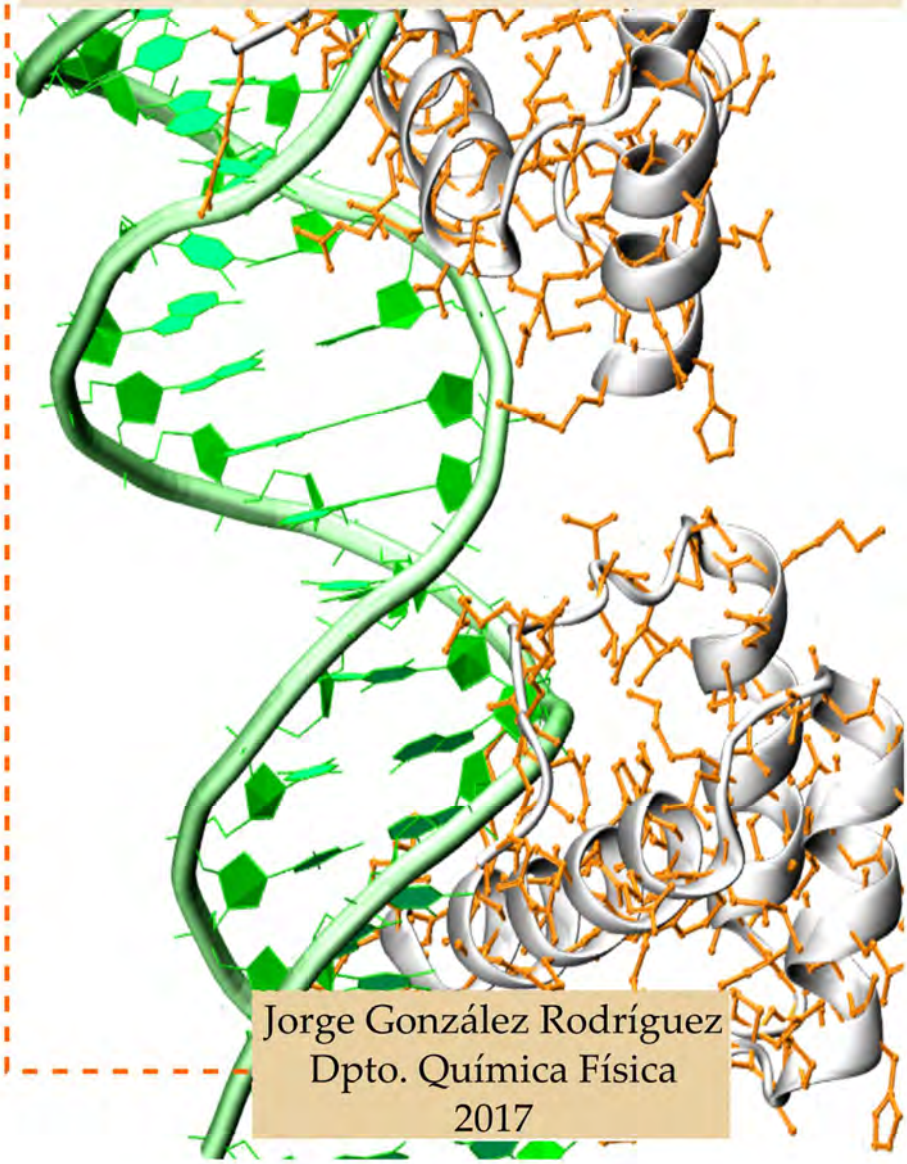
eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
FACULTY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF THE BASQUE COUNTRY

Aggregation of biological building blocks: a theoretical and spectroscopic study



Jorge González Rodríguez
Dpto. Química Física
2017

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
FACULTY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF THE BASQUE COUNTRY

AGGREGATION OF BIOLOGICAL BUILDING BLOCKS: A THEORETICAL AND SPECTROSCOPIC STUDY

By

Jorge González Rodríguez

Leioa, May 2017

*A **Thesis** submitted in Fulfillment of the Requirements for the degree of*

DOCTOR OF PHILOSOPHY

In the Department of Chemical Physics

INDEX

ABSTRACT

DNA and proteins are involved in biological processes which make possible life. On the one hand, the four DNA bases constitute the alphabet of life. Their typical appearance includes stacked Watson-Crick pairs to form DNA strands, reinforced by the existence of the sugar backbone. However, the presence of unusual structures, such as Hoogsteen pairs, triplexes and quadruplexes, raises the question of which are the intrinsic preferences of the DNA aggregation and their role in the appearance of the first DNA strands. On the other hand, proteins are chains of amino acids, whose interactions with DNA strands are the basis of a lot of vital processes, such as the nucleosome formation. Nevertheless, the mechanism which guides the nucleosome formation, based in DNA-histone interactions, is still bewildering, despite the amount of statistical studies of the contacts between DNA and histones. So in order to elucidate some of these questions, we consider three main goals in this thesis:

The first goal is the characterization of the non-covalent interactions in biological species and their aggregates by a combination of DFT calculations and the Non-Covalent Interaction approach.

The second goal is the elaboration of a reductionist model applying theoretical calculations, based in amino acid-DNA base interactions, in order to explain the mechanism of the histone-DNA formation. Thus, the conformational landscape of three different amino acid of each type, according to the nature of their side chain and the amount of contacts with the DNA base, has been explored to understand the leading aggregation forces. The correlation with the statistical crystallographic observations has also been investigated.

Finally, the third goal is the study of the intrinsic preferences of the DNA base self-aggregation process, analyzing the tautomerical landscape of the DNA bases, focusing on the competition between hydrogen bond and dispersive contributions. A combination of theoretical calculations and spectroscopic techniques has been used to understand the self-aggregation of purines (guanine) and pyrimidines (cytosine) in gas phase. Furthermore, the influence of the methylation in the position where the sugar joins to the DNA base has also been also tested and how it affects the final aggregation process.

Keywords: DNA base, amino acid, histone-DNA complex, amino acid-DNA aggregate, building block, spectroscopy, REMPI, IR/UV spectroscopy, gas-phase, DFT, non-covalent interactions, NCI plot.

RESUMEN

Proteínas, ácidos nucleicos, azúcares y lípidos son considerados los constituyentes biológicos de la célula. Sus complejas relaciones mediante procesos físicos y químicos resultan en lo que se conoce como vida. En esta tesis doctoral se abordará el estudio de dos de estos constituyentes biológicos: ácidos nucleicos (ADN) y proteínas.

El ADN está constituido por una repetición de unidades fundamentales llamadas nucleótidos, formados por una base nitrogenada unida a un esqueleto de azúcar-fosfato.[1] Asimismo, su apariencia normal dentro de la célula incluye la combinación de dos hebras de azúcar-fosfato que se mantienen unidas mediante el apareamiento entre bases nitrogenadas complementarias. Así, adenina (A), timina (T), guanina (G) y citosina (C) son las cuatro bases nitrogenadas presentes en el ADN que constituyen lo que se conoce como "alfabeto de la vida", pues su combinación y secuencia permite el almacenamiento de todo el material genético de la célula. La estructura de cada componente celular es el resultado de la información almacenada en la secuencia de nucleótidos de los ácidos nucleicos de una célula. La razón de la elección de estas cuatro bases para construir el ADN en lugar de otras presentes en la célula con una estructura similar sigue siendo materia de debate. Algunas teorías apuntan a que la fotoestabilidad de estos compuestos, con estados excitados con tiempos de vida cortos, fue clave en la selección de las bases de ADN.[2] Sin embargo, otros metabolitos presentan una fotoestabilidad similar. Otras teorías apuntan a razones estructurales. La presencia de varios grupos polares en la molécula (>NH, $-NH_2$, >C=O) permite establecer múltiples enlaces de hidrógeno entre bases complementarias, así como la aromaticidad entre pares de bases apiladas, que permite establecer interacciones de tipo $\pi\cdots\pi$, refuerza la estabilidad de la estructura. Por lo tanto, la razón para elegir estas cuatro bases nitrogenadas puede estar condicionada por las interacciones intramoleculares presentes en el ADN, que le otorgan una gran estabilidad.[3,4]

A pesar de que la disposición más común de las bases nitrogenadas es el apareamiento de bases descrito por Watson y Crick (C-G/A-T), se han encontrado otras estructuras para formar la doble hélice de ADN que abarcan desde el apareamiento de bases de forma inusual (pares de Hoogsteen) a estructuras de más de dos bases nitrogenadas (tripletes o cuadrupletes).[5] Así, los tripletes pueden estar formados por bases complementarias ($C{\equiv}G{\cdot}C^+$ o $T{\cdot}A{=}T$)[6] o incluso por una misma base

(adenina). Además, existen zonas en el ADN ricas en una determinada base. Por ejemplo, los telómeros,[7] extremos de los cromosomas, presentan una secuencia concreta de bases nitrogenadas (TTAGGG) rica en guanina, que puede doblarse formando cuadrupletes de guanina.[8] Por otro lado, existen zonas ricas en citosina (*i-motifs*)[9] formadas por pares de citosinas ($C \cdot C^+$) combinadas de forma antiparalela. Toda esta variabilidad de estructuras nos hace preguntarnos acerca de las preferencias intrínsecas de las bases de ADN en el proceso de agregación, es decir, cual es la tendencia natural de las bases de ADN en este proceso.

Con el objetivo de analizar las propiedades de agregación inherentes a las bases de ADN, es necesario eliminar aquellos factores que puedan enmascarar los resultados, tal como el efecto del disolvente. Una forma de eliminar sus consecuencias es trabajar bajo condiciones aisladas en fase gaseosa mediante el uso combinado de ablación láser y jets supersónicos, que permiten crear las condiciones adecuadas para la agregación de moléculas.[10] Por un lado, la ablación láser permite desorber el material sólido, transformándolo en fase gaseosa sin ningún tipo de degradación. Así, se utiliza un láser pulsado para evaporar la muestra, con el fin de ser analizada mediante técnicas espectroscópicas. Por otro lado, el uso de expansiones supersónicas permite enfriar la mayoría de la población molecular al estado fundamental vibro-rotacional, simplificando su espectroscopía. Estas expansiones supersónicas son formadas fácilmente mediante una válvula pulsada, que permiten liberar un gas a alta presión hacia una cámara de vacío a través de un pequeño orificio o *nozzle*.

Una vez que se forman los agregados moleculares, se aplican las técnicas espectroscópicas con resolución de masas de ionización multifotónica resonante (REMPI) y de doble resonancia (IR/UV) para obtener información estructural relevante de las especies estudiadas. Así, dependiendo de la fuente láser, se puede obtener información de la región del espectro infrarrojo. En este estudio, abarcaremos la región del espectro de 200-4000 cm^{-1}, amplio intervalo de frecuencias, gracias a los experimentos realizados con el láser de electrones libres del laboratorio FELIX y el láser infrarrojo OPO/OPA del laboratorio de la UPV/EHU.

Sin embargo, a pesar de la cantidad de información proporcionada por estas dos técnicas, su interpretación requiere del apoyo de precisos cálculos teóricos. Por ejemplo, los sistemas analizados en esta tesis doctoral contienen varios grupos NH, cuyas vibraciones aparecen muy próximas en el espectro infrarrojo. Además, estos agregados presentan un panorama conformacional muy complicado con numerosos mínimos

locales que dan lugar a espectros muy similares, por lo que son necesarias predicciones teóricas de alto nivel para poder interpretar correctamente los espectros experimentales. Por otra parte, la metodología teórico-experimental aplicada en esta tesis permite abordar sistemas en los que las interacciones no covalentes están presentes.

El estudio de esta tesis doctoral se ha centrado en analizar el comportamiento natural de agregación de dos de las cuatro bases nitrogenadas: citosina[11] y guanina. Así, se han controlado las condiciones experimentales para formar trímeros de guanina y tetrámeros de citosina, analizando su espectroscopía infrarroja y comparándola con cálculos teóricos para ambos sistemas. Además, estos resultados se han comparado con los de sus derivados metilados, un modelo más realista, pues se impide la unión entre bases por los puntos de metilación, ocupados por el azúcar en las estructuras biológicas. Los resultados obtenidos han puesto de relieve que los tautómeros cetónicos de citosina, guanina y sus formas metiladas son las formas de agregación preferidas para estas bases, ello puede ser la razón por la que estas formas tautoméricas sean elegidas para el almacenamiento del material genético. Además, el proceso de agregación está gobernado por un patrón de enlaces de hidrógeno, que es diferente para cada base. Por un lado, la citosina adopta estructuras planas ensambladas por enlaces de hidrógeno entre N1-H···N3 y C2=O···H-NH. La metilación de la citosina impide la unión por la posición N1 por lo que la agregación está guiada por un patrón de doble enlace de hidrógeno HN-H···N3. Por otro lado, la complejidad tautomérica de la guanina dificulta la asignación del espectro experimental. Así, tanto la forma más estable de la guanina como su derivado metilado adoptan estructuras casi planas a través de un patrón de enlaces de hidrógeno N1-H···O=C6 y HN-H···O=C6. Sin embargo, también se ha observado la combinación con otras formas tautoméricas, exhibiendo estructuras planas con otros puntos de contacto. Finalmente, la competencia entre estructuras de apilamiento y estructuras de enlace de hidrógeno es diferente para citosina y guanina. Así, ambas bases muestran dímeros planos con una diferencia de energía alta con respecto a los dímeros apilados. Sin embargo, mientras que la diferencia de estabilidad se mantiene para agregados superiores de guanina (y su derivado metilado), en citosina esta diferencia disminuye ya en trímeros, invirtiéndose la estabilidad en los tetrámeros, cuyas estructuras apiladas están favorecidas entálpicamente. No obstante, los tetrámeros planos son más probables desde el punto de vista entrópico, lo que sugiere una competencia entre ambas tendencias reflejada en el espectro infrarrojo. Además, en la citosina metilada, esta competencia ya se observa en la formación del trímero.

El segundo tipo de unidades estructurales estudiadas en la tesis son las proteínas. Se trata de cadenas polipeptídicas formadas por una secuencia variable de aminoácidos que adopta una determinada estructura tridimensional. Estos aminoácidos se diferencian por la naturaleza de su cadena lateral, pudiendo tener cadenas polares, apolares, cargadas y aromáticas. Así, la naturaleza de las cadenas laterales determina la estructura tridimensional de la proteína y su función biológica.

Las interacciones entre el ADN y las proteínas son abundantes en la célula, siendo claves en muchos procesos biológicos como la replicación y la transcripción del ADN, los procesos de reconocimiento y metilación, unión y reparación del ADN, etc.[1] Uno de los procesos biológicos más interesantes y enigmáticos basados en los contactos entre el ADN y las proteínas es la formación del nucleosoma.[12] Esta superestructura está formada por la asociación de un octámero de histonas envuelto por una superhélice de 146 pares de bases de ADN, cuyo resultado es la formación de la cromatina de las células eucariotas.

Así, el mecanismo que guía la formación del complejo histona-DNA no es claro y existen varios interrogantes acerca de este mecanismo. ¿Tiene la histona una secuencia específica de aminoácidos que guía al ADN a doblarse durante la interacción o es el ADN el que presenta una secuencia en alguna sección que puede doblarse alrededor de las histonas?¿Puede cualquier hebra de ADN doblarse alrededor de cualquier histona o hay diferentes histonas para diferentes secuencias de ADN?

Existen estudios estadísticos acerca de la predisposición de contacto entre los aminoácidos y las bases de ADN[13] que muestran una gran disparidad en el número de contactos entre el residuo y las bases de ADN dependiendo de la naturaleza del aminoácido. Aquellos con cadenas laterales polares y cargadas presentan un mayor número de contactos que los de cadenas laterales no polares o aromáticos. Estas diferencias pueden apuntar a la existencia de un lenguaje especializado que usan las histonas para reconocer una secuencia específica del ADN con el fin de maximizar su interacción con él y compensar la energía necesaria para doblar la hebra de ADN.

El estudio se ha centrado en entender más profundamente las fuerzas que guían las interacciones entre el ADN y las histonas. Sin embargo, el gran tamaño del sistema, compuesto por miles de átomos, hace imposible analizarlo mediante cálculos mecanocuánticos. Por lo tanto, es necesario aplicar una aproximación reduccionista para abordar dicha superestructura, dividiendo las interacciones ADN – histona en sus unidades fundamentales: bases de ADN – aminoácido, que pueden ser

estudiadas con gran detalle utilizando métodos mecanocuánticos. Así, se ha construido un modelo de pares de base de ADN – aminoácidos para analizar las fuerzas que guían la naturaleza de esta interacción. Los cálculos se inician con una exhaustiva exploración del panorama conformacional de cada sistema base de ADN - aminoácido utilizando Mecánica Molecular, seguido de cálculos mecanocuánticos basados en la Teoría del Funcional de la Densidad (DFT) y una posterior caracterización de las interacciones no covalentes existentes en los agregados mediante el método de las Interacciones No Covalentes (NCI), basado en el análisis de la densidad electrónica y su gradiente reducido.

