabajo ‘*Effective Parallelization of Non-bonded Interactions Kernel for Virtual Screening on GPUs*’ [49], publicado en las actas del “5th International Conference on Practical Applications of Computational Biology & Bioinformatics” se muestra el resultado de la implementación del cálculo del potencial electrostático en la GPU, obteniendo un *speed-up* de 260x frente a su versión secuencial.

Como ya se comentó en la Sección 1.2, además del desarrollo de este cálculo, se realizó la implementación del cálculo de SASA en GPU, dando como resultado **MURCIA** (*Molecular Unburied Rapid Calculation of Individual Areas*). El cual fue usado para calcular la contribución entrópica a la energía libre de las interacciones ligando-proteína, obteniendo una aceleración de dos órdenes de magnitud en el estudio titulado ‘*Accelerated Conformational Entropy Calculations Using Graphic Processing Units*’ [53], publicado en “*Journal of Chemical Information and Modeling*”.

Teniendo en cuenta el alto rendimiento obtenido por la GPU en el cálculo del potencial electrostático, se decidió comparar ésta contra dos infraestructuras de un centro de supercomputación, las cuales tienen un precio muy superior. Los resultados

fueron sorprendentes, ya que la aceleración conseguida en la GPU superaba con creces los resultados obtenidos en la máquina de memoria compartida e igualaba a los resultados obtenidos con 128 *cores* de la máquina de memoria distribuida. Es importante destacar que para hacer la comparación lo más equitativa posible se optimizaron los códigos que se ejecutaron en el centro de supercomputación, tal como se presentó en “*20th Euromicro International Conference on Parallel, Distributed and Network-Based Processing*” con el trabajo titulado ‘*Parallelization of Virtual Screening in Drug Discovery on Massively Parallel Architectures*’ [50].

Con el objetivo de ampliar el análisis comparativo de la GPU frente a otras arquitecturas, se publicó un estudio en el que se analiza el rendimiento y el consumo energético del cálculo del potencial electrostático en diversas plataformas. Este estudio, titulado ‘*Energy and performance evaluation of Virtual Screening methods on different parallel computing systems*’ [46], fue publicado en “*Concurrency and Computation: Practice and Experience*”. En este artículo se mostraba la GPU como clara vencedora frente al resto de arquitecturas en cuanto a rendimiento y consumo energético.

Posteriormente, se desarrolló un modelo de rendimiento/coste para la ejecución de la aplicación BINDSURF, que permite comprobar si es más adecuado ejecutar las simulaciones en una plataforma local o usando la computación en la nube, en términos de rendimiento y coste. Este trabajo se publicó en “*Concurrency and Computation: Practice and Experience*” con el título “*A performance/cost model for a CUDA drug discovery application on physical and public cloud infrastructures*’ [52].

El último artículo fruto de esta tesis ha sido recientemente enviado a la revista “*High-Performance Computing and Big Data in Omics-based Medicine*” para su evaluación. En este artículo se muestra la plataforma de computación voluntaria BOINC [1] como una alternativa eficiente para la ejecución de la aplicación BINDSURF [48]; y además muy atractiva, ya que tiene un coste virtualmente gratuito. Hay que destacar que este trabajo es público a través del proyecto ibercivis³.

1.4 Conclusiones y Trabajo Futuro

En la primera década del siglo XXI, la ley de Moore, que ha guiado el desarrollo de procesadores en los últimos cincuenta años, fue puesta en entredicho por la comunidad científica. Esto fue debido principalmente a las limitaciones físicas del silicio que provocaron un cambio en la tendencia en el desarrollo de procesadores, tomando como principal estandarte para este cambio el paralelismo. Esta transición ha situado a la programación (masivamente) paralela como la única manera de extraer el máximo rendimiento de las nuevas plataformas de consumo; siendo esto fundamental para abordar los retos científicos de la actualidad. Desafortunadamente, estos retos plantean problemas cuyas necesidades computacionales están fuera del alcance de una sola máquina. Simulaciones como las tratadas en esta Tesis Doctoral, necesitan escalar a grandes centros de cálculo, cuyos costes, sólo están al alcance de grandes instituciones y gobiernos. Sin embargo, la actual situación socio-económica obliga a la utilización

³<http://www.ibercivis.es/>

eficiente de los recursos. Herramientas como la computación en la nube o voluntaria ofrecen una alternativa para aprovechar al máximo los recursos computacionales de una manera flexible, rápida, económica y respetuosa con el medio ambiente.