Para estudiar la validez del modelo, se han investigado 9 aminoácidos para determinar si el carácter de la cadena lateral es determinante en el proceso de agregación. Se han seleccionado tres aminoácidos no polares con baja tendencia a formar contactos (alanina, valina e isoleucina), los dos aminoácidos polares con mayor tendencia a formar contactos (asparagina y glutamina), el aminoácido con cadena lateral cargada con mayor número de contactos (arginina) y los tres aminoácidos aromáticos (fenilalanina, triptófano y tirosina). Estos aminoácidos se modificaron añadiendo un enlace peptídico en sus extremos para reproducir su existencia en una cadena peptídica. Por otro lado, se seleccionaron las cuatro bases nitrogenadas presentes en el ADN, metilando la posición de unión al azúcar para eliminar dicho punto de contacto, no presente cuando forman parte de una hebra de ADN.

Así, en una primera etapa, se analizó el panorama conformacional de los componentes aislados para seleccionar una metodología de cálculo adecuada y determinar las interacciones intramoleculares. Se observó que los aminoácidos mostraban una tendencia a adoptar ciertas conformaciones cíclicas de 5 y 7 miembros (β y γ), estabilizadas mediante interacciones no covalentes (enlaces de hidrógeno y fuerzas de van der Waals). Además, la naturaleza de la cadena lateral del aminoácido condiciona la estructura final del mismo. Así, mientras que los aminoácidos no polares apenas presentan interacciones con su cadena lateral, para los aminoácidos polares y aromáticos estas interacciones son primordiales, estableciendo enlaces de hidrógeno o interacciones con la densidad electrónica de cadena lateral. Por otro lado, las bases nitrogenadas apenas muestran interacciones intramoleculares, siempre debidas a la constricción del anillo aromático. Finalmente, se testearon varios métodos de cálculo con el objetivo de seleccionar el más adecuado para abordar el estudio de los agregados de ADN y proteínas, escogiéndose el método M06-2X, debido a su idoneidad para abordar

sistemas en los que las fuerzas dispersivas son predominantes junto con su bajo coste computacional.

Seleccionado el método de cálculo, se abordó el estudio de la agregación de pares de aminoácido - base de ADN, estudiando el panorama conformacional y las propiedades de agregación de la combinación de los 9 aminoácidos previamente estudiados con las 4 bases nitrogenadas del ADN. Se observó que dos tipos de interacciones intramoleculares, enlaces de hidrógeno y fuerzas dispersivas de van der Waals, controlan las propiedades de agregación de estos sistemas, así como patrones diferentes de agregación, en función de la naturaleza de la cadena lateral del aminoácido. Así, los aminoácidos no polares tienden a establecer enlaces de hidrógeno con los grupos polares del esqueleto peptídico, colocando su cadena lateral alejada de la base. Sin embargo, la cadena lateral de los aminoácidos polares juega un papel principal en el proceso de agregación, interaccionando directamente con la base de ADN, que también se estabiliza por interacciones con el esqueleto peptídico. Los aminoácidos aromáticos interaccionan de manera similar a los polares, aunque en este caso la cadena lateral aromática se coloca en paralelo a la base de ADN para establecer interacciones de tipo π-π. Por otro lado, se ha observado este mismo comportamiento en la agregación con bases complementarias. Así, los agregados con citosina y guanina presentan una energía de enlace mayor con mayor número de enlaces de hidrogeno y menor contribución por fuerzas dispersivas que los complejos con adenina y timina. Finalmente, se abordó un estudio de correlación de las interacciones individuales estudiadas y su comparación sobre los puntos de contacto presentes en bibliografía, reflejada en la energía de enlace de los pares estudiados. Se observó que existía una clara correlación para aminoácidos no polares y polares, reproduciendo los resultados estadísticos en todos los casos excepto en la asparagina, que presenta una energía de enlace menor de la esperada. Sin embargo, esta correlación no se reproduce para los aminoácidos aromáticos, cuyos valores de energía de enlace son superiores a lo esperado debido al establecimiento de interacciones de carácter π con la cadena lateral (no posibles cuando hay apilamiento de bases).

Además, se aumentó el tamaño del sistema estudiado para reproducir la doble hélice de bases de ADN así como considerar la existencia de puntos de contacto entre el aminoácido y la base no accesibles debido a la complementariedad de bases. Para ello, se enfrentaron los 9 aminoácidos estudiados a los pares de bases propuestos por Watson y Crick: A-T y C-G. Se observó que las conclusiones extraídas anteriormente se mantienen cuando se aumenta el tamaño del

sistema. Así, se reprodujeron los mismos tres patrones de agregación anteriormente indicados, observándose apenas diferencias entre la tendencia de contactos y la energía de enlace de los agregados. Es decir, la correlación sigue siendo buena para aminoácidos polares y no polares, aunque sigue sin ser adecuada para aminoácidos aromáticos, ya que el aumento del tamaño del sistema no permite resolver las restricciones estéricas de la cadena lateral aromática.

En conclusión, el balance entre los enlaces de hidrógeno y las interacciones de van der Waals que controla la forma final y la energía de enlace de los pares aminoácido – base de ADN está convenientemente considerado en el método de cálculo aplicado (MM+DFT, en concreto M06-2X). Los resultados computacionales ofrecen una buena caracterización de las interacciones dominantes presentes en la interacción ADN – histona, siendo una herramienta útil para la predicción de las propiedades de agregación del ADN. Así, la caracterización de las interacciones entre aminoácidos y bases de ADN puede ayudar a esclarecer el lenguaje que guía el proceso de reconocimiento de los aminoácidos de una determinada histona y las secuencias específicas del ADN.[14]

1 D. L. Nelson and M. M. Cox, *Principles of Biochemistry*, Omega, New York, 4th edn., 2000.

2 E. T. Kool, J. C. Morales and K. M. Guckian, *Angew. Chemie - Int. Ed.*, 2000, **39**, 990–1009.

3 C. Plützer, I. Hünig, K. Kleinermanns, E. Nir and M. S. De Vries, *ChemPhysChem*, 2003, **4**, 838–842.

4 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2002, **4**, 740–750.

5 J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, **40**, 5893–5909.

6 S. M. Mirkin and M. D. Frank-Kamenetskii, *Annu. Rev. Biomol. Struc*, 1994, **23**, 541–576.

7 E. H. Blackburn, *Nature*, 1991, **350**, 569–73.

8 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–80.

9 H. A. Day, P. Pavlou and Z. A. E. Waller, *Bioorganic Med. Chem.*, 2014, **22**, 4407–4418.

10 M. S. de Vries, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, eds. A. M. Rijs and J. Oomens, Springer International Publishing, Cham, 2015, pp. 271–297.

11 J. Gonzalez, I. Usabiaga, P. F. Arnaiz, I. Leon, R. Martínez, J. Millán and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2017, (DOI: 10.1039/C6CP08476A).

12 M. M. Müller and T. W. Muir, *Chem. Rev.*, 2015, **115**, 2296–2349.

13 R. Sathyapriya, M. S. Vijayabaskar and S. Vishveshwara, *PLoS Comput. Biol.*, 2008, **4**, e1000170.

14 J. González, I. Baños, I. León, J. Contreras-García, E. J. Cocinero, A. Lesarri, J. A. Fernández and J. Millán, *J. Chem. Theory Comput.*, 2016, **12**, 523–534.

CHAPTER 1

INTRODUCTION

1. - INTRODUCTION

DNA and proteins, sugars and lipids are the building blocks of the cell, where they play specialized roles. They present complex relationship and are involved in different physical and chemical processes that result in what we know as life.

DNA is formed by a combination of two strands of a sugar-phosphate backbone that maintain held together a long and variable sequence of four molecules, cytosine (C), guanine (G), adenine (A) and thymine (T), known as DNA bases. The four DNA bases constitute what has been called "the alphabet of life" or the genetic code, as their combination into DNA and this into chromosomes is the mechanism used by most living beings to pass the blueprints to build the new generation of organisms to their descendants.[1,2] In human beings, the information required to synthetize the nearly 100,000 proteins that may be found in a cell are encoded in three billion base pairs.[3] Thus, the structure of every cellular component is the result of the information packaged in the nucleotide sequence of a cell's nucleic acids.

The reason why these four molecules were chosen to build the DNA instead of others present in the cell is still a material of discussion. Nucleotides have a variety of roles in cellular metabolism.[1,2,4] For example, adenine is also the base to build adeninetri/diphosphate (ATP/ADP), one of the energy tokens of the cell metabolism.[1,5] Likewise guanine is an essential component of guanosinetri/diphosphate (GTP/GDP),[6] an essential metabolism for the function of G-protein coupled receptors (GPCR), a wide family of membrane receptors.[7–10] Thus, it seems that nucleotides have several functions in a cell, apart from being essential components of DNA. However, many other metabolites exist inside the cell, some of them with remarkably similar structures that could in principle be used to build DNA.[1,2] Some authors hypothesize with the natural selection occurred in a primitive earth due to the strong UV radiation that presumably bathed it, in absence of the filters of the present day's atmosphere.[3] This hypothesis is supported by a number of studies on the photodissociation or photoinduced degradation of DNA bases which remarks that the four bases and their nucleotide derivatives present short electronic exited state lifetime.[11–21]

Indeed, photostability is an important property of the DNA bases. Even with present UV radiation doses there is an unneglectable probability of the UV to induce mutations in our DNA that result in cancer appearance.[22–24] However, there are still many other metabolites that

present short excited state lifetimes, of the order of those presented by the DNA bases and that become even shorter when forming a supermolecule such as a protein or DNA itself. Certainly, as molecules join together to grow a biological polymer, no matter their individual nature, the accumulation of electronic and vibrational states results in very short lifetimes.[15,18]

Thus, there may be additional reasons for life to have chosen CGAT as its alphabet. It is worthy to note the remarkably stable structure that combination of these four bases produces. They present several polar groups (>NH, -NH_2, >C=O) that allow them to form strong and multiple hydrogen bonds between complementary bases (C≡G, A=T).[25,26] Furthermore, the four bases are aromatic and therefore they can establish π···π interactions.[27] The result is a very stable structure,[1,3,4,28,29] able even to survive outside the nuclear cell and at high temperatures. Actually, it enables forensic sciences to use DNA as strong evidences of the presence of individuals in a given place, or to sequence ancient DNA from archeological sites. It is not unlikely that the reason for life to choose CGAT was in part conditioned by the intermolecular interactions that those molecules present and that give DNA part of its high stability.[3]

The first accepted representation of the DNA was proposed by Watson and Crick in 1953 and it was based in two strands of DNA paired as an antiparallel double helix,[30] as **Figure 1.1** shows. This model was immediately accepted by the scientific community due to its elegance to explain the genetic replication and the agreement with the Schrodinger's "aperiodic crystal" concept throughout the complementarity of bases, which maintains the size and the shape of the coding unit and ensures to find its complementary partner.[28]

However, these canonical dimeric structures (C-G/A-T) are not the only ones found in DNA. In 1963, Hoogsteen found evidences of other types of pairing,[31] and that were typical in RNAs. This alternative pairing was called Hoogsteen pairing and was based on the interaction of the pyridine base on the other side of the purine base, interacting with purine's N7.

This was not the only unusual structure found in other works, where the existence of trimers and even tetramers of DNA bases was demonstrated, although in many cases they involved protonation of at least one of the participants or incorporation of a metal cation.[32] Triplex DNA bases can be formed by a Hoogsteen pairing of a protonated cytidine residue with the guanosine of a GC, generating the C≡G·C^+. Likewise, TA pairs can also form T·A=T triplexes by the pairing of an additional

thymine molecule.[2,33–36] Protein Data Bank (PDB) is an online storehouse of crystallographic structural information of biological macromolecules, such as nucleic acids and proteins.[37,38] **Figure 1.2** shows an example of structure (PDB: 1BWG)[36] which presents triplexes of T·A=T and C^{+}·G≡C. Furthermore, triplexes formed by repetition of the same base were also identified in combination with other aggregates, such as adenine triplexes.[39]

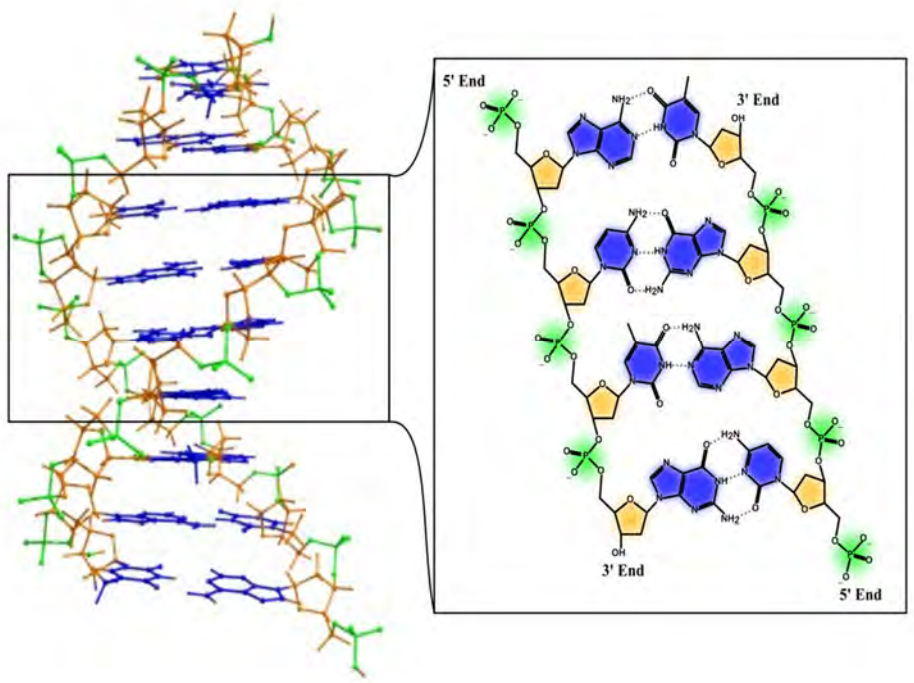


Figure 1.1 *DNA model proposed by Watson and Crick* [30] *and schematic of a section of a DNA double strand. Phosphates are depicted in green; pentoses in orange and DNA bases in blue.*

DNA strands can also form quadruplexes. Human telomeres, a kind of "end-of-the-road" signal for DNA replication, are formed by repetition of TTAGGG sequences.[40–42] These very rich guanine DNA strands are able to bend forming quadruplexes of guanines (G-quadruplexes), usually stabilized by a K^{+} cation. **Figure1.3a** exhibits the structure of PDB id: 2KZD[43] with the G-quadruplexes marked in green.[44–50]

Furthermore, sequences rich in cytosines can also form different tetrameric structures called i-motifs.[51–54] These i-motifs may be constituted by one, two or four different strands and consist of two parallel

duplexes of intercalated $C{\cdot}C^+$ base pairs combined in an antiparallel way.[52] The sequences of cytosines can contain from 3 to 12 bases, with/without thymine residues at various points in the sequence.[55–57] **Figure 1.3b** shows the shape of the d(TC_5) i-motif (PDB ID: 225D) [52] with the $C{\cdot}C^+$ base pair shaded in pink. Other tetrads of adenine and thymine were also reported.[58]

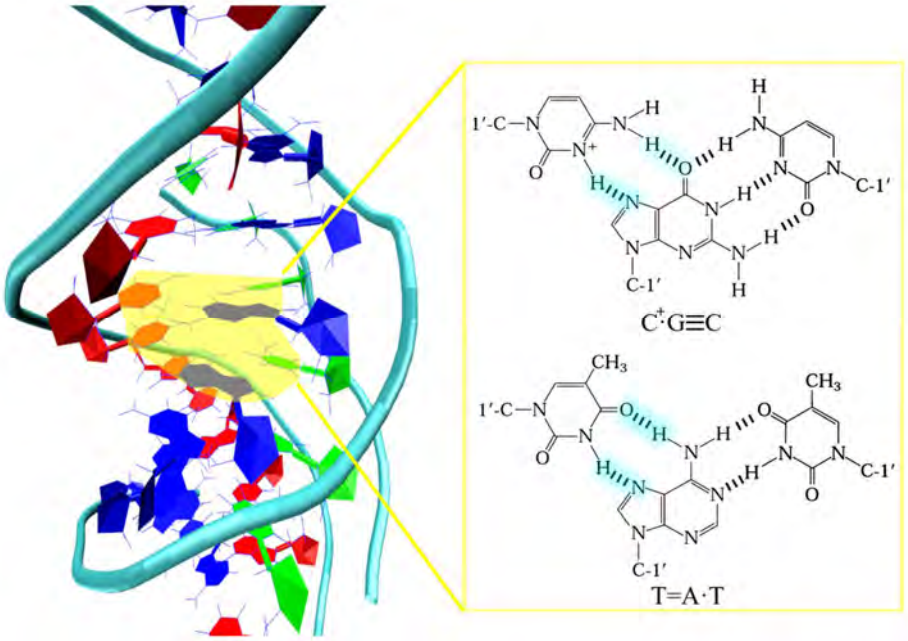


Figure 1.2 *Representation of the DNA Triple helix (PDB ID: 1BWG)[36] with the DNA triplexes T·A=T and C^+·G≡C shaded in yellow. Green and blue strands are related with parallel strands and the red one corresponds with the antiparallel strand. The Hoogsteen pairing is marked in blue.*

These observations raise the questions of which are the real intrinsic preferences of the DNA bases in the aggregation processes and which is the role of such preferences in the appearance of the first DNA strands. The objective of this work is therefore to shed some light on the natural aggregation preferences of DNA bases.

In order to answer the above questions and analyze the inherent properties of DNA base aggregation, it is necessary to eliminate those factors which can mask the results, as the solvent effect.[27] A way to remove solvent effect is to study the systems in isolation conditions, for

example in the gas phase, using a combination of laser ablation and a supersonic expansion to create the ideal conditions for the molecules to aggregate. Employment of supersonic jets together with laser spectroscopy opened the field to significantly larger systems.[59,60] At room temperature, the several vibronic states of the molecules are populated, whose transitions congest the spectrum. With the use of supersonic expansions, most of the population of the molecules is cooled to the ground vibro-rotational state, simplifying their spectroscopy.[61]

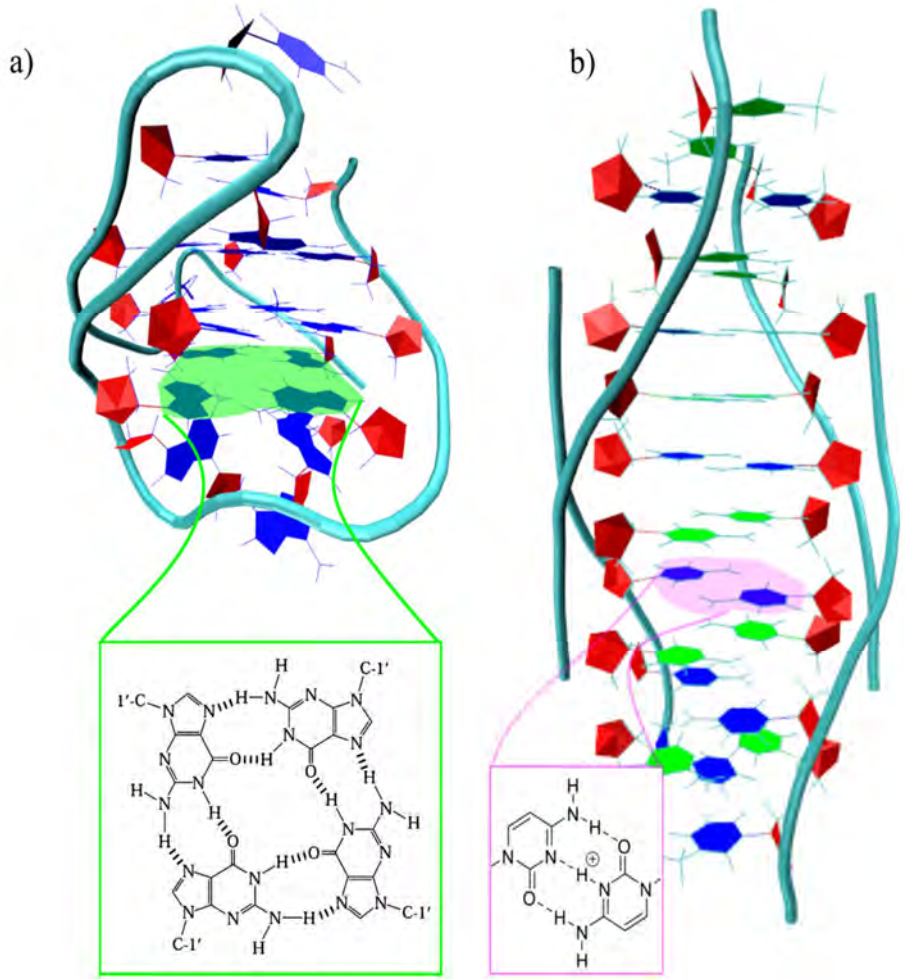


Figure 1.3 a) *Representation of the B-DNA structure (PDB id: 2KZD)*[43] *with the G-quadruplex marked in green.* ***b)*** *Structure of the d(TC5) intermolecular i-motif (PDB id: 225D)*[52] *with the* $C{\cdot}C^{+}$ *base pair shaded in pink. Blue and green depicts base pairs at different height.*

Jets are easily formed connecting a reservoir of gas at high pressure with a vacuum chamber, through a small orifice (nozzle). The pressure difference pushes the molecules through the narrow exit, where the collisions transform their ro-vibrational energy into translational energy.[60–62] Thus, the molecules exit the nozzle at supersonic velocity, forming a molecular beam with a relatively high molecular density, but in which the molecules are isolated, as all of them travel in the same direction. Development of laser spectroscopy in jets also permitted to tackle the spectroscopy of molecular aggregates, as the low temperature and isolation conditions in the expansion are ideal to study supramolecular systems formed by non-covalent interactions.[63,64]

Once the aggregates are formed, mass-resolved spectroscopic techniques, resonance enhanced multiphoton ionization (REMPI) and IR/UV double resonance techniques are used to obtain reliable structural information from the species. Indeed, mass-resolved IR/UV has demonstrated to be a powerful technique. Depending on the laser source, it is possible to extract information from the fingerprint region up to the NH and OH stretch region, covering from 200-4000 cm^{-1}, although not in a single experiment. While the ~2800-4000 cm^{-1} window is easily achievable with table-top lasers, which can even extend their emission window down to ~2200 cm^{-1} with the appropriate non-linear optics, regions below require of complicated and large installations such as free-electron lasers, like the one at FELIX facility, to be probed. Nevertheless, the region of the stretching vibrations has proven to be the most informative one, at least for molecules containing hydroxyl and amino groups. Such groups are prone to establish intermolecular interactions and even the most subtle ones produce measurable perturbations in the stretch vibrations that result in spectral shifts that usually correlate with the strength of the interaction.

However, all the structural information obtained using IR/UV is very often difficult to interpret and usually requires of the guidance of theoretical calculations for their interpretation. For example, the systems tackled in this work contain several NH groups, whose stretch vibrations lay close together. In addition, larger systems have complicated conformational landscapes with numerous local minima, whose spectra are very similar and therefore accurate theoretical predictions of their respective IR spectra are required to assign the experimental spectra.

Despite its complexity, if this combined approach is properly applied it allows one to tackle the spectroscopy of numerous systems in which non-covalent interactions are present.[65–71]

Thus, to achieve the above-proposed goal of understanding the natural aggregation preferences of DNA bases, we used a combination of laser spectroscopy in jets and quantum mechanical calculations, to explore cytosine and guanine aggregates, in which there is a competition between H-bonding and stacking interactions.[70] Despite the complexity of the systems, all possible tautomers were taken into account not to leave aside relevant conformations. Furthermore, the influence of the substitution of the hydrogen position to form the nucleoside in the final aggregation process was also explored.

Following with the cell organization, while DNA stores the genetic code, proteins are the nano-machines that perform all the duties in the cell. They are polypeptide chains formed by a variable sequence of amino acids, which differ in the side chain, and that adopt a functional three-dimensional structure (a functional structure). The nature of the side chains determines the biological function and the specific three-dimensional structure of the protein.[1,2,4,72]

The interactions between proteins and DNA are ubiquitous in the cell and many essential biological processes are based in DNA-protein contacts, such as DNA replication and transcription, methylation and recognition processes, splicing and repair, etc.[73–75] For example, DNA replication takes place by the interaction between proteins called DNA polymerases with DNA strands. These enzymes produce identical replicates of an original DNA double strand. A large number of studies with the structures of DNA polymerases may be found in the PDB.[76–80]

DNA methylation is an epigenetic modification to repress gene expression and affects a wide variety of different biological pathways. For example, Ten-eleven translocation (TETs) proteins interact with DNA removing these modifications and silencing the onset of the disease.[81–87] But DNA methylations are not the only modification which DNA can experience and other enzymes are involved in slicing and repair modifications and DNA lesions.[88,89] Thus, DNA mismatch,[90–94] base excision,[95–100] DNA oxidation[100–103]or DNA alkylation[104–111] are typical harms of DNA strands. When any of them happens, a checkpoint is activated in order to recruit specialized proteins to repair the damage. For example, **Figure 1.4** depicts the interaction between a DNA polymerase (PDB ID: 5DLG)[111] with an O^4-methylthymidine of a DNA strand.

One of the most puzzling and interesting examples of these DNA-protein contacts is the nucleosome formation by association of a histone supercomplex with a DNA strand, which results in the eukaryotic chromatin.[112–116] Such superstructures are formed by a superhelix of 147

base pairs of DNA wrapped around a histone octamer, as X-ray studies show.[117,118] **Figure 1.5** depicts the crystal structure of the nucleosome core particle (PDB ID: 1KXE).[118] Apart from having a clear structural function, they play a regulatory role in the expression of DNA[119–123] and their structural changes can be inherited by the next generations.[124]

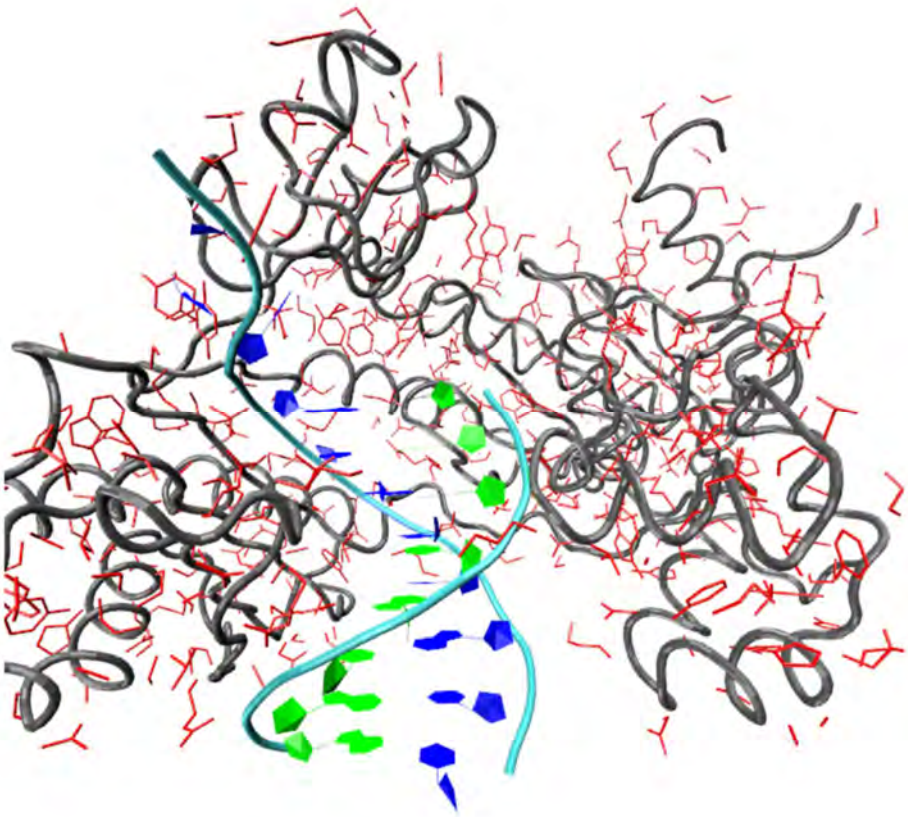


Figure 1.4 *Representation of the DNA polymerase (PDB ID: 5DLG),*[111] *interacting with a DNA double strand. Green and blue: DNA strand; grey: DNA polymerase; red: side chains of the DNA polimerase amino acids.*

The mechanism that guides formation of the histone-DNA complex is not clear. Does the histone have a specialized sequence of amino acids that guides DNA during the process, forcing it to bend, or is DNA that presents a special sequence in the sections that may bend around the histone? Can any DNA strand wrap around any histone or are there different histones for different DNA sequences? Statistical studies on the

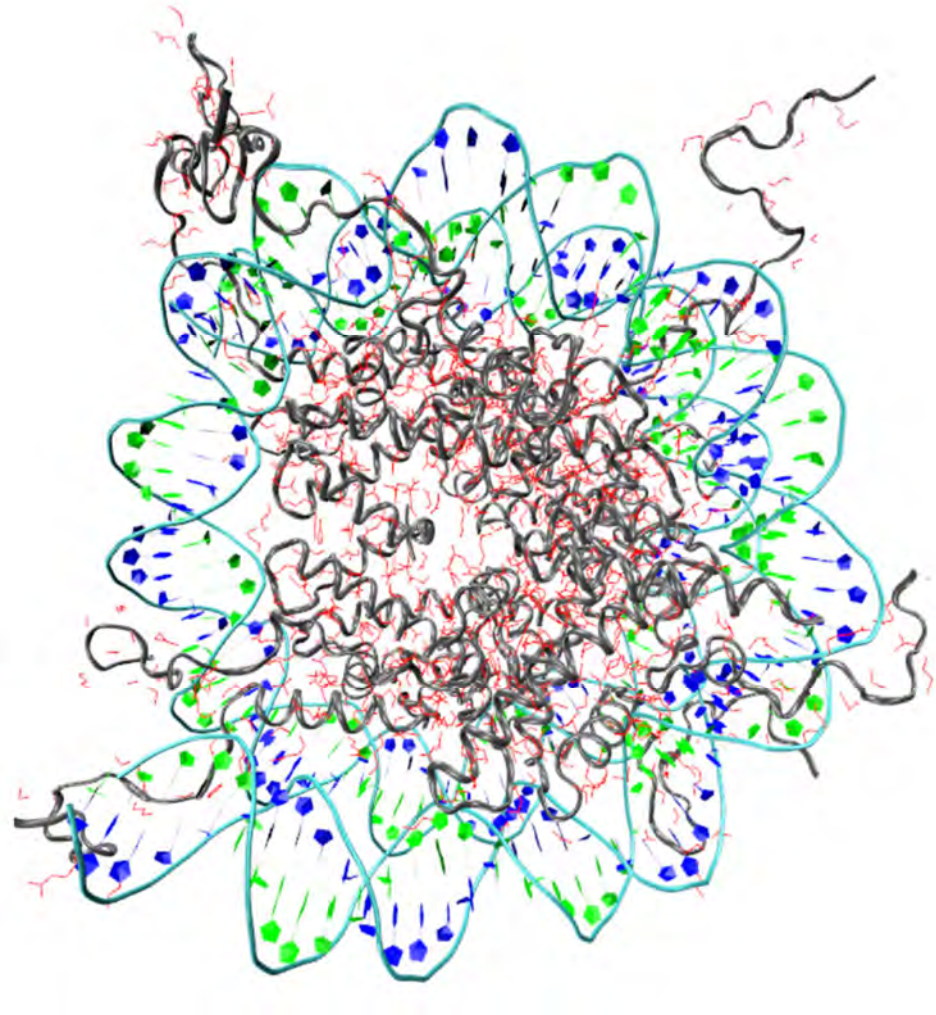


Figure 1.5 *Structure of the nucleosome (PDB ID: 1KX5.)*[118] *Green and blue: DNA strand; grey: histone octamer; red: side chains of the histone octamer amino acids.*

contact propensity between the amino acids of the histones and the DNA bases show large disparity in the number of contacts between the residue of the amino acids and the nucleobases,[125–131] depending on the nature of the amino acid: those with polar or charged lateral chains present a significantly larger number of contacts than those with non-polar or aromatic residues. **Figure 1.6** shows the variation on the number of contacts between amino acid – DNA base obtained from ref. 131. Such differences may be pointing to the existence of a specialized language that

the histones use to recognize a specific sequence of DNA to interact with, or to a specialized design of the histones to maximize the interaction with DNA and to compensate for the desolvation energy and for the stress induced in the DNA strand due to the bending. In such case, the sequence of amino acids of the histones would be determined by the DNA segment to which a given histone is designed to face those amino acids which has a higher number of contacts with the DNA strand, adapting its 3D-structure to the DNA sequence.

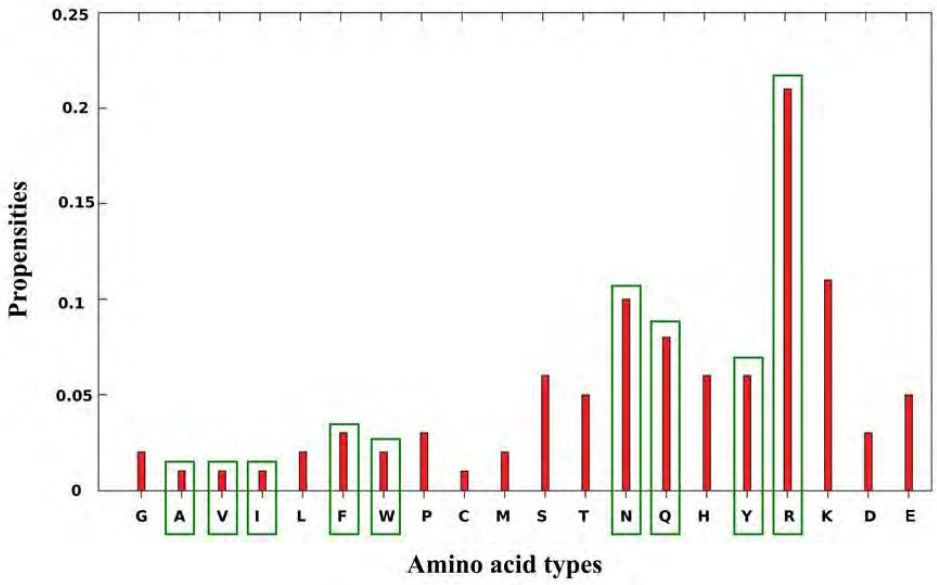


Figure 1.6 *Propensities of the amino acids with the DNA bases.*[131] *The green boxes note the selected amino acids for the study.*

One of the objectives of this work was to gain more in-depth knowledge on the DNA-histone interactions. However, the huge size of the system, formed by thousands of atoms, makes impossible to approach the problem using quantum mechanics calculations. In order to address this superstructure, a reductionist approach was applied, dividing the histone-DNA interactions into amino acid-DNA base pairs, which may be studied in great detail using high-level quantum chemistry procedures.