En esta Tesis Doctoral evaluamos el actual paradigma de computación descrito anteriormente, utilizando como caso de estudio un problema de alta repercusión en la sociedad como es la simulación de cribado virtual, utilizada para el descubrimiento de nuevos fármacos. El estudio ha recorrido los distintos niveles de procesamiento; partiendo por un análisis exhaustivo de las diferentes alternativas a nivel de chip disponibles en el mercado, pasando por la evaluación de los mismos en un entorno clúster, hasta escalar a niveles de computación en la nube y computación voluntaria.

Como conclusión de este estudio podemos afirmar que las GPUs están a la vanguardia del desarrollo de aplicaciones científicas con patrones de cómputo masivamente paralelos y altas demandas computacionales, como es el cribado virtual. Además, nos atrevemos a afirmar, con los números mostrados en esta tesis doctoral y los resultados de estas plataformas en los últimos años, que se debería valorar su uso en la gran mayoría de campos científicos que demanden gran capacidad de cómputo. Es posible, sin embargo, que la migración de estas plataformas impliquen replantear el problema desde su origen, pero sin lugar a dudas, esto forma parte del pensamiento computacional, imprescindible en el desarrollo de aplicaciones científica en el estado actual de la computación de altas prestaciones.

Por otro lado, si las ejecuciones que se quieren realizar son a mayor escala, se han de valorar como posibles alternativas al uso de un clúster de cómputo, la utilización de computación en la nube y la computación voluntaria. La computación en la nube puede ser una opción muy interesante si el cómputo que se desea realizar se desarrolla en periodos de tiempo intermitentes, ya que la no utilización de los recursos locales implica que la inversión económica no quede justificada. Por otro lado, la opción de usar una plataforma de computación voluntaria es muy atractiva, siempre y cuando, la aplicación que se está paralelizando admita ser portada a una plataforma de este tipo; pudiendo conseguir computación de altas prestaciones a un coste cero.

Con respecto al trabajo futuro, a corto plazo, se pretende probar como esta metodología de estudio y los resultados obtenidos en ella se pueden aplicar a otras aplicaciones de Biología Computacional y de Bioinformática con similares resultados. Además, habría que contemplar en éste y los próximos estudios otras nuevas arquitecturas computacionales, como es el caso del Intel Xeon Phi; y otros lenguajes de programación emergentes, como puede ser OpenACC. Por otro lado, a corto plazo también podrían ser estudiadas las soluciones híbridas de computación en la nube, donde sólo se accedan los recursos extra que provee la nube bajo determinadas condiciones, como puede ser una demanda de recursos que la plataforma local no es capaz de satisfacer por sí misma.

Finalmente, nos gustaría subrayar que esta Tesis Doctoral ha contribuido al desarrollo de una aplicación de VS y que el uso de la misma debería de ayudar a encontrar nuevos candidatos a fármacos de manera eficiente en términos de rendimiento, energía y costes económicos.

Capítulo 2

Publicaciones que Componen esta Tesis Doctoral

2.1 *Accelerated Conformational Entropy Calculations Using Graphic Processing Units*

Título	<i>Accelerated Conformational Entropy Calculations Using Graphic Processing Units</i>
Autores	Q. Zhang, J. Wang, G.D. Guerrero, J.M. Cecilia, J.M. García, Y. Li, H. Pérez-Sánchez y T. Hou
Tipo	Revista
Revista	<i>Journal of Chemical Information and Modeling</i>
Páginas	2057-2064
Año	2013
Mes	Julio
Resumen	<i>Conformational entropy calculation, usually computed by normal-mode analysis (NMA) or quasi harmonic analysis (QHA), is extremely time-consuming. Here, instead of NMA or QHA, a solvent accessible surface area (SASA) based model was employed to compute the conformational entropy, and a new fast GPU-based method called MURCIA (Molecular Unburied Rapid Calculation of Individual Areas) was implemented to accelerate the calculation of SASA for each atom. MURCIA employs two different kernels to determine the neighbors of each atom. The first kernel (K1) uses brute force for the calculation of the neighbors of atoms, while the second one (K2) uses an advanced algorithm involving hardware interpolations via GPU texture memory unit for such purpose. These two kernels yield very similar results. Each kernel has its own advantages depending on the protein size. K1 performs better than K2 when the size is small and vice versa. The algorithm was extensively evaluated for four protein data sets and achieves good results for all of them. This GPU-accelerated version is ~600 times faster than the former sequential algorithm when the number of the atoms in a protein is up to 10⁵.</i>
DOI	http://dx.doi.org/10.1021/ci400263t
URL	http://pubs.acs.org/doi/abs/10.1021/ci400263t?journalCode=jcisd8
Estado	Publicado

2.2 Toward Energy Efficiency in Heterogeneous Processors: Findings on Virtual Screening Methods

Título	<i>Toward Energy Efficiency in Heterogeneous Processors: Findings on Virtual Screening Methods</i>
Autores	G.D. Guerrero, J.M. Cebrián, H. Pérez-Sánchez, J.M. García, M. Ujaldón y J.M. Cecilia
Tipo	Revista
Revista	<i>Concurrency and Computation: Practice and Experience</i>
Año	2013
Mes	Agosto
Resumen	<i>The integration of the latest breakthroughs in computational modeling and high performance computing (HPC) has leveraged advances in the fields of healthcare and drug discovery, among others. By integrating all these developments together, scientists are creating new exciting personal therapeutic strategies for living longer that were unimaginable not that long ago. However, we are witnessing the biggest revolution in HPC in the last decade. Several graphics processing unit architectures have established their niche in the HPC arena but at the expense of an excessive power and heat. A solution for this important problem is based on heterogeneity. In this paper, we analyze power consumption on heterogeneous systems, benchmarking a bioinformatics kernel within the framework of virtual screening methods. Cores and frequencies are tuned to further improve the performance or energy efficiency on those architectures. Our experimental results show that targeted low-cost systems are the lowest power consumption platforms, although the most energy efficient platform and the best suited for performance improvement is the Kepler GK110 graphics processing unit from Nvidia by using compute unified device architecture. Finally, the open computing language version of virtual screening shows a remarkable performance penalty compared with its compute unified device architecture counterpart.</i>
DOI	http://dx.doi.org/10.1002/cpe.3119
URL	http://onlinelibrary.wiley.com/doi/10.1002/cpe.3119/abstract
Estado	Publicado

2.3 A Performance/Cost Model for a CUDA Drug Discovery Application on Physical and Public Cloud Infrastructures

Título	<i>A Performance/Cost Model for a CUDA Drug Discovery Application on Physical and Public Cloud Infrastructures</i>
Autores	G.D. Guerrero, R.M. Wallace, J.L. Vázquez-Poletti, J.M. Cecilia, J.M. García, D. Mozos y H. Pérez-Sánchez
Tipo	Revista
Revista	<i>Concurrency and Computation: Practice and Experience</i>
Año	2013
Mes	Agosto
Resumen	<i>Virtual Screening (VS) methods can considerably aid drug discovery research, predicting how ligands interact with drug targets. BINDSURF is an efficient and fast blind VS methodology for the determination of protein binding sites, depending on the ligand, using the massively parallel architecture of graphics processing units (GPUs) for fast unbiased prescreening of large ligand databases. In this contribution, we provide a performance/cost model for the execution of this application on both local system and public cloud infrastructures. With our model, it is possible to determine which is the best infrastructure to use in terms of execution time and costs for any given problem to be solved by BINDSURF. Conclusions obtained from our study can be extrapolated to other GPU-based VS methodologies.</i>
DOI	http://dx.doi.org/10.1002/cpe.3117
URL	http://onlinelibrary.wiley.com/doi/10.1002/cpe.3117/abstract
Estado	Publicado

Capítulo 3

Relevancia de la Tesis Doctoral

Accelerated Conformational Entropy Calculations Using Graphic Processing Units

El artículo titulado ‘*Accelerated Conformational Entropy Calculations Using Graphic Processing Units*’, ha sido publicado en la revista “*Journal of Chemical Information and Modeling*”, cuya relevancia y factor de impacto pueden ser observados a continuación:

ISI Web of KnowledgeSM

Journal Citation Reports[®]



2012 JCR Science Edition

Rank in Category: Journal of Chemical Information and Modeling

Journal Ranking ⁱ

For **2012**, the journal **Journal of Chemical Information and Modeling** has an Impact Factor of **4.304**.