Thus, a model of amino acid-DNA base pair was constructed in this work to study the leading forces which guide the aggregation process.[132] This model was based in an exhaustive exploration of the conformational landscape of each dimer using molecular mechanics, followed by quantum

mechanical calculations based in density-functional theory. The non-covalent interactions of the aggregates were also characterized by the non-covalent interaction (NCI) approach.[133]

In this computational work, a set of non-polar (alanine, valine and isoleucine), polar (asparagine, glutamine), charged (arginine) and aromatic (phenylalanine, tryptophan and tyrosine) amino acids were selected to elucidate if the nature of the side chain amino acid is relevant in the aggregation process. Different DFT methods, which take into account dispersive contribution, were tested for comparison. [134–136] The individual interactions were compared to the statistical observations in order to correlate both magnitudes and clarify the role played by the nature of the amino acids of the histone in the DNA-histone interactions and if its sequence follows an intelligent design that maximizes the affinity of the DNA strand for the correct portion of the histone.

1.1 References:

1 B. Alberts, A. Johnson, D. Morgan, M. Raff and J. Lewis, *Molecular Biology of the Cell*, Taylor & Francis Inc, New York, 2014.

2 D. L. Nelson and M. M. Cox, *Principles of Biochemistry*, Omega, New York, 4th edn., 2000.

3 E. T. Kool, J. C. Morales and K. M. Guckian, *Angew. Chemie - Int. Ed.*, 2000, **39**, 990–1009.

4 L. Stryer, J. M. Berg and J. L. Tymoczko, *Biochemistry*, Reverté, New York, 5th edn., 2002.

5 M. Klingenberg, *Biochim. Biophys. Acta - Biomembr.*, 2008, **1778**, 1978–2021.

6 H. R. Bourne, D. A. Sanders and F. McCormick, *Nature*, 1991, **349**, 117–127.

7 D. M. Rosenbaum, S. G. F. Rasmussen and B. K. Kobilka, *Nature*, 2009, **459**, 356–363.

8 Q. Bin Gao and Z. Z. Wang, *Protein Eng. Des. Sel.*, 2006, **19**, 511–516.

9 V. Katritch, V. Cherezov and R. C. Stevens, *Annu Rev Pharmacol Toxicol*, 2013, **53**, 531–556.

10 N. S. Schonenbach, S. Hussain and M. A. O'Malley, *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology*, 2015, **7**, 408–427.

11 S. Lobsiger, M. Etinski, S. Blaser, H. M. Frey, C. Marian and S. Leutwyler, *J. Chem. Phys.*, 2015, **143**, 0–12.

12 C. G. Triandafillou and S. Matsika, *J. Phys. Chem. A*, 2013, **117**, 12165–12174.

13 D. Mishra and S. Pal, *J. Mol. Struct.-THEOCHEM*, 2009, **902**, 96–102.

14 O. Kostko, K. Bravaya, A. Krylov and M. Ahmed, *Phys. Chem. Chem. Phys.*, 2010, **12**, 2860–2872.

15 A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, *Proc. Natl. Acad. Sci.*, 2005, **102**, 20–23.

16 K. Kleinermanns, D. Nachtigallová and M. S. De Vries, *Int. Rev. Phys. Chem.*, 2013, **32**, 308–342.

17 C. E. Crespo-Hernández, L. Martínez-Fernández, C. Rauer, C. Reichardt, S. Mai, M. Pollum, P. Marquetand, L. González and I. Corral, *J. Am. Chem. Soc.*, 2015, **137**, 4368–4381.

18 C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández and B. Kohler, *Annu. Rev. Phys. Chem.*, 2009, **60**, 217–239.

19 E. Mburu and S. Matsika, *J. Phys. Chem. A*, 2008, **112**, 12485–12491.

20 Z. Gengeliczki, M. P. Callahan, N. Svadlenak, C. I. Pongor, B. Sztáray, L. Meerts, D. Nachtigallová, P. Hobza, M. Barbatti, H. Lischka and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5375–5388.

21 A. Beckstead, Y. Zhang, M. S. de Vries and B. Kohler, *Phys. Chem. Chem. Phys.*, 2016, **18**, 24228–24238.

22 G. P. Pfeifer, Y.-H. You and A. Besaratinia, *Mutat. Res.*, 2005, **571**, 19–31.

23 V. O. Melnikova and H. N. Ananthaswamy, *Mutat. Res.*, 2005, **571**, 91–106.

24 L. Daya-Grosjean and A. Sarasin, *Mutat. Res.*, 2005, **571**, 43–56.

25 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2002, **4**, 740–750.

26 C. Plützer, I. Hünig, K. Kleinermanns, E. Nir and M. S. De Vries, *ChemPhysChem*, 2003, **4**, 838–842.

27 A. S. Mahadevi and G. N. Sastry, *Chem. Rev.*, 2016, **116**, 2775–2825.

28 S. A. Benner, *ACS Cent. Sci.*, 2016, acscentsci.6b00344.

29 H. Karabıyık, R. Sevinçek and H. Karabıyık, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15527–38.

30 F. H. C. Watson, J. D.; Crick, *Nature*, 1953, 171, 737–738.

31 K. Hoogsteen, *Acta Crystallogr.*, 1963, **16**, 28–38.

32 J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, **40**, 5893–5909.

33 S. M. Mirkin and M. D. Frank-Kamenetskii, *Annu. Rev. Biomol. Struc*, 1994, **23**, 541–576.

34 L. Lavelle and J. R. Fresco, *Nucleic Acids Res.*, 1995, **23**, 2692–2705.

35 J. L. Asensio, T. Brown and A. N. Lane, *Nucleic Acids Res.*, 1998, **26**, 3677–3686.

36 J. L. Asensio, T. Brown and A. N. Lane, *Structure*, 1999, **7**, 1–11.

37 www.rcsb.org.

38 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.

39 J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 2440–2450.

40 E. H. Blackburn, *Nature*, 1991, **350**, 569–73.

41 S. Misri, S. Pandita, R. Kumar and T. Pandita, *Cytogenet. Genome Res.*, 2009, **122**, 297–307.

42 R. J. O'Sullivan and J. Karlseder, *Nat Rev Mol Cell Biol*, 2010, **11**, 171–181.

43 K. W. Lim, L. Lacroix, D. J. E. Yue, J. K. C. Lim, J. M. W. Lim and A. T. Phan, *J. Am. Chem. Soc.*, 2010, **132**, 12331–12342.

44 Z. Y. Kan, Y. Lin, F. Wang, X. Y. Zhuang, Y. Zhao, D. W. Pang, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2007, **35**, 3646–3653.

45 H. Han and L. H. Hurley, *Trends Pharmacol. Sci.*, 2000, **21**, 136–142.

46 V. Setnička, J. Nový, S. Böhm, N. Sreenivasachary, M. Urbanová and K. Volka, *Langmuir*, 2008, **24**, 7520–7527.

47 Q. Wang, J. Q. Liu, Z. Chen, K. W. Zheng, C. Y. Chen, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2011, **39**, 6229–6237.

48 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–80.

49 M. Rueda, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2006, **128**, 3608–3619.

50 V. Sychrovský, Z. Sochorová Vokáčová and L. Trantírek, *J. Phys. Chem. A*, 2012, **116**, 4144–4151.

51 M. Guéron and J. L. Leroy, *Curr. Opin. Struct. Biol.*, 2000, **10**, 326–331.

52 H. A. Day, P. Pavlou and Z. A. E. Waller, *Bioorganic Med. Chem.*, 2014, **22**, 4407–4418.

53 A. T. Phan and J.-L. Mergny, *Nucleic Acids Res.*, 2002, **30**, 4618–4625.

54 S. Benabou, R. Ferreira, A. Aviño, C. González, S. Lyonnais, M. Solà, R. Eritja, J. Jaumot and R. Gargallo, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, **1840**, 41–52.

55 X. Han, J. L. Leroy and M. Guéron, *J. Mol. Biol.*, 1998, **278**, 949–965.

56 A. T. Phan, M. Guéron and J. L. Leroy, *J. Mol. Biol.*, 2000, **299**, 123–44.

57 K. Kanaori, N. Shibayama, K. Gohda, K. Tajima and K. Makino, *Nucleic Acids Res.*, 2001, **29**, 831–40.

58 B. Pan, Y. Xiong, K. Shi, J. Deng and M. Sundaralingam, *Structure*, 2003, **11**, 815–823.

59 A. Kantrowitz and J. Grey, *Rev. Sci. Instrum.*, 1951, **22**, 328–332.

60 D. H. Levy, *Annu. Rev. Phys. Chem.*, 1980, **31**, 197–225.

61 M. V. Johnston, *Trends Anal. Chem.*, 1984, **3**, 58–61.

62 T. A. Miller, *Science*, 1984, **223**, 545–553.

63 R. E. Smalley, L. Wharton and D. H. Levy, *Acc. Chem. Res.*, 1977, **10**, 139–145.

64 G. Meijer, M. S. de Vries, H. E. Hunziker and H. R. Wendt, *Appl. Phys. B*, 1990, **51**, 395–403.

65 I. Usabiaga, J. González, P. F. Arnáiz, I. León, E. J. Cocinero and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12457–12465.

66 I. Leon, I. Usabiaga, J. Millan, E. J. Cocinero, A. Lesarri and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16968–16975.

67 I. León, E. J. Cocinero, J. Millán, A. M. Rijs, I. Usabiaga, A. Lesarri, F. Castaño and J. A. Fernández, *J. Chem. Phys.*, 2012, **137**, 74303.

68 E. Aguado, I. León, E. J. Cocinero, A. Lesarri, J. A. Fernández and F. Castaño, *Phys. Chem. Chem. Phys.*, 2009, **11**, 11608–11616.

69 I. León, J. González, J. Millán, F. Castaño and J. A. Fernández, *J. Phys. Chem. A*, 2014, **118**, 2568–2575.

70 J. Gonzalez, I. Usabiaga, P. F. Arnaiz, I. Leon, R. Martínez, J. Millán and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2017, (DOI: 10.1039/C6CP08476A).

71 I. Usabiaga, J. González, I. León, P. F. Arnaiz, E. J. Cocinero and J. A. Fernández, *J. Phys. Chem. Lett.*, 2017, **8**, 1147–1151.

72 J. S. Richardson, *The Anatomy & Taxonomy of Protein Structure*, 2007.

73 N. M. Luscombe, S. E. Austin, H. M. Berman and J. M. Thornton, *Genome Biol.*, 2000, **1**, 1–10.

74 H. Zhao, J. Wang, Y. Zhou and Y. Yang, *PLoS One*, 2014, **9**, 26–28.

75 S. A. Coulocheri, D. G. Pigis, K. A. Papavassiliou and A. G. Papavassiliou, *Biochimie*, 2007, **89**, 1291–1303.

76 M. a Schumacher, N. K. Tonthat, S. M. Kwong, N. B. Chinnam, M. a Liu, R. a Skurray and N. Firth, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 9121–6.

77 Y. Gao and W. Yang, *Science*, 2016, **352**, 1334–7.

78 T. Oyama, S. Ishino, T. Shirai, T. Yamagami, M. Nagata, H. Ogino, M. Kusunoki and Y. Ishino, *Nucleic Acids Res.*, 2016, **44**, 9505–9517.

79 A. G. Baranovskiy, N. D. Babayeva, Y. Zhang, J. Gu, Y. Suwa, Y. I. Pavlov and T. H. Tahirov, *J. Biol. Chem.*, 2016, **291**, 10006–10020.

80 R. Fernandez-Leiro, J. Conrad, S. H. W. Scheres and M. H. Lamers, *Elife*, 2015, **4**, 1–16.

81 W. A. Pastor, L. Aravind, A. Rao, G. G. Lin and J. G. Scott, *Nat Rev Mol Cell Biol*, 2013, **14**, 341–356.

82 X. Lu, B. S. Zhao and C. He, *Chem. Rev.*, 2015, **115**, 2225–2239.

83 K. Williams, J. Christensen and K. Helin, *EMBO Rep.*, 2012, **13**, 28–35.

84 R. M. Kohli and Y. Zhang, *Nature*, 2013, **502**, 472–479.

85 L. Scourzic, E. Mouly and O. A. Bernard, *Genome Med.*, 2015, **7**, 1–16.

86 H. Wu and Y. Zhang, *Genes Dev.*, 2011, **25**, 2436–2452.

87 K. D. Rasmussen and K. Helin, *Genes Dev.*, 2016, **30**, 733–750.

88 R. D. Wood, M. Mitchell and T. Lindahl, *Mutat. Res.*, 2005, **577**, 275–283.

89 S. Schneider, S. Schorr and T. Carell, *Curr. Opin. Struct. Biol.*, 2009, **19**, 87–95.

90 T. A. Kunkel and D. A. Erie, *Annu. Rev. Biochem.*, 2005, **74**, 681–710.

91 P. Hsieh and K. Yamane, *Mech. Ageing. Dev.*, 2008, **129**, 391–407.

92 J. Bowers, P. T. Tran, A. Joshi, R. M. Liskay and E. Alani, *J. Mol. Biol.*, 2001, **306**, 957–968.

93 S. D. Lee and E. Alani, *J. Mol. Biol.*, 2006, **355**, 175–184.

94 P. K. Patel, A. S. R. Koti and R. V. Hosur, *Nucleic Acids Res.*, 1999, **27**, 3836–3843.

95 G. Dianov and J. Parsons, *DNA Repair*, 2007, **6**, 454–460.

96 K. Hitomi, S. Iwai and J. A. Tainer, *DNA Repair*, 2007, **6**, 410–428.

97 D. Wilson III and V. Bohr, *DNA Repair*, 2007, **6**, 544–559.

98 I. V Kovtun and C. Mcmurray, *DNA Repair*, 2007, **6**, 517–529.

99 P. Fortini and E. Dogliotti, *DNA Repair*, 2007, **6**, 398–409.

100 S. S. David, V. L. O'Shea and S. Kundu, *Nature*, 2007, **447**, 941–950.

101 A.-L. Lu, X. Li, P. M. Wright and D. Chang, *Cell Biochem. Biophys.*, 2001, **35**, 141–170.

102 P. Ø. Falnes, A. Klungland and I. Alseth, *Neuroscience*, 2007, **145**, 1222–1232.

103 M. S. Cooke, M. D. Evans, M. Dizdaroglu and J. Lunec, *FASEB J.*, 2016, **17**, 1195–1214.

104 A. Y. Lau, O. D. Scha, L. Samson, G. L. Verdine and T. Ellenberger, *Cell*, 1998, **95**, 249–258.

105 P. Ø. Falnes and T. Rognes, *Res. Microbiol.*, 2003, **154**, 531–538.

106 B. Sedgwick, P. A. Bates, J. Paik, S. C. Jacobs and T. Lindahl, *DNA Repair*, 2006, **6**, 429–442.

107 O. Sundheim, V. A. Talstad, C. B. Vågbø, G. Slupphaug and H. E. Krokan, *DNA Repair*, 2008, **7**, 1916–1923.

108 F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vaagbø, B. Kavli, M. S. Bratlie, J. Peña-Diaz, M. Otterlei, G. Slupphaug and H. E. Krokan, *DNA Repair*, 2004, **3**, 1389–1407.

109 E. A. Mullins, R. Shi, Z. D. Parsons, P. K. Yuen, S. S. David, Y. Igarashi and B. F. Eichman, *Nature*, 2015, **527**, 254–258.

110 D. S. SHIN and J. A. TAINER, *Nature*, 2015, **525**, 8–9.

111 D. K. O'Flaherty, A. Patra, Y. Su, F. P. Guengerich, M. Egli and C. J. Wilds, *Chem. Sci.*, 2016, **7**, 4896–4904.

112 K. Luger and T. J. Richmond, *Curr Opin Struct Biol*, 1998, **8**, 33–40.

113 A. J. Andrews and K. Luger, *Annu. Rev. Biophys.*, 2011, **40**, 99–117.

114 R. K. McGinty and S. Tan, *Chem. Rev.*, 2015, **115**, 2255–2273.

115 M. Biswas, K. Voltz, J. C. Smith and J. Langowski, *PLoS Comput. Biol.*, 2011, **7**, 2–13.

116 A. R. Cutter and J. J. Hayes, *FEBS Lett.*, 2015, **589**, 2914–2922.

117 K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, A. W. M Auml Der, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251–260.

118 C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond, *J. Mol. Biol.*, 2002, **319**, 1097–1113.

119 J. A. Downs, M. C. Nussenzweig and A. Nussenzweig, *Nature*, 2007, **447**, 951–958.

120 B. D. Strahl and C. D. Allis, *Nature*, 2000, **403**, 41–45.

121 T. Jenuwein and C. D. Allis, *Science*, 2001, **293**, 1074–1081.

122 K. Luger, *Chromosom. Res.*, 2006, **14**, 5–16.

123 S. L. Berger, *Nature*, 2007, **447**, 407–412.

124 G. Felsenfeld and M. Groudine, *Nature*, 2003, **421**, 448–453.

125 S. Jones, P. van Heyningen, H. M. Berman and J. M. Thornton, *J. Mol. Biol.*, 1999, **287**, 877–896.

126 N. M. Luscombe, R. a Laskowski and J. M. Thornton, *Nucleic Acids Res.*, 2001, **29**, 2860–2874.

127 N. M. Luscombe and J. M. Thornton, *J. Mol. Biol.*, 2002, **320**, 991–1009.

128 D. Lejeune, N. Delsaux, B. Charloteaux, A. Thomas and R. Brasseur, *Proteins*, 2005, **61**, 258–271.

129 B. Schneider, J. Černý, D. Svozil, P. Čech, J. C. Gelly and A. G. De Brevern, *Nucleic Acids Res.*, 2014, **42**, 3381–3394.