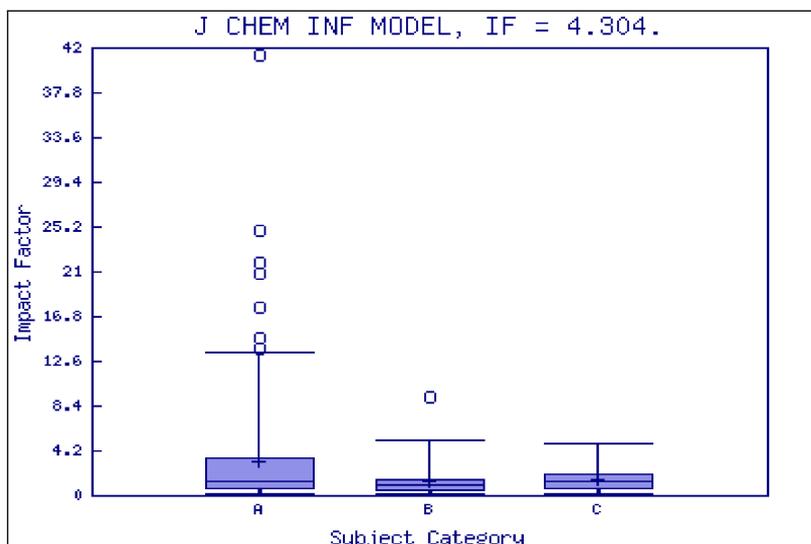
This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
CHEMISTRY, MULTIDISCIPLINARY	152	30	Q1
COMPUTER SCIENCE, INFORMATION SYSTEMS	132	4	Q1
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	100	3	Q1

Category Box Plot ⁱ

For **2012**, the journal **Journal of Chemical Information and Modeling** has an Impact Factor of **4.304**.

This is a box plot of the subject category or categories to which the journal has been assigned. It provides information about the distribution of journals based on Impact Factor values. It shows median, 25th and 75th percentiles, and the extreme values of the distribution.



Toward Energy Efficiency in Heterogeneous Processors: Findings on Virtual Screening Methods

El artículo titulado ‘*Toward Energy Efficiency in Heterogeneous Processors: Findings on Virtual Screening Methods*’, ha sido publicado en la revista “*Concurrency and Computation: Practice and Experience*”, cuya relevancia y factor de impacto pueden ser observados a continuación:

ISI Web of KnowledgeSM

Journal Citation Reports[®]



2012 JCR Science Edition

Rank in Category: **CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE**

Journal Ranking ⓘ

For **2012**, the journal **CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE** has an Impact Factor of **0.845**.

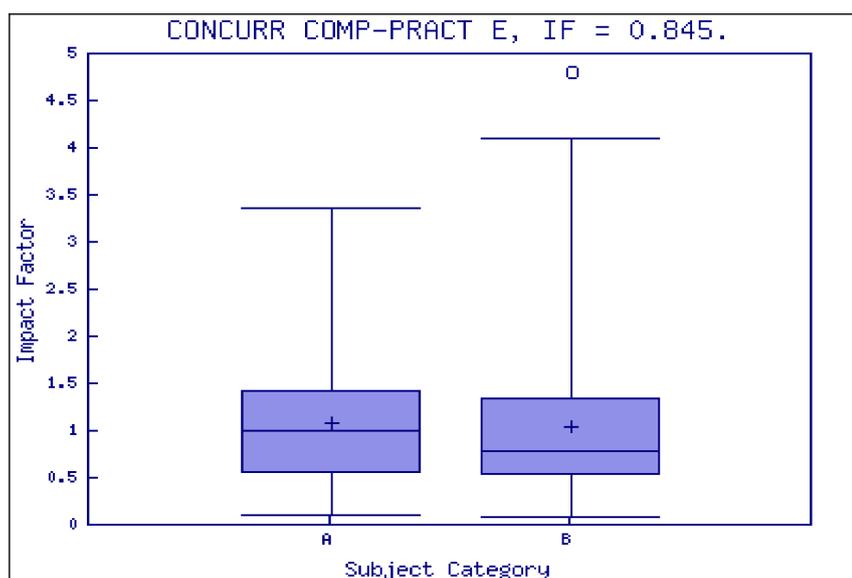
This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
COMPUTER SCIENCE, SOFTWARE ENGINEERING	105	63	Q3
COMPUTER SCIENCE, THEORY & METHODS	100	42	Q2

Category Box Plot ⓘ

For **2012**, the journal **CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE** has an Impact Factor of **0.845**.

This is a box plot of the subject category or categories to which the journal has been assigned. It provides information about the distribution of journals based on Impact Factor values. It shows median, 25th and 75th percentiles, and the extreme values of the distribution.



A Performance/Cost Model for a CUDA Drug Discovery Application on Physical and Public Cloud Infrastructures

El artículo titulado ‘*A Performance/Cost Model for a CUDA Drug Discovery Application on Physical and Public Cloud Infrastructures*’, ha sido publicado en la revista “*Concurrency and Computation: Practice and Experience*”, cuya relevancia y factor de impacto pueden ser observados a continuación:

ISI Web of KnowledgeSM

Journal Citation Reports[®]



2012 JCR Science Edition

Rank in Category: CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE

Journal Ranking ⁱ

For **2012**, the journal **CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE** has an Impact Factor of **0.845**.

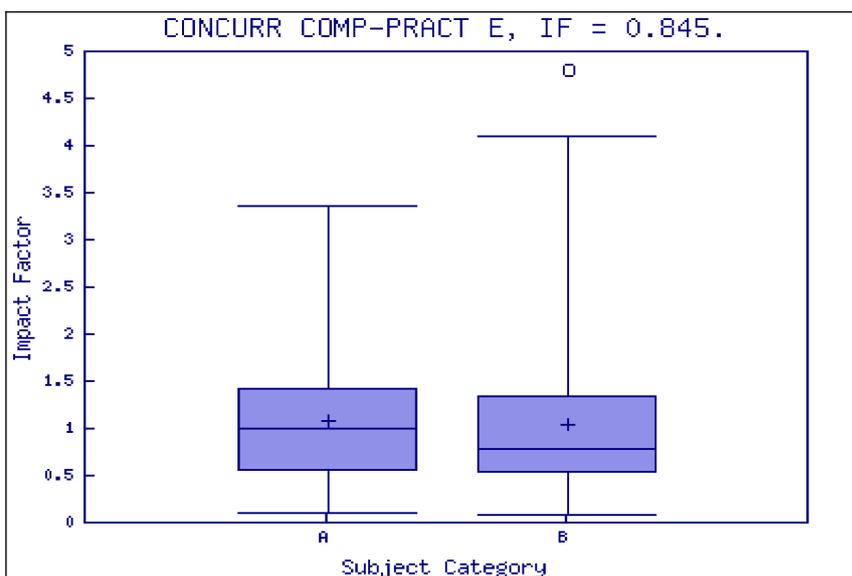
This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
COMPUTER SCIENCE, SOFTWARE ENGINEERING	105	63	Q3
COMPUTER SCIENCE, THEORY & METHODS	100	42	Q2

Category Box Plot ⁱ

For **2012**, the journal **CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE** has an Impact Factor of **0.845**.

This is a box plot of the subject category or categories to which the journal has been assigned. It provides information about the distribution of journals based on Impact Factor values. It shows median, 25th and 75th percentiles, and the extreme values of the distribution.



Capítulo 4

Bibliografía

4.1 Referencias

- [1] David P. Anderson. BOINC: A System for Public-Resource Computing and Storage. In *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing (GRID 2004)*, pages 4–10, 2004.
- [2] David P Anderson and Gilles Fedak. The Computational and Storage Potential of Volunteer Computing. In *Cluster Computing and the Grid (CCGRID 2006)*, volume 1, pages 73–80, 2006.
- [3] Andreas Berl, Erol Gelenbe, Marco Di Girolamo, Giovanni Giuliani, Hermann De Meer, Minh Quan Dang, and Kostas Pentikousis. Energy-Efficient Cloud Computing. *The Computer Journal*, 53(7):1045–1051, 2010.
- [4] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, January 2000.
- [5] Shekhar Borkar. Thousand Core Chips: A Technology Perspective. In *Proceedings of the 44th annual Design Automation Conference*, pages 746–749, 2007.
- [6] Leonardo Dagum and Ramesh Menon. OpenMP: An Industry Standard API for Shared-Memory Programming. *Computational Science & Engineering*, 5(1):46–55, 1998.
- [7] Hadi Esmaeilzadeh, Emily Blem, Renee St Amant, Karthikeyan Sankaralingam, and Doug Burger. Dark Silicon and The End of Multicore Scaling. In *Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA 2011)*, pages 365–376, 2011.
- [8] Xiaobo Fan, Wolf-Dietrich Weber, and Luiz Andre Barroso. Power Provisioning for a Warehouse-Sized Computer. In *Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA 2007)*, pages 13–23, 2007.