130 R. Sathyapriya, K. V. Brinda and S. Vishveshwara, *J. Chem. Inf. Model.*, 2006, **46**, 123–129.

131 R. Sathyapriya, M. S. Vijayabaskar and S. Vishveshwara, *PLoS Comput. Biol.*, 2008, **4**, e1000170.

132 J. González, I. Baños, I. León, J. Contreras-García, E. J. Cocinero, A. Lesarri, J. A. Fernández and J. Millán, *J. Chem. Theory Comput.*, 2016, **12**, 523–534.

133 R. Chaudret, B. de Courcy, J. Contreras-Garcia, E. Gloaguen, A. Zehnacker-Rentien, M. Mons and J.-P. Piquemal, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9876–9891.

134 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

135 Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.

136 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 1–19.

CHAPTER 2

EXPERIMENTAL AND COMPUTATIONAL METHODS

2. - EXPERIMENTAL AND COMPUTATIONAL METHODS

The spectroscopic techniques used in this thesis required of a complicated set-up which combined several components: laser sources, dye mixtures, pulsed valves, vacuum pumps, synchronization devices, etc. were combined to generate the ideal conditions for the spectroscopic experiments. On the other hand, conformational exploration and data analysis were carried out using several computer programs.

In this chapter, the details of the experimental and computational procedures used along the thesis will be described.

2.1 Experimental methods

2.1.1 Laser sources

The samples studied in this work are difficult to sublimate or evaporate by heating. Therefore, the use of a laser desorption set up was required. Two different laser sources were used for this purpose:

- **Ultra Nd/YAG (Quantel):** 30 mJ/pulse at 1064 nm, working at 10 Hz, 8 ns of pulse duration and 3 mm of beam diameter. This laser was available at UPV/EHU laboratory.

- **Polaris II Nd:YAG laser (New Wave Research Inc.):** 50 mJ/pulse at 1064 nm, working at 10 Hz, 5 ns of pulse duration and 3 mm of beam diameter. This laser was available at FELIX laboratory.

Also, several tunable lasers were combined in different ways to perform the different spectroscopic techniques.

- **YG-980 Nd-YAG laser (Quantel):** 700 mJ/pulse at 532 nm, working at 10 Hz, 8 ns of pulse duration and 10 mm of beam diameter, used to pump a TDL-90 dye laser (Quantel).

- **TDL-90 Dye laser (Quantel):** ~2 mJ/pulse after doubling and mixing, working at 10 Hz, 10 ns of pulse duration and ~0.2 cm^{-1} of beam linewidth. This laser was available at UPV/EHU laboratory.

- **Brilliant B Nd-YAG laser (Quantel):** 400 mJ/pulse at 532 nm or 200 mJ/pulse at 355 nm, working at 10 Hz, 4.6 ns of pulse duration and

8 mm of beam diameter. This laser was used to pump a Scanmate dye laser or Fine Adjustment Pulsare Pro–S dye laser.

- **Scanmate dye laser (Lambda-Physik):** ~2 mJ/pulse after doubling, 10 ns of pulse duration and ~0.15 cm^{-1} of beam linewidth. This laser was available at UPV/EHU laboratory.

- **PULSARE Pro–S Dye laser (Fine Adjustment):** ~2 mJ/pulse after doubling, ~5 ns of pulse duration and ~0.06 cm^{-1} of beam linewidth. This laser was available at UPV/EHU laboratory.

- **Surelite (Continum):** ~800 mJ/pulse at 1064 nm, working at 10 Hz, 6 ns of pulse duration and 6 mm of beam diameter. This laser was used to pump the OPO/OPA laser (LaserVision).

- **OPO/OPA (LaserVision):** ~10 mJ/pulse after doubling, working at 5 Hz, ~5 ns of pulse duration and ~3.5 cm^{-1} of beam linewidth. Available at UPV/EHU laboratory.

- **Quanta-Ray Pulsed Nd:YAG laser (SpectraPhysics):** ~200 mJ/pulse at 532 nm, working at 10 Hz, ~5 ns of pulse duration and 5 mm of beam diameter, used to pump a Radiant Dye Laser.

- **Radiant Dye laser (NarrowScan):** ~2 mJ/pulse after doubling unit, working at 10 Hz, ~5 ns of pulse duration and ~0.1 cm^{-1} of beam linewidth. This laser was also available at FELIX laboratory.

- **Free-Electron Laser:** the radiation of a free-electron laser (FEL) was used during the experiments carried out in FELIX laboratory. The free-electron laser was able to generate ~100 mJ of mid and far-IR wavelengths, between 3600-66 cm^{-1}. It runs at 5/10 Hz and had pulse duration of 10 μs. The spectral linewidth was ~0.5% of the IR frequencies.

2.1.2 Laser dyes

Laser dyes are organic compounds used as a gain media to generate coherent radiation when they are irradiated by a high-power light source. Their structure, characterized by the presence of several conjugated bonds, is designed to present strong absorptions in a range of wavelengths in UV and visible region, overlapping with the emission bands of the most frequently used pump lasers.[1,2] Depending on the nature of the dye laser and the solvent used, different emission wavelengths can be reached, providing a versatile tool to address different regions of the electromagnetic spectra, shown in **Figure 2.1**.[3]

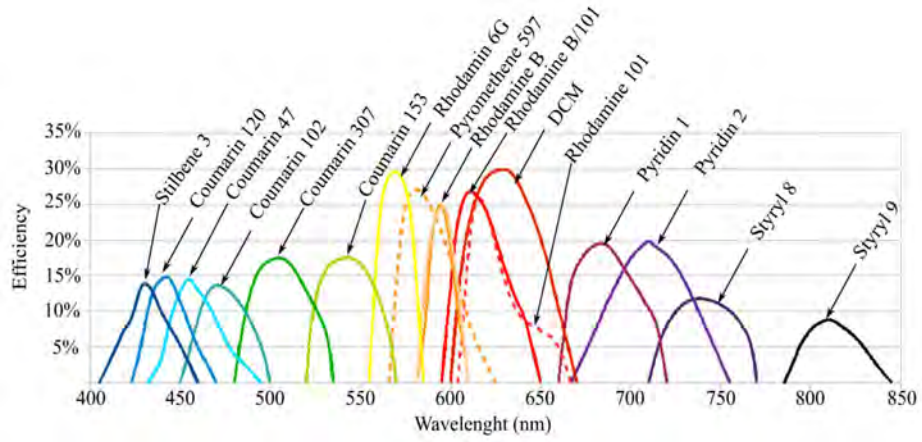


Figure 2.1 *Emission curves of different dyes.*[3]

In this thesis, several dyes were used to cover a broad spectral region. **Table 2.1** collects all the dyes used during this work, with the ideal concentration for the oscillator (C_{osc}) and for the amplificator (C_{amp}).

Table 2.1 *Dyes and dyes mixtures used during this work.*

Dye labels	MW (g mol^{-1})	Pump source (nm)	Solvent	Cosc (M)	Camp (M)	λrange (cm^{-1})
Coumarin 153	309.29	Nd:YAG(355)	MeOH	$2.1\ 10^{-3}$	$4.0\ 10^{-4}$	36000-37000
Rhodamine B	543.02	Nd:YAG(532)	MeOH	$2.1\ 10^{-4}$	$3.1\ 10^{-5}$	34200-33300
Rhodamine B +**Rhodamine 101**	543.02 591.05	Nd:YAG(532)	MeOH	$6.3\ 10^{-5}$ $1.4\ 10^{-5}$	$1.6\ 10^{-5}$ $2.4\ 10^{-6}$	33900-33000
Rhodamine 101	591.05	Nd:YAG(532)	MeOH	$3.6\ 10^{-4}$	$1.9\ 10^{-4}$	33300-32600

2.1.3 Experimental set-up

Two different set-ups were used for mass-resolved excitation spectroscopy. They are based in a time-of-flight (TOF) mass spectrometer equipped with a laser desorption source. Some differences, related to the components of those systems, will be explained.

2.1.3.1 UPV/EHU set-up

Figure 2.2 shows a schematic of the set-up used in the study of the aggregates of DNA bases in the laboratory of the UPV/EHU.[4–8] It was built around a time of flight mass spectrometer equipped with a laser desorption system and contained several components: a source chamber, where the sample was evaporated and picked up by a supersonic jet created with a pulsed valve, a ionization chamber, where the molecules were ionized by a UV laser, and the flight tube of the mass spectrometer, were the different species were separated according to their m/z ratio.

A scheme of the desorption system is shown in **Figure 2.3**. The sample, mixed with carbon nanotubes (Multi-Walled Carbon Nanotubes, purity >90%, 10-30 mm diameter, Sun NanoTech Co Ltd.) was placed on the surface of a cylindrical sample holder (4.5 mm^{ϕ} and 10 mm long) attached to a rotatory axis, actioned by a stepper motor, which was controlled by an external PC. The sample was desorbed using an IR laser (see Section 2.1.1) at the precise time at which the pulsed valve (General valve, Series 9, with a nozzle of 0.5 mm and an aperture time of 160 μs operated at 10 Hz) opened, so the supersonic expansion of Ar (10 bars of typical pressure) picked the ablated material, cooling it and creating the adequate conditions to form the molecular aggregates. The jet entered the

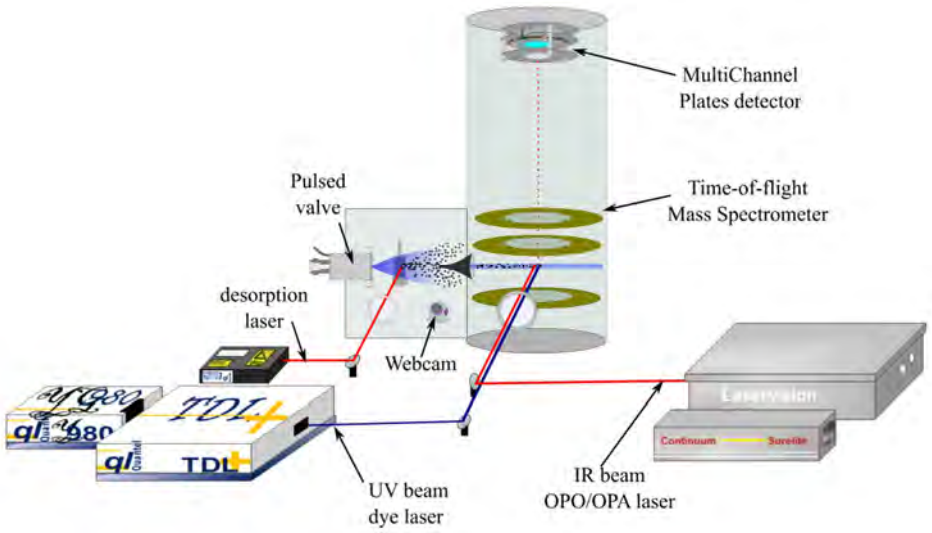


Figure 2.2 *Schematic of the experimental set-up available at UPV/EHU laboratory with some of its component highlighted.*

ionization region of the TOF through a 2 mm skimmer, which selected the coolest region of the beam. Alignment of the IR desorption was monitored during the expansion using a webcam placed inside the desorption chamber. The addition of the web cam may look as a modest improvement, but it was of a great help during the alignment stage, allowing us to obtain superior S/N ratios.

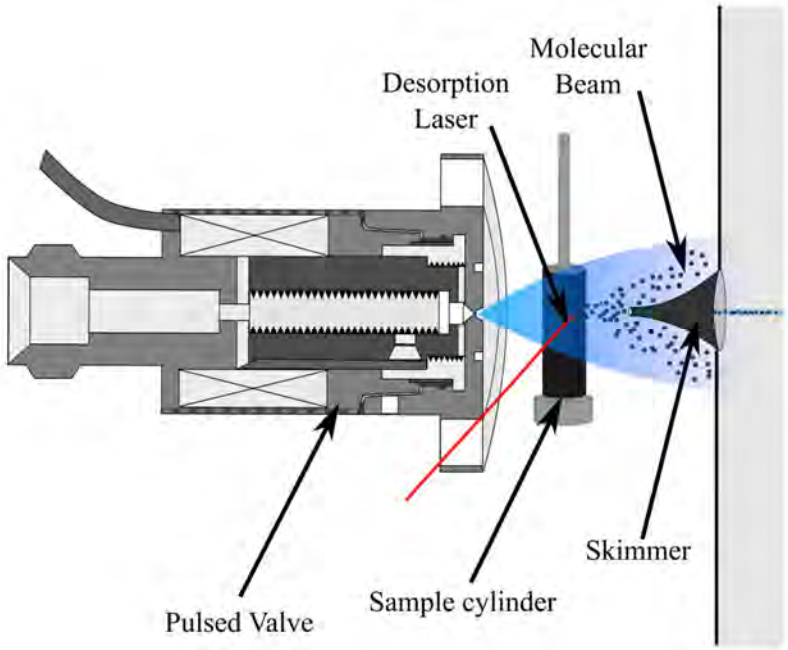


Figure 2.3 *Desorption system of the UPV/EHU set-up.*

Once the molecular jet entered the ionization chamber, it was interrogated using IR and/or UV lasers and either pump/probe or double resonance techniques. Such experimental techniques always involved an excitation/ionization step, using the $S_1 \leftarrow S_0$ // $D_0^+ \leftarrow S_1$ transitions of the chromophore, as the mass spectrometer can only analyze ions. The ions generated were accelerated and sent into the TOF tube, **Figure 2.4**, using the electric fields created by three parallel plates maintained at voltages of ~4000, ~3700 and 0 V (repulsive plate, extraction plate and acceleration plate respectively). Once the acceleration plate were surpassed, the ions entered the free-flight region. All the ions experience a similar force, that depends on their charge and the field applied. At the same time, the force is proportional to their masses. Consequently, lighter ions are accelerated more than heavier ones. During the free-flight region, this differential acceleration results in a physical separation of the ions according to their

m/z ratio, so the lighter ions arrive earlier to the detector, followed by ions with increasing m/z. The detector consisted of two MultiChannel Plates (MCP C-0701, 18 $mm^{Ø}$) and the ions generate a pulsed signal of ~4 mV/ion. The signal was recorded and averaged with the aid of a digital oscilloscope (Tektronics TDS 3032 of 300 MHz or Agillent Infinitum DSO9104A of 1 GHz) and transmitted to a PC for storage and further analysis.

The set-up was maintained under high vacuum (~10^{-5} mbar) during the experiments with the aid of two Leybold Vacuum turbomolecular pumbs: (TurboVac TW 701) of 680 L s^{-1} and (TurboVac 151C) of 150 L s^{-1} and a Varian rotatory Vane Vacuum Pump (D25B).

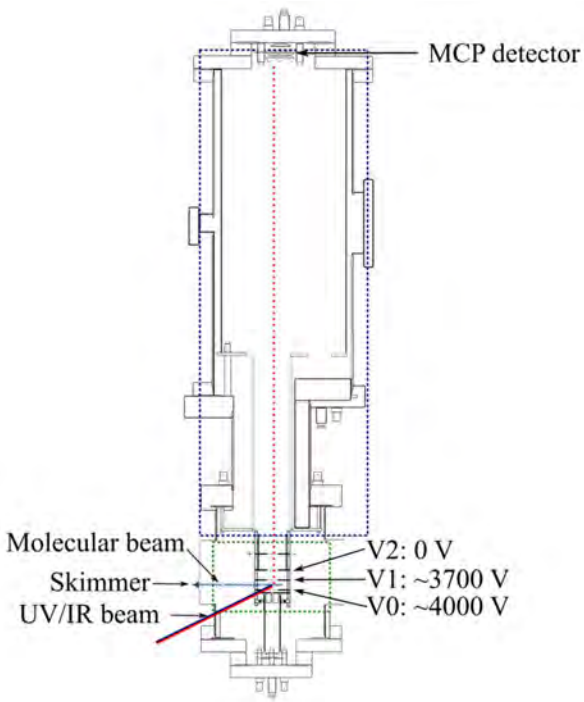


Figure 2.4 *Scheme of the Time-of-flight mass spectrometer of the UPV/EHU set-up. Green dot line depicts the ionization region and blue dot lines depicts the detection region. Red dot line shows the flight of the ions to the detector.*

Finally, all the events during the experiments must happen in a narrow time window. Tight control of the events was achieved using a pulsed Stanford delay generators (DG535). A schematic with the synchronization of the relevant events is shown in **Figure 2.5**.