-
- [9] Michael Garland and David B Kirk. Understanding Throughput-Oriented Architectures. *Communications of the ACM*, 53(11):58–66, 2010.
- [10] David S. Goodsell and Arthur J. Olson. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Structure, Function, and Bioinformatics*, 8(3):195–202, 1990.
- [11] Carl Hewitt. ORGs for Scalable, Robust, Privacy-Friendly Client Cloud Computing. *IEEE Internet Computing*, 12(5):96–99, 2008.
- [12] C. Hodgman, A. French, and D. Westhead. *BIOS Instant Notes in Bioinformatics*. Instant Notes. Taylor & Francis, 2009.
- [13] James Jeffers and James Reinders. *Intel Xeon Phi Coprocessor High Performance Programming*. Newnes, 2013.
- [14] Gerhard Klebe. Virtual Ligand Screening: Strategies, Perspectives and Limitations. *Drug discovery today*, 11(13):580–594, 2006.
- [15] Konstantin V. Klenin, Frank Tristram, Timo Strunk, and Wolfgang Wenzel. Derivatives of Molecular Surface Area and Volume: Simple and Exact Analytical Formulas. *Journal of Computational Chemistry*, 32(12):2647–2653, 2011.
- [16] Victor W. Lee, Changkyu Kim, Jatin Chhugani, Michael Deisher, Daehyun Kim, Anthony D. Nguyen, Nadathur Satish, Mikhail Smelyanskiy, Srinivas Chennupaty, Per Hammarlund, Ronak Singhal, and Pradeep Dubey. Debunking the 100X GPU vs. CPU Myth: An Evaluation of Throughput Computing on CPU and GPU. In *Proceedings of the 37th Annual International Symposium on Computer Architecture (ISCA 2010)*, pages 451–460, 2010.
- [17] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable Parallel Programming with CUDA. *Queue*, 6(2):40–53, 2008.
- [18] Mayur R. Palankar, Adriana Iamnitchi, Matei Ripeanu, and Simson Garfinkel. Amazon S3 for Science Grids: A Viable Solution? In *Proceedings of the 2008 International Workshop on Data-aware Distributed Computing (DADC 2008)*, pages 55–64, New York, NY, USA, 2008.
- [19] Horacio Pérez-Sánchez and Wolfgang Wenzel. Implementation of an Effective Non-Bonded Interactions Kernel for Biomolecular Simulations on the Cell Processor. In *GI Jahrestagung*, pages 721–729, 2009.
- [20] D. Pham, S. Asano, M. Bolliger, M.N. Day, H.P. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, M. Riley, D. Shippy, D. Stasiak, M. Suzuki, M. Wang, J. Warnock, S. Weitzel, D. Wendel, T. Yamazaki, and K. Yazawa. The Design and Implementation of a First-Generation CELL Processor. In *IEEE International Solid-State Circuits Conference (ISSCC 2005)*, volume 1, pages 184–592, February 2005.

-
- [21] Matt Pharr and Randima Fernando. *GPU GEMS 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation*. Addison-Wesley Professional, 2005.
- [22] B. Shoichet. Virtual Screening of Chemical Libraries. *Nature*, 432(7019):862–5, 2004.
- [23] Marc Snir. *MPI – The Complete Reference: The MPI Core*, volume 1. MIT press, 1998.
- [24] John E Stone, David Gohara, and Guochun Shi. OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems. *Computing in science & engineering*, 12(3):66–72, 2010.
- [25] Irene Sánchez-Linares, Horacio Pérez-Sánchez, José M. Cecilia, and José M. García. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinformatics*, 13(14):1–14, 2012.
- [26] Sandra Wienke, Paul Springer, Christian Terboven, and Dieter an Mey. OpenACC – First Experiences with Real-World Applications. In *Euro-Par 2012 Parallel Processing*, pages 859–870. Springer, 2012.