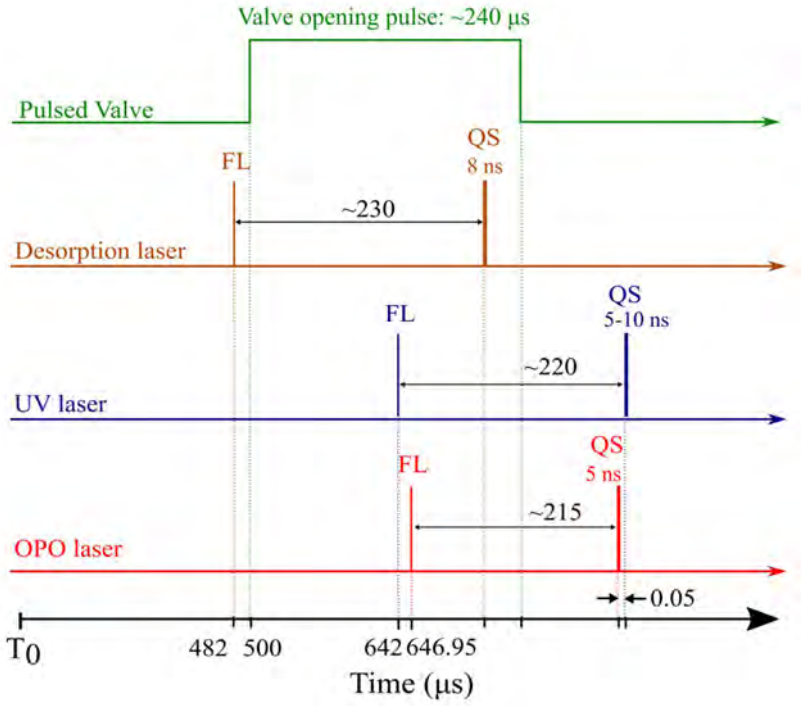


Figure 2.5 *Timeline of the main events during a typical IR/UV experiment using UPV/EHU set-up. FL: flashlamp triggering, QS: qswitch opening. The temporal width of the beams is also indicated.*

2.1.3.2 FELIX set-up:

Figure 2.6 shows a schematic of the set-up used in the FELIX laboratory.[9–11] This set-up was in essence similar to that described in the previous section; however, some differences may be noticed, related to the pulsed valve, the design to the desorption chamber and the time-of-flight mass spectrometer.

The desorption system is depicted in **Figure 2.7**. The sample is mixed with carbon black and placed over the top of a graphite sample bar (50 x 15 x 1 mm), attached to a sample holder moved horizontally by a stepper motor controlled by an external PC. The sample was brought into gas phase by a pulsed IR laser (see section 2.1.1) in vertical configuration. When the pulsed valve (Jordan, 0.5 mm of nozzle and 60 ns of aperture time, working at 10 Hz) created the supersonic jet expansion of Ar (2-3 bar typical pressure), the desorbed material was picked, cooled and carried throughout a skimmer of 2 mm into the ionization chamber.

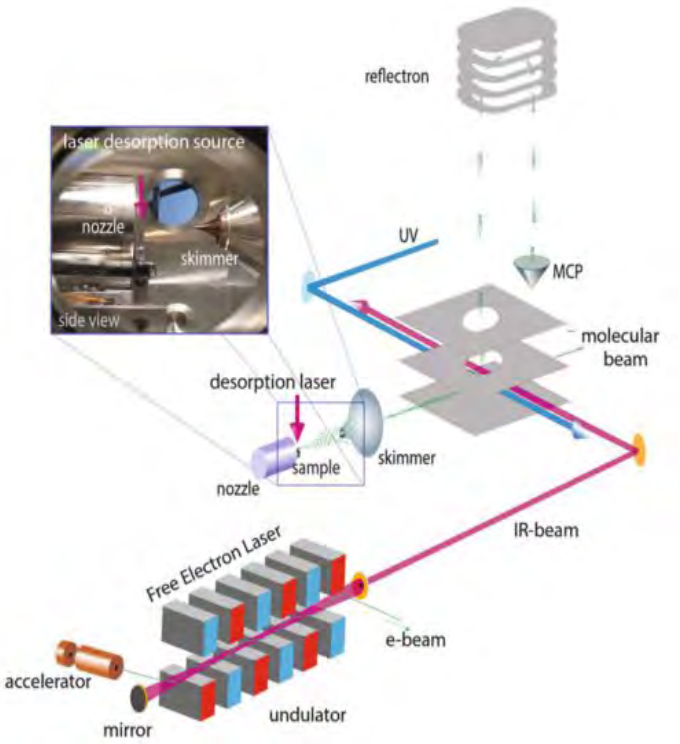


Figure 2.6 *Schematic of the molecular beam set-up, available at FELIX laboratory. It is equipped with a laser desorption source and a reflectron TOF mass spectrometer.*[9]

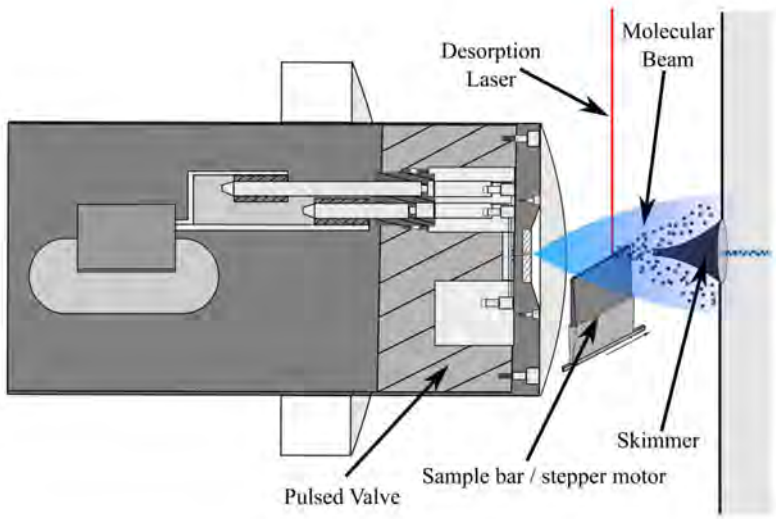


Figure 2.7 *Desorption system in the experimental set-up of the FELIX laboratory.*

After the sample molecules entered the ionization chamber, they were interrogated using UV and/or IR lasers. Once the molecules were photoionized, these ions were accelerated and sent into the TOF tube of a reflectron mass analyzer (D-850 TOF, Jordan Co. **Figure 2.8)**, using three parallel plates whose voltages were maintained at ~3400, ~2650 and 0 V. An extra plate (+/-) generated a voltage of ~160 V to slightly perturb the linear movement of the ions so they arrived with the correct angle to be reflected back and focused into the MCP detector.

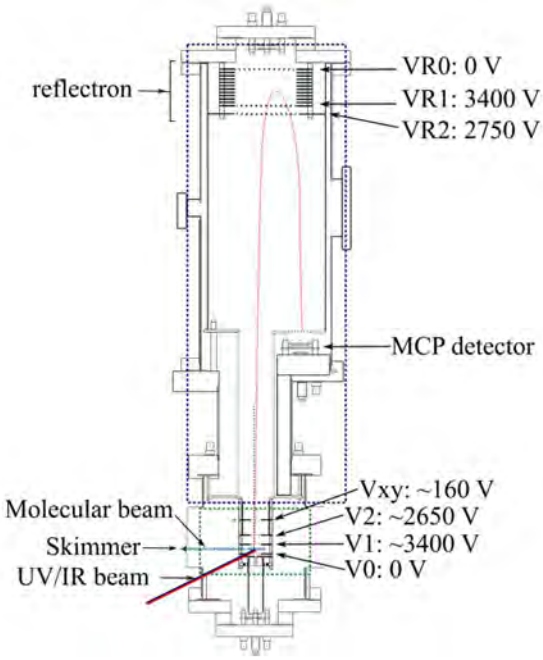


Figure 2.8 *Scheme of the ionization and detection chamber of the FELIX set-up. Green dot line depicts the ionization region and blue dot lines depicts the detection region. Red dot line shows the flight of the ions to the detector.*

The purpose of the reflection field is to extend the flight distance and to increase mass resolution. The field acts a spring. Therefore, the ions of the same m/z that are slightly faster carry higher kinetic energy and penetrate deeper in the field and therefore are reflected with some delay, while slow ions penetrate less and exit faster. In this way, the initial velocity distribution of the ions is reduced and the mass resolution, defined as m/Δm, increases. Three plates at voltages of 0, 2750 and 3400 V create the reflecting field to focus the ions back into the microchannel

plate detector (MCP C726, 40 mm, Jordan Co.) and the signal generated was routed into a digital oscilloscope (Tektronkix 3054B) or a digital data acquisition card (National Instruments).

The three chambers were maintained under high vacuum by two turbopumps (Pfeiffer TMH 261) of 210 L/s for the ionization and detection chamber (1.3 x 10^{-7} and 7.1 x 10^{-8} mbar respectively), and another turbopump (Pfeiffer TPH 1201 UP) of 1250 L/s for the desorption chamber (1.7 x 10^{-7} mbar). Furthermore, Delay generators (Stanford DG645/DG535) were used to tightly control all the events. A schematic with the synchronization of all the relevant events is shown in **Figure 2.9**.

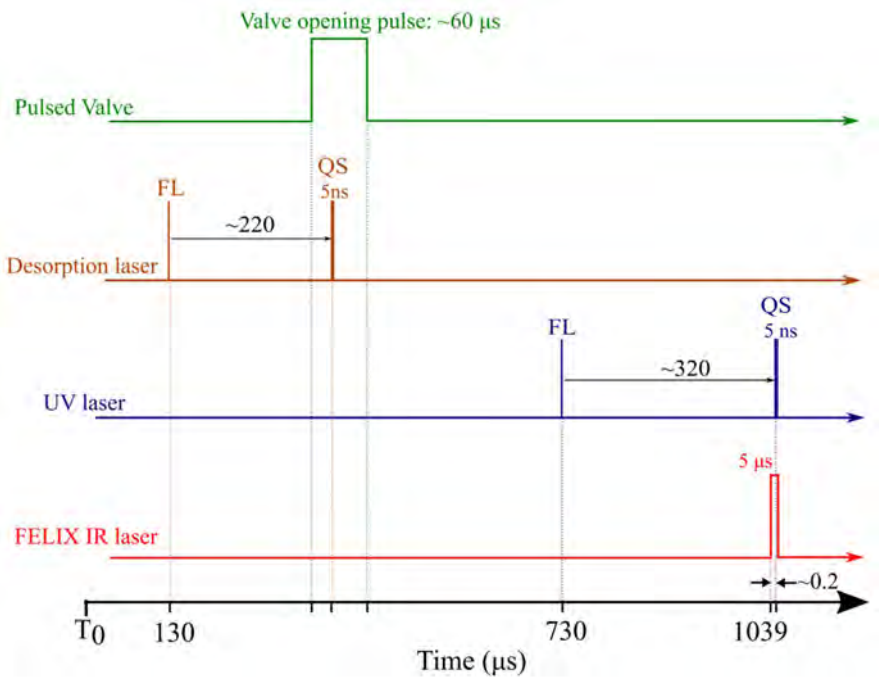


Figure 2.9 *Timeline of the main events synchronized in a typical IR/UV experiment at Felix set-up. FL: flashlamp triggering, QS: qswitch opening. The temporal width of the beams is also indicated.*

2.2 Computational methods

Different software packages were used during this work for the treatment of the results, conformational analysis and quantum chemical calculations.

All the calculations were carried out using the computing resources available in the Scientific Computing Service (IZO-SGI) of the University of the Basque Country. This service has 1926 cores (1774 xenon cores, 128 Itanium2 cores and 24 Opteron cores) with RAM ranging from 16 to 512 GB per node and the connection to the service is made using a Virtual Private Network (VPN).

2.2.1 Conformational analysis and structural predictions

Several programs were used for the analysis of the conformational landscape and structural predictions.

- **Schrödinger suite:** is a computational chemistry package that includes several applications used in the analysis of the conformational landscape of the molecules and molecular aggregates.[12] This suite includes:

1. Maestro: is a full-featured molecular visualization environment. It is used in combination with Macromodel tool as an interface for building and visualizing 3-dimensional chemical models. Three versions were used during this work: *9.0*, *9.2* and *9.3*.[13]

2. MacroModel: is a force-field-based molecular modeling program with applicability to a wide range of chemical systems. It can be used for the understanding of chemical structure, energetics, and dynamics of systems ranging from small molecules to proteins. A wide collection of methods is available for conformational searching, allowing for an efficient sampling of the Potential Energy Surface (PES) of low-energy structures. This program was used to explore the conformational landscape of the systems, obtaining hundreds or even thousands of possible conformers of each system.[14]

3. XCluster: is a MacroModel tool for the analysis and visualization of the clustering of molecular conformations, classifying hundreds or even thousands of conformers into a few families of structures using molecular similarity criteria.[15]

- **Gaussian:** is a computational chemistry software to carry out quantum chemical calculations. Starting from the fundamental laws of quantum mechanics, it predicts the energies, molecular structures, vibrational frequencies and molecular properties of molecules and reactions in a wide variety of chemical environments. This program was used to perform structural optimizations and normal-mode analysis on the

different conformational isomers of the studied system. The version used was Gaussian09 D01.[16]

- **NCIplot:** is a program for plotting non-covalent interactions regions, classifying all the non-covalent interactions using a color code.[17,18] It was used during the analysis of the intra- and inter-molecular interactions.

2.2.2 Data analysis

Three different programs were used during this work for data analysis.

- **Excel (Microsoft):** this program, included in Microsoft Office suite, was used for the analysis of experimental and theoretical data.

- **Origin (Microcal):** is a program for the data analysis and a graphical representation of the results.[19]

- **GNUplot**: is a portable command-line driven graphing utility to visualize mathematical functions and data interactively. It was used for the construction of 2-D NCIplot diagrams.[20]

2.2.3 Graphical representation

For the graphical representation of the structures, two different programs were used.

- **Chemcraft:** is a graphical program designed by G. A. Andrienko[21] to visualize the results from quantum chemistry computations. It is a convenient tool for visualizing the Gaussian outputs and preparing new jobs for calculation.

- **VMD**: is a molecular visualization program designed by the Theoretical and Computational Biophysics Group (University of Illinois at Urbana-Champaign) for displaying, animating, and analyzing large biomolecular systems using 3-D graphics.[22,23] It was used for the construction of 3-D NCIplots diagrams.

2.3 References:

1 F. P. Schäfer, in *Dye Lasers*, ed. F. P. Schäfer, Springer Berlin Heidelberg, Berlin, Heidelberg, 1973, pp. 1–85.

2 K. H. Drexhage, in *Dye Lasers*, ed. F. P. Schäfer, Springer Berlin Heidelberg, Berlin, Heidelberg, 1973, pp. 144–193.

3 http://www.radiant-com/index.php/products/las.

4 I. Leon, E. J. Cocinero, J. Millán, S. Jaeqx, A. M. Rijs, A. Lesarri, F. Castaño and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2012, **14**, 4398–4409.

5 I. Usabiaga, J. González, P. F. Arnáiz, I. León, E. J. Cocinero and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12457–12465.

6 A. Longarte, C. Redondo, J. A. Fernández and F. Castaño, *J. Chem. Phys.*, 2005, **122**, 1643041–11.

7 J. Gonzalez, I. Usabiaga, P. F. Arnaiz, I. Leon, R. Martínez, J. Millán and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2017, (DOI: 10.1039/C6CP08476A).

8 I. Usabiaga, J. González, I. León, P. F. Arnaiz, E. J. Cocinero and J. A. Fernández, *J. Phys. Chem. Lett.*, 2017, **8**, 1147–1151.

9 A. M. Rijs and J. Oomens, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Springer International Publishing, Cham, 1st edn., 2015.

10 S. Jaeqx, W. Du, E. J. Meijer, J. Oomens and A. M. Rijs, *J. Phys. Chem. A*, 2013, **117**, 1216–1227.

11 A. M. Rijs, E. R. Kay, D. A. Leigh and W. J. Buma, *J. Phys. Chem. A*, 2011, **115**, 9669–9675.

12 www.schrodinger.com.

13 Maestro 9.0 UserManual, 2009.

14 MacroModel user manual, 2013.

15 XCluster Macromodel Manual, 2012.

16 www.gaussian.com.

17 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625–632.

18 http://www.lct.jussieu.fr/pagesperso/contrera.

19 http://www.originlab.com/Origin.

20 http://gnuplot.sourceforge.net.

21 http://www.chemcraftprog.com.

22 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 27–38.

23 http://www.ks.uiuc.edu/Research/vmd.

CHAPTER 3

METHOLODOGY

3. - METHODOLOGY

3.1 Experimental techniques

In this section, the different mass-resolved spectroscopic techniques used during this work will be described. These techniques are based in multiphoton ionization processes (MPI),[1–3] in which a molecule absorbs more than one resonant photons to be ionized:

$$Y + n^* h\nu \longrightarrow Y^+ + e^-$$

where Y and Y^+ are the neutral and the ion molecule respectively. The photons can come from a single laser beam passing through the sample or may be provided by two or more lasers beams, giving rise to different spectroscopic techniques.

Once the molecules are ionized, they are accelerated in the mass spectrometer, being separated and selected by their m/z ratio and detected in the multichannel plate detector.

During this work, two different spectroscopic techniques, REMPI and IR/UV were used as explained below. Additional information may be found in refs. 3–7.

3.1.1 REMPI spectroscopy

Resonant Enhanced MultiPhoton Ionization (REMPI) spectroscopy is a powerful technique to ionize molecules and study their electronic structure. This technique is a modification of Laser Induced Fluorescence (LIF) spectroscopy that incorporates a mass-selective detection. The combination with mass spectrometry allows addressing complex systems by discriminating different species according to their m/z ratio.

In the REMPI spectroscopy, a molecule is ionized from the ground state (S_0) via an intermediate excited state (S_1) by the absorption of more than one photon. Depending on the number of photons absorbed to reach the excited state (n photons) and the ion required (m photons), a (n+m) REMPI spectroscopy is applied. Furthermore, the photons can be generated by one or more lasers, depending on the energetic requirements of the sample: one-color process when the photons are isoenergetic (λ_n =

λ_n) and come from the same laser source; two-color spectroscopy when the photons have different wavelengths ($\lambda_n \neq \lambda_n$) and come from two different laser sources, etc. A schematic of the different REMPI processes is shown in **Figure 3.1**. **Figure 3.1a** shows the excitation scheme for one-color, two photons (1+1) REMPI, while **Figure 3.1b** depicts a one-color, 2+1 process. **Figure 3.1c** represents a two-color process. The latter technique is useful for those samples in which fragmentation congests the spectrum. During this work, only one-color REMPI spectroscopy (1+1) has been used.

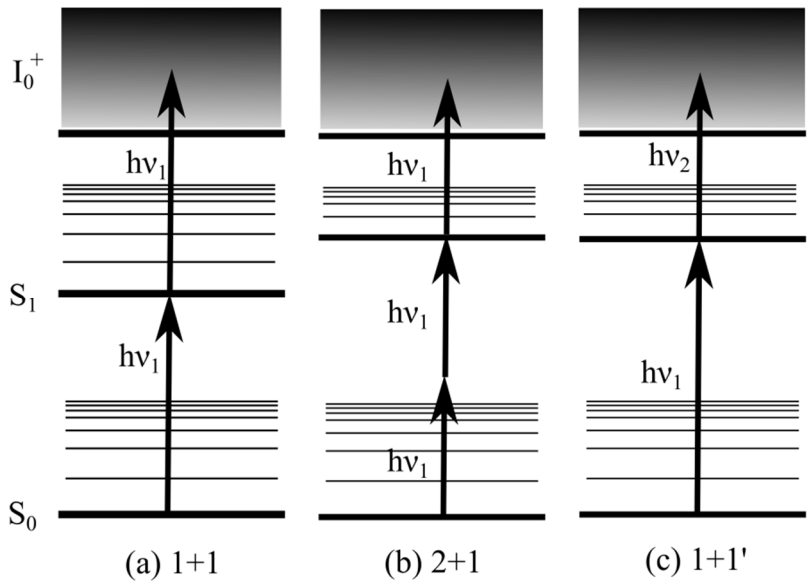


Figure 3.1 *Representation of different REMPI processes.* ***(a)*** *One-color (1+1)* ***(b)*** *One-color (2+1)* ***(c)*** *Two-color (1+1').*

In (1+1) REMPI spectroscopy, the molecule reaches the excited state with the first photon and it is ionized by a second photon from the same laser. A UV-chromophore in the molecule of interest is required to observe the UV transitions. Several conditions need to be accomplished to produce REMPI absorption. On the one hand, a resonant excited state is necessary for the multiphoton process and the ionization, but not sufficient, although a real S_1 excited state enhances the probability of the excitation process, resulting in significantly stronger signal. On the other hand, the excited state should be sufficiently long-lived (in the order of nanoseconds) for the absorption of the second photon to take place. Thus,

when these conditions are met, the UV spectrum of the $S_1 \leftarrow S_0$ transition of the neutral structure is recorded as a function of the excitation wavelength. **Figure 3.2** shows a scheme of one-color REMPI absorptions which accomplish or not all the conditions.[1] While ν_1 and ν_3 do not produce any detectable signal in the spectrum, ν_2 reaches a resonant excited state which produces a high-enough intensity to be measurable.

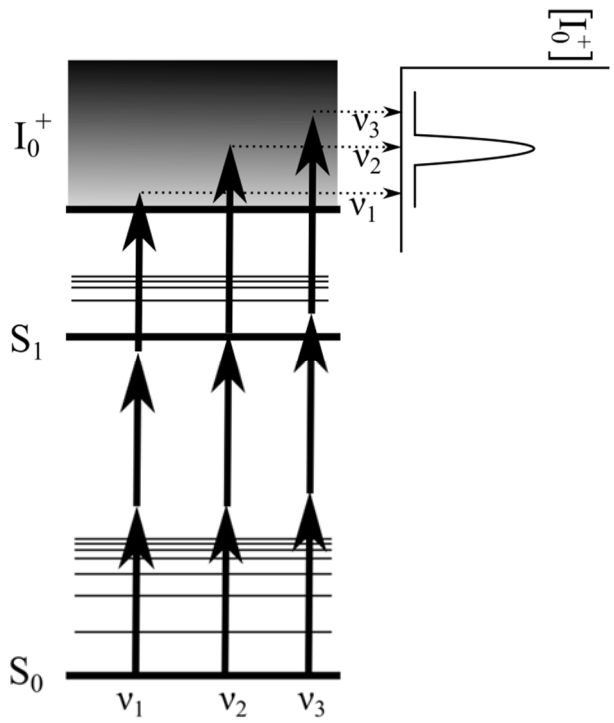


Figure 3.2 *Representation of one-color (2+1) REMPI absorptions. ν_1 and ν_3 photons are not resonant absorptions and ν_2 photon corresponds to a resonant absorption.*

3.1.2 IR/UV spectroscopy

IR/UV spectroscopy is a double resonance technique designed to obtain the mass-resolved vibrational spectrum of the molecules. As a double resonance technique, two different laser beams are used: the depopulation of the ground state of a molecule is induced with a first laser (burn laser) and a second laser is used to probe the desired region (probe laser).

Thus, in the IR/UV ion dip technique, the UV beam used as a probe laser is fixed on a transition of the REMPI spectrum, in order to obtain a constant signal. The tunable IR laser beam is fired a few nanoseconds prior to the UV pulse and is used to scan the IR region. When the IR laser hits a resonant transition, the ground state is depopulated and the signal from the probe laser decreases. **Figure 3.3** shows a schematic of the IR/UV process.

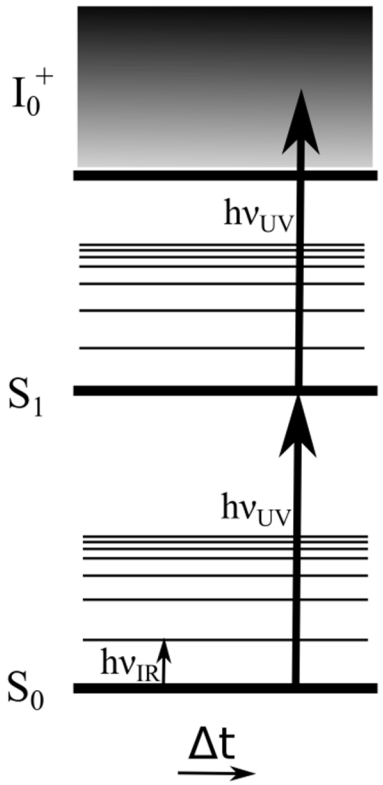


Figure 3.3 *Representation of the IR/UV ion dip process.*

It is worthy to mention that this is not a background-free technique but a differential technique in which the final spectrum is obtained as the difference between REMPI signals with and w/o the IR laser. Thus, the IR laser runs at 5 Hz whereas the UV laser runs at 10 Hz, allowing for a real-time subtraction. Furthermore, the correction of the baseline is carried out by monitoring the fluctuations of the baseline. **Figure 3.4** shows the raw signal obtained with/without depletion and the subtraction of the signals.

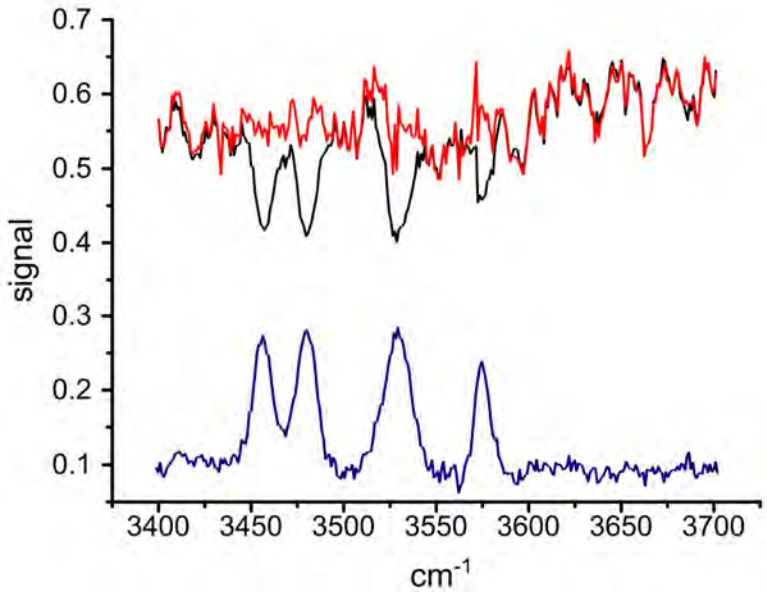


Figure 3.4 *Schematic of the signal recorded with (black trace) and without (red trace) depletion by the absorption of the IR beam. The blue trace represents the substracction of both signals with the IR peaks observed.*

3.2 Computational chemistry

Computational chemistry is a branch of chemistry that uses the computer resources to tackle chemical problems. It uses mathematical algorithms and the physical and chemical fundaments to create models to shed light on empirical phenomena. Thanks to the advances in computing in the last decades, systems of increasing complexity can be addressed through theoretical calculations, providing information about the physical and chemical properties of the systems: potential energy surfaces, geometries, vibrational modes, intermolecular and intramolecular forces, etc.[2,8,9] Thus, computational chemistry is a powerful tool for researchers to explain the experimental observations.

During this work, several theoretical methods have been used to explore the conformational landscape of the systems, predict their vibrational spectra and analyze different physical properties. In this section, all the different theoretical methodologies will be explained. As in the previous section, additional information can be found in refs. 8–13.

3.2.1 Molecular Mechanics

Molecular Mechanics (MM) is a useful tool to carry out conformational searches and automatic explorations of potential energy surface of large and complicated systems with a low computational cost. It is based on the laws of classical physics to describe the potential energy surface and divides the system in pairs of atom-atom interactions, angles and dihedrals throughout a series of equations and parameters called Force Field. In a force field, the interactions between atoms are replaced by equations that describe the behavior of similar chemical groups in other molecules. These equations contain certain adjustable parameters that are filled using crystallographic data. Thus, the total energy is expressed as the sum of the list of terms collected in **Equation 3.1**. These terms can be divided in three groups: **bonded terms** describe the behavior of chemical bonds and the movement of their members (E_{bond}, E_{angle}, $E_{out\text{-}of\text{-}plane}$ and $E_{torsion}$); **non-bonded terms**, related to the interactions between non-bonded atoms (E_{vdW} and $E_{electrostatic}$); and **crossing terms**, produced by the influence of the other atoms and the chemical groups in the vicinity into the bonds, angles and dihedrals (E_{cross}: $E_{stretch\text{-}bend}$, $E_{stretch\text{-}torsion}$, etc).

$$E = E_{bond} + E_{angle} + E_{out\text{-}of\text{-}plane} + E_{torsion} + E_{vdW} + E_{electrostatic} + E_{cross} \quad \textbf{(3.1)}$$

The different terms that form the exact functional depend on the force field. Thus, force fields are classified in three different types:

- First class: they consider that valence terms are isolated and do not have crossing terms. Furthermore, torsion and bond terms are parametrized as harmonic potentials. For example, AMBER or CHARMM are first class force fields.

- Second class: they take into account cubical and higher terms to describe torsion and crossing terms. MM3 and CFF93 are examples of second class force fields.

- Third class: they are second class force fields that also include chemical effects as hyperconjugation or electronegativity. MM4 and MMFFs are examples of this kind of force fields.

Thus, the selection of the adequate force field is critical in the performance of the conformational analysis. The force field selected for this work is the Merck-Molecular Force Field (MMFFs), with a

modification for energy minimization studies.[14–20] This force field can be applied to a wide collection of molecules, as it was parameterized to describe accurately conformational and non-bonded interactions and it is appropriate to threat gas phase systems without solvent.

The conformational search methodology is a combination of several methods: Monte-Carlo (MC) algorithm, Mixed-Torsional Large scale Low Mode Sampling (MT-LLMOD) and Polak-Ribiere Conjugate Gradient (PRCG) methodology.[21] The MC algorithm is a probabilistic mechanism that generates a set of structures by the progressive variation of bond distances, angles and dihedrals in order to find all the possible conformers. After that, MT-LLMOD method selects this set of structures and perturbs them in the direction of one of the low-frequency normal modes to look for additional low energy conformation. Finally, those perturbed structures are energy minimized using the Polak-Ribiere first derivatives method and added to the list if the energy difference with the most stable conformer is within the selected energy window.

It is worthy to mention that the large amount of structures found in the conformational search (sometimes hundreds or thousands of structures with minimum geometrical differences) needs to be classified into families before applying quantum mechanics calculations. Such classification is carried out using the full distance matrix and three-dimensional superposition.[22] With this procedure, the number of structures is reduced from thousands to less than one hundred families. So, representative members of each family may be subjected to full optimization.

3.2.2 Quantum Mechanics

Quantum Mechanics (QM) calculations are used to obtain a quantitative description of the molecular properties of the structures obtained by Molecular Mechanics. The goal of the QM calculations is to solve the Schrödinger equation, **Equation 3.2**, in order to obtain a precise description of the electronic structure of the system. However, it is not possible to find analytical solutions of this equation for atoms or molecules bigger than hydrogen-like systems and approximate solutions may be found instead. These solutions involve the calculation of a large number of integrals and equations, providing several methods which can be classified into three groups: semi-empirical methods, *ab-initio* methods and density functional theory (DFT) methods. During this work, *ab-initio*

and DFT methods were tested in order to obtain reliable results at a reasonable computational resources cost.

$$\hat{H}\,\Psi = E\,\Psi \qquad \textbf{(3.2)}$$

Ab-initio methods describe the molecular structure based on the approximation of the electronic wavefunction throughout theoretical principles and fundamental constants, without using any experimental data. Thus, *ab-initio* calculations include several different methods: Hartree-Fock (HF), Multi-Configurations Self Consistent Field (MCSCF), Configuration Interaction (CI), Coupled Cluster (CC), etc. One of the most broadly used *ab-initio* method is the Moller-Plesset perturbation theory (MPn), a post-HF method which takes into account the effects of electronic correlation using the perturbation theory. Thus, second order Moller-Plesset (MP2) calculations were extensively used in quantum chemistry to describe the structure of small molecules. However, these *ab-initio* methods are computationally expensive for large systems and other methodologies were addressed.

DFT calculations include electronic correlation effects throughout electron density functionals. The formulation of the Hohenberg-Kohn theorem in 1964 laid the foundations of the DFT methodology.[23,24] This theorem established the existence of a functional to define accurately the density and energy of the ground state, although they did not provide the configuration of this functional. Thus, several functionals have appeared during the last years with different analytical forms. Perhaps, the most remarkable difference between them is the parametrization of the part of the functional that describes the interactions due to dispersive forces.

Thus, the large size of the systems tackled in this work makes DFT calculations the method of choice to meet both criteria: accuracy and affordable computing cost. In fact, two different DFT hybrid methods were mainly used: B3LYP and M06-2X.

B3LYP is the gradient-corrected exchange-correlation density functional developed by Becke in 1993.[25–28] This functional is the first example of hybrid functional that combines HF and DFT theory thanks to the incorporation of the Becke's three-parameters (B3) to the LYP functional, developed by Lee, Yang and Park in 1988. This functional was commonly used in chemistry in the last twenty years and was considered to describe successfully numerous systems. However, B3LYP functional does not take into account dispersive contribution, not describing accurately those systems in which dispersive forces are prevailing, such as

stacking systems. Nowadays, its use has been restricted to serve as comparison with other functionals that take into account dispersive interactions.

Several years later, Thrular developed Minnesota functionals M05 and M06 which take into account the dispersive contribution.[29,30] M06-2X is a meta-hybrid density functional, parametrized only for non-metallic atoms, and that is appropriate to address systems in which non-covalent interactions are predominant.[31] During this work, M06-2X has been selected to perform geometrical optimizations and normal-mode analysis of the structures to obtain the minima in the potential energy surface. The theoretical IR spectrum of the structures was also extracted from the normal-mode analysis to compare with the experimental results.

Recently, Grimme developed an empirical method to account for dispersive interactions and implemented this correction for several functionals, including B3LYP. By convention, these correlation corrected DFTs are labelled with a –D, as DFT-D methods.[32,33] For example, B3LYP-D is the correlation corrected version of B3LYP. During this work, Grimme's D3 correction has been used with the B3LYP-D functional in order for comparison with M06-2X and B3LYP results.