4.2 Publicaciones

- [27] Gregorio Bernabé, **Ginés D. Guerrero**, and Juan Fernández. CUDA and OpenCL implementations of 3D Fast Wavelet Transform. In *IEEE Third Latin American Symposium on Circuits and Systems (LASCAS 2012)*, pages 1–4, Feb 2012.
- [28] Juan M. Cebrian, **Ginés D. Guerrero**, and José M. García. Energy Efficiency Analysis of GPUs. In *High-Performance, Power-Aware Computing (HPPAC 2012)*, pages 1008–1016, Shanghai, China, May 2012.
- [29] José M. Cecilia, José M. García, **Ginés D. Guerrero**, Miguel A. Martínez del Amor, Mario J. Pérez-Jiménez, and Manuel Ujaldón. P Systems Simulations on Massively Parallel Architectures. *Third International Workshop on Parallel Architectures and Bioinspired Algorithms (WPABA 2010)*, pages 17–26, September 2010.
- [30] José M. Cecilia, José M. García, **Ginés D. Guerrero**, Miguel A. Martinez del Amor, Ignacio Perez-Hurtado, and Mario J. Perez-Jimenez. Simulation of P systems with active membranes on CUDA. *Briefings in Bioinformatics*, 11(3):313–322, May 2010.
- [31] José M. Cecilia, José M. García, **Ginés D. Guerrero**, Miguel A. Martínez del Amor, Mario J. Pérez-Jiménez, and Manuel Ujaldón. P systems simulations

- on massively parallel architectures. *Third International Workshop on Parallel Architectures and Bioinspired Algorithms*, pages 17–26, September 2010.
- [32] José M. Cecilia, **Ginés D. Guerrero**, José M. García, Miguel A. Martínez del Amor, Ignacio Pérez-Hurtado, and Mario J. Pérez-Jiménez. A massively parallel framework using P systems and GPUs. In *Proceedings of the Symposium on Application Accelerators in High Performance Computing (SAAHPC 2009)*, Illinois, USA, July 2009.
- [33] José M. Cecilia, **Ginés D. Guerrero**, José M. García, Miguel A. Martínez del Amor, Ignacio Pérez-Hurtado, and Mario J. Pérez-Jiménez. Simulation of P Systems with Active Membranes on CUDA. In *Proceedings of 2009 International Workshop on High Performance Computational Systems Biology (HiBi 2009)*, pages 61–70, October 2009.
- [34] José M. Cecilia, **Ginés D. Guerrero**, José M. García, Miguel A. Martínez del Amor, Mario J. Pérez-Jiménez, and Manuel Ujaldón. Enhancing the Simulation of P Systems for the SAT Problem on GPUs. In *Proceedings of the Symposium on Application Accelerators in High Performance Computing (SAAHPC 2010)*, Knoxville, USA, July 2010.
- [35] José M. Cecilia, José M. García, **Ginés D. Guerrero**, Miguel A. Martínez del Amor, Ignacio Pérez-Hurtado, and Mario J. Pérez-Jiménez. Implementing P systems parallelism by means of GPUs. *Lecture Notes in Computer Science*, 5957:227–241, January 2010.
- [36] José M. Cecilia, José M. García, **Ginés D. Guerrero**, Miguel A. Martínez del Amor, Ignacio Pérez-Hurtado, and Mario J. Pérez-Jiménez. Simulating a P system based efficient solution to SAT by using GPUs. *The Journal of Logic and Algebraic Programming*, 79(6):317–325, 2010.
- [37] José M. Cecilia, José M. García, **Ginés D. Guerrero**, Miguel A. Martínez del Amor, Mario J. Pérez-Jiménez, and Manuel Ujaldón. The GPU on the simulation of cellular computing models. *Soft Computing*, 16(2):231–246, February 2012.
- [38] José M. Cecilia, José M. García, **Ginés D. Guerrero**, and Manuel Ujaldón. Evaluating the SAT Problem on P Systems for Different High-Performance Architectures. *The Journal of Supercomputing*, pages 1–25, 2014.
- [39] Miguel A. Martínez del Amor, Ignacio Pérez-Hurtado, Mario J. Pérez-Jiménez, José M. Cecilia, **Ginés D. Guerrero**, and José M. García. Simulating Active Membrane Systems Using GPUs. *10th Workshop on Membrane Computing*, pages 369–384, August 2009.
- [40] Miguel A. Martínez del Amor, Ignacio Pérez-Hurtado, Mario J. Pérez-Jiménez, José M. Cecilia, **Ginés D. Guerrero**, and José M. García. Simulation of

- Recognizer P Systems by Using manycore GPUs. In *Seventh Brainstorming Week on Membrane Computing*, volume 2, pages 369–384, July 2009.
- [41] Moises Hernández, **Ginés D. Guerrero**, José M. Cecilia, José M. García, Alberto Inuggi, and Stamatis N. Sotiropoulos. Accelerating Fibre Orientation Estimation from Diffusion Weighted Magnetic Resonance Imaging Using GPUs. In *Proceedings of the 20th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP 2012)*, pages 622–626, Munich, Germany, 2012.
- [42] Moisés Hernández, **Ginés D. Guerrero**, José M. Cecilia, José M. García, Alberto Inuggi, Saad Jbabdi, Timothy E. J. Behrens, and Stamatis N. Sotiropoulos. Accelerating Fibre Orientation Estimation from Diffusion Weighted Magnetic Resonance Imaging Using GPUs. *PLoS ONE*, 8(4):e61892, 04 2013.
- [43] Horacio Pérez-Sánchez, **Ginés D. Guerrero**, José M. García, Jorge Pena, José M. Cecilia, Gaspar Cano, Sergio Orts-Escolano, and José García Rodríguez. Improving Drug Discovery using a Neural Networks based Parallel Scoring Function. In *International Joint Conference on Neural Networks (IJCNN 2013)*, pages 1–5, 2013.
- [44] Irene Sánchez-Linares, Horacio Pérez-Sánchez, **Ginés D. Guerrero**, José M. Cecilia, and José M. García. Accelerating multiple target drug screening on GPUs. In *Proceedings of the 9th International Conference on Computational Methods in Systems Biology (CMSB 2011)*, pages 95–102, New York, NY, USA, 2011.
- [45] **Ginés D. Guerrero**, José M. Cecilia, José M. García, and Horacio Pérez-Sánchez. Impact of Implicit Solvation Models on Database Enrichment in GPU based Blind Virtual Screening. In *International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2013)*, pages 731–734, 2013.
- [46] **Ginés D. Guerrero**, Juan M. Cebrián, Horacio Pérez-Sánchez, José M. García, Manuel Ujaldón, and José M. Cecilia. Toward Energy Efficiency in Heterogeneous Processors: Findings on Virtual Screening Methods. *Concurrency and Computation: Practice and Experience*, 2013.
- [47] **Ginés D. Guerrero**, José M. Cecilia, Antonio Llanes, José M. García, Martyn Amos, and Manuel Ujaldón. Comparative Evaluation of Platforms for Parallel Ant Colony Optimization. *The Journal of Supercomputing*, pages 1–12, 2014.
- [48] **Ginés D. Guerrero**, Baldomero Imbernón, Horacio Pérez Sánchez, Francisco Sanz, José M. García, and José M. Cecilia. A Volunteer Computing Platform for the Discovery of Bioactive Compounds: Opportunities and Trade-Offs. *BioMed Research International*, 2014. To be accepted.
- [49] **Ginés D. Guerrero**, Horacio Pérez-Sánchez, Wolfgang Wenzel, José Cecilia, and José García. Effective Parallelization of Non-bonded Interactions Kernel for Virtual Screening on GPUs. In *5th International Conference on Practical Applications*

- of Computational Biology & Bioinformatics (PACBB 2011)*, volume 93, pages 63–69, 2011.
- [50] **Ginés D. Guerrero**, Horacio E. Perez-Sánchez, José M. Cecilia, and José M. García. Parallelization of Virtual Screening in Drug Discovery on Massively Parallel Architectures. In *Proceedings of the 20th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP 2012)*, pages 588–595, Munich, Germany, 2012.
- [51] **Ginés D. Guerrero**, Horacio Pérez Sánchez, Francisco Sanz, and José M. Cecilia. A GPU based Volunteer Computing Platform for the Discovery of Bioactive Compounds. In *Proceedings of International Conference on Computational and Mathematical Methods in Science and Engineering (CMMSE 2013)*, pages 1557–1160, 2013.
- [52] **Ginés D. Guerrero**, Richard M. Wallace, José L. Vázquez-Poletti, José M. Cecilia, José M. García, Daniel Mozos, and Horacio Pérez-Sánchez. A Performance/Cost Model for a CUDA Drug Discovery Application on Physical and Public Cloud Infrastructures. *Concurrency and Computation: Practice and Experience*, 2013.
- [53] Qian Zhang, Junmei Wang, **Ginés D. Guerrero**, José M. Cecilia, José M. García, Youyong Li, Horacio Pérez-Sánchez, and Tingjun Hou. Accelerated Conformational Entropy Calculations Using Graphic Processing Units. *Journal of Chemical Information and Modeling*, 53(8):2057–2064, 2013.