3.2.3 Basis set

Basis set is a group of mathematical functions whose combination is able to describe the molecular orbitals of the systems. In this work, double and triple ζ Pople's basis set were used in all the calculations.[34–37] Thus, 6-31+G and 6-311++G(d,p) were used for optimization and normal-mode analysis of the structures with the previously mentioned functionals. The notation of the basis set may be found elsewhere. For example, for 6-311++G(d,p) means:

- 6 Gaussian functions are used to describe the inner-shell orbitals (core orbitals).
- The hypen (-) denotes a split-valence set.
- 31 (double ζ) and 311 (triple ζ) are related with the number of Gaussian functions used to describe the valence shell. Double ζ implies two basis functions for each atomic orbital and triple ζ implies three basis functions for each atomic orbital.
- (d,p) refer to the use of polarized functions. They add a set of p orbitals to H and d orbitals to heavy atoms.

- The plus signs (++) mean that diffuse functions (functions with small exponents that extend to regions of the space very distant to the nuclei) will be used in H and heavy atoms.

3.2.4 Binding energy calculation

Binding energy is the energy required to dissociate a bond or a molecular aggregate into its components. The calculation of the binding energy of an aggregate is obtained by subtracting the energy of the components from the total energy of the complex. However, some corrections need to be done to the calculation of the resulting energy values.

The first correction is due to the basis set superposition error (BSSE). This error is produced due to the higher number of basis set used for the complex than for the description of its components. This leads to an artificial extra stabilization of the dimer compared to the monomers. The procedure to solve this error was developed by Boys and Bernardi in 1970 and it is called the Counterpoise Procedure,[38] **Equation 3.3**.

$$\Delta E^{CP} = BSSE = -E_A(AB) - E_B(AB) + E_A^*(AB) + E_B^*(AB) \qquad \textbf{(3.3)}$$

where $E_A(AB)$ and $E_B(AB)$ are the energy of the monomers in the configuration adopted in the dimer and $E_A^*(AB)$ and $E_B^*(AB)$ represent the energy of the monomers A and B computed with the dimer basis set. This procedure is translated in the BSSE correction and the value, positive by definition, should be subtracted from the total binding energy.

The second correction is the zero-point energy (ZPE) correction, and accounts for the vibrational energy storage in any molecule even at 0 K, within the harmonic oscillator approximation. It is a direct consequence of the Heisenberg uncertainty principle. This correction must be applied both to the complex and to its components.

Thus, the final expression for the binding energy calculation (E_{BE}) is collected in **Equation 3.4** subtracting the BSSE correction to the energy of the dimer, with the zero point correction (E_{AB}^0) and it has been used for the calculation of the binding energies during this work.

$$E_{BE} = E_{AB}^0 - BSSE \qquad \textbf{(3.4)}$$

3.2.5 Gibbs free energy diagrams

Entropy (S) and Gibbs free energy (G) provide information about the energetics and the behavior of the systems with the temperature. These two magnitudes are related with the enthalpy (H) by the **Equation 3.5**.

$$\Delta G = \Delta H - T \cdot \Delta S \quad \textbf{(3.5)}$$

The study of the relative Gibbs free energy with the temperature can provide information about the stability of the different structures of the same molecule and the temperature in which an aggregate becomes unstable. The representation of these two magnitudes generates diagrams which easily show which conformers are more likely to appear in a range of temperature. Thus, the values of the Gaussian outputs and the tools supplied by the National Institute of Standards and Technology (NIST),[39] which provides the values of the variation of entropy (ΔS) and enthalpy (E_{ddH}) with the temperature, were used to construct these diagrams for isolated molecules, using the **Equation 3.6** and the stability of the species (E_{Elec}), ZPE corrected.[40,41]

$$\Delta G_{monomer} = [E_{Elec} + ZPE + E_{ddH}] - T\,\Delta S \quad \textbf{(3.6)}$$

When the Gibbs free energy of a complex ($\Delta G_{complex}$) is taken into account, BSSE correction must be also included (E_{BSSE}). The relative Gibbs free energy of the aggregate (ΔG) is then calculated by the **Equations 3.7** and **3.8**.

$$\Delta G_{complex} = [E_{Elec} + ZPE + E_{ddH+} E_{BSSE}] - T\,\Delta S \quad \textbf{(3.7)}$$

$$\Delta G = \Delta G_{complex} - \sum \Delta G_{monomer} \quad \textbf{(3.8)}$$

It is worthy to mention that the calculation of ΔG assumes that the system is in thermodynamic equilibrium to allow one to compare with biological systems. However, the experimental conditions in the gas expansion, where the rotational temperature is ~3-5 K and vibrational temperature is ~100-150 K, are not in equilibrium. A similar calculation may be found in the **Appendix 8.3** for cytosine systems, assuming a rotational temperature of 4 K, without meaningful differences to the ΔG

diagrams calculated as previously expressed. Therefore, such calculation demonstrates that the vibrational entropic term is the main source of differences between the entropy of the different isomers for a given stoichiometry.

3.2.6 Non-covalent interaction approach:

The Non-Covalent Interaction (NCI) approach[42–46] evaluates the intermolecular interactions, based on the behavior of the electron density (ρ) and its reduced density gradient (s), **Equation 3.9**, where $C_F = 2\,(3\pi^2)^{1/3}$ is a constant.

$$s = \frac{|\nabla\rho|}{c_F\,\rho^{4/3}} \qquad \textbf{(3.9)}$$

The technique focuses on regions far from the nuclei where the electron density is perturbed by the interaction between nearby groups, which produces a change in its reduced density gradient, generating a density critical point between them. In the analysis of weak interactions, NCI approach places an emphasis on those regions where electron density is weak and its reduced density gradient is close to zero. However, the representation of the electron density vs. its reduced gradient shows those critical points as troughs, not distinguishing between attractive and repulsive interactions. Thus, in order to distinguish those interactions, the sign of the second eigenvalue of the Hessian matrix of the electron density (λ_2) is used to generate the 2D NCI plots. In these 2D plots, the reduced gradient (s) is plotted as a function of the electron density (ρ), multiplied by the sign of the λ_2. Thus, the critical points are separated according to the topology of the electron density and the nature of the interactions:

- Strong stabilizing interactions, such as hydrogen bonds, appear when $\rho > 0.01$ a.u. and $\lambda_2 < 0$.

- Strong destabilizing interactions, such as steric crowding, appear when $\rho > 0.01$ a.u. and $\lambda_2 > 0$.

- Delocalized weak interactions, such as van der Waals interactions, appear when $\rho < 0.01$ a.u. and $\lambda_2 \sim 0$.

3-D NCI plots are the spatial representation of the non-covalent interactions obtained in the 2D plots. Thus, these plots generate isosurfaces of the reduced gradient of density between the atoms, showing all the non-covalent interactions and classifying them by BGR color code:

Blue for stabilizing interactions, Green for weak interactions and Red for destabilizing interactions. Thus, the intensity of the color tone is related with the strength of the interactions. For example, hydrogen bonds depicted in deep blue are more intense that those ones depicted in light blue. These isosurfaces are generated by the analysis of local properties on a cubic grid constructed within the NCIPlot program.[43] This grid is constructed with the geometry of the molecule, with a threshold of ±2 Å from the outermost x,y,z coordinates to ensure that the isosurface is contained in the cube, and the increments 0.1 a.u. in each direction, generating a grid of 0.1 x 0.1 x 0.1 a.u.

This approach was used during this work to evaluate the nature of the non-covalent interactions between DNA base – amino-acid aggregates[46] and DNA base aggregates.[47] Furthermore, the intramolecular interactions of amino acids were also studied with this methodology.

3.2.7 Hydrogen bond considerations:

The hydrogen bond term has evolved over the years since Latimer and Rodebush mentioned it for the first time in 1920.[48,49] From this moment, the evolution of the understanding about hydrogen bonding has been huge and scientists feel that there is no an universal and accepted definition for the hydrogen bond.[50] In 2011, the IUPAC stablished some criteria to define an interaction as a hydrogen bond.[51] In this thesis, the IUPAC criteria were followed, but making some additional considerations.

As the IUPAC states, the hydrogen bond is "an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom (Y) or a group of atoms in the same or a different molecule, in which there is evidence of bond formation." [50,51] No maximum length between Y···H atoms was indicated in the IUPAC criteria. In this work, the maximum distance to consider an interaction as hydrogen bond is 2.200 Å. Larger interactions were considered as dispersive forces. Moreover, the IUPAC criteria about X-H···Y angle for hydrogen bonds ranges from 180º to 110º. Due to this range is wide, in this thesis the maximum angle to consider an interaction as a hydrogen bond is 140º. Angles lower than 140º were considered due to dispersive forces. However despite these two considerations, the analysis of the electron density obtained from the NCI approach also help us to elucidate the nature of the different interactions investigated and to confirm their characterization as hydrogen bond or dispersive forces.

3.3 References:

1 A. Requena and J. Zuñiga, *Espectroscopía*, Pearson Educación, Madrid, 4th edn., 2004.

2 I. León, PhD Thesis, University of the Basque Country, Leioa, 2011.

3 W. Demtröder, *Laser Spectroscopy*, Springer-Verlag, Berlin, 4th edn., 2002.

4 J. M. Hollas, *Modern Spectroscopy*, John Wiley & Sons, Reading, 4th edn., 2004.

5 A. M. Rijs and J. Oomens, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Springer International Publishing, Cham, 1st edn., 2015.

6 G. Tranter, J. Holmes and J. Lindon, *Encyclopedia of Spectroscopy and Spectrometry, Part 2 (M-Z)*, Academic Press, London, 1st edn., 2000.

7 H. H. Telle, A. G. Ureña and R. J. Donovan, *Laser Chemistry: Spectroscopy, Dynamics and Applications*, John Wiley & Sons, Ltd, Chichester, 1st edn., 2007.

8 J. Leszczynski, *Handbook of Computational Chemistry*, Springer Netherlands, Dordrecht, 1st edn., 2012.

9 J. B. Foresman and A. Frisch, *Exploring Chemistry With Electronic Structure Methods*, Gaussian, Inc., Pittsburgh, 2nd edn., 1996.

10 J. Andrés and J. Beltrán, *Química teórica y computacional*, Universitat Jaume I, Castellón de la Plana, 4th edn., 2001.

11 A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, Dover Publications Inc., Mineola, 1st edn., 1982.

12 D. B. Cook, *Handbook of computational quantum chemistry*, Dover Publications Inc., Mineola, 2nd edn., 2005.

13 P. W. Atkins and R. S. Friedman, *Molecular quantum mechanics*, Oxford University Press, Oxford, 4th edn., 2005.

14 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.

15 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 520–552.

16 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 553–586.

17 T. A. Halgren and R. B. Nachbar, *J. Comput. Chem.*, 1996, **17**, 587–615.

18 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 616–641.

19 T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 720–729.

20 T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 730–748.

21 MacroModel user manual, 2013.

22 XCluster Macromodel Manual, 2012.

23 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, 864–871.

24 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, 1133–1138.

25 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.

26 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

27 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.

28 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

29 Y. Zhao and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2701–2705.

30 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

31 Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.

32 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1789–1799.

33 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 1–19.

34 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.

35 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.

36 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.

37 M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265–3269.

38 F. Boys, S.F.; Bernardi, *Mol. Phys.*, 1970, **19**, 553–556.

39 http://www.nist.gov/mml/csd/informatics_research.

40 K. k Irikura and D. J. Fruip, *Computational Thermochemistry. Prediction and Estimation of Molecular Thermodynamics.*, American Chemical Society, Washington DC, 1999, vol. 10.

41 I. Usabiaga, J. González, P. F. Arnáiz, I. León, E. J. Cocinero and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12457–12465.

42 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.

43 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625–632.

44 J. R. Lane, J. Contreras-García, J. P. Piquemal, B. J. Miller and H. G. Kjaergaard, *J. Chem. Theory Comput.*, 2013, **9**, 3263–3266.

45 R. Chaudret, B. de Courcy, J. Contreras-Garcia, E. Gloaguen, A. Zehnacker-Rentien, M. Mons and J.-P. Piquemal, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9876–9891.

46 J. González, I. Baños, I. León, J. Contreras-García, E. J. Cocinero, A. Lesarri, J. A. Fernández and J. Millán, *J. Chem. Theory Comput.*, 2016, **12**, 523–534.

47 J. Gonzalez, I. Usabiaga, P. F. Arnaiz, I. Leon, R. Martínez, J. Millán and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2017, (DOI: 10.1039/C6CP08476A).

48 W. M. Latimer and W. H. Rodebush, *J. Am. Chem. Soc.*, 1920, **42**, 1419–1433.

49 L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, Cornell University Press, Ithaca, 1st edn., 1960.

50 E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.*, 2011, **83**, 1619–1636.

51 E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.*, 2011, **83**, 1637–1641.

CHAPTER 4

AMINO ACIDS AND DNA BASES

4. - AMINO ACIDS AND DNA BASES

Amino acids are the building blocks of proteins. There are 20 natural amino acids in humans, which differ in the nature of their side chain. The amino acid sequence of a protein determines its three-dimensional structure and its biological function.[1–5] On the other hand, DNA bases are the building blocks of DNA, the molecule where genetic information is stored.[6]

In this section, we present a study on the conformational behavior of isolated amino acids and DNA bases, as a previous step to tackle in following chapters the interaction between both types of molecules. Here, we characterized the intramolecular non-covalent interactions that shape the amino acids using several theoretical methods, so the most convenient one can be selected for the studies on molecular aggregates.

Three different types of amino acids were selected for the study, according to the nature of their lateral chain. Hence, alanine (A), valine (V) and isoleucine (I) were chosen as amino acids with non-polar side chain; asparagine (N) and glutamine (Q) were selected as polar side chain amino acids, while, arginine (R) was chosen as a representative member of the amino acids with charged side chain. This amino acid may be considered as an extreme case of polar amino acid and therefore its study will be presented as part of the polar amino acids section. Finally, phenylalanine (F), tryptophan (W) and tyrosine (Y) represented the amino acids with aromatic side chain. On the other hand, cytosine (C), thymine (T), adenine (A) and guanine (G) were selected as pyrimidinic and puric DNA bases, leaving aside only uracil, which is found in RNA but not in DNA.

The final goal is to study the protein-DNA interaction, which takes places between individual pairs of amino acids and DNA bases. Thus, the amino acids are integrated in the protein, forming peptidic bonds and the terminal carboxylic and amino groups are not available for the interaction, but they are forming amide groups. Therefore, to build a system closer to the real situation, the amino and acid groups of the amino acids were replaced by two amide groups at each end of the molecule. Likewise, the methylated derivatives of the DNA bases were used, to block the nitrogen atom where the deoxyribose molecule is attached in DNA.

Computed structures were named as aXn and mYn, where aX, X = A, V, I, N, Q, R, F, W and Y refers to the amino acid derivative and mY, Y = C, G, A or T refers to the methylated DNA base. In both cases, n = 1,

2, 3… represents the relative stability of the structure, starting with the most stable.

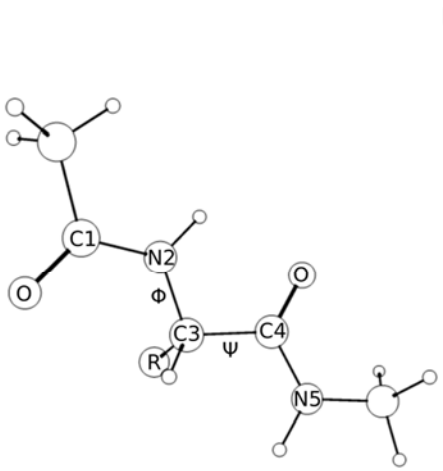
A significantly large amount of calculations was carried out on the selected systems, to ensure a faithful representation of the conformational landscape. Despite that they were grouped into families, presenting all the structures or even representative structures for all the families along the text is not practical. Therefore, only the most stable structure and an additional structure for the second most stable conformation of each amino acid will be shown. Nevertheless, the 15 most stable structures of each amino acid may be found in Appendix 8.1, together with their relative energy. NCI plots[7–11] on the two structures will be presented in this section, while the 2D-NCI plot (ρ vs. s representations) of the two most stable conformations of each species are collected in the Appendix 8.1.

The 15 most stable structures of each amino acid were selected to calculate Gibb's free energy in the interval of temperatures: from 0 to 700 K, following the procedure described in section 3.2.5.

As described in Section 3.2.1, MMFFs force field[12] was used during the conformational exploration stage, setting the energy window in 30 kJ mol^{-1}. Full optimization of selected structures was carried out using several computational methods: M06-2X, B3LYP, B3LYP-D and MP2, but only the structures obtained at M06-2X/6-311++G(d,p) are shown.

4.1 Amino acids

There is a great deal of theoretical and experimental studies about the conformational behavior of the amino acids in the literature.[13–29] In this section, the conformational analysis of the three types of amino acids above-mentioned was carried out in order to characterize the intramolecular non-covalent interactions. A scheme with the number and the atom labeling for the selected amino acids is presented in **Figure 4.1.** The torsional dihedral angles ϕ and ψ denote the 9 possible conformations, whose labelling is collected in refs. 4,30.



R= $-CH_3$ (Alanine)

$-CH(CH_3)_2$ (Valine)

$-CH(CH_3)-CH_2-CH_3$ (Isoleucine)

$-CH_2-C(=O)-NH_2$ (Asparagine)

$-CH_2-CH_2-C(=O)-NH_2$ (Glutamine)

$-CH_2-CH_2-CH_2-NH-C(=NH_2^+)-NH_2$ (Arginine)

$-CH_2-C_6H_5$ (Phenylalanine)

$-CH_2-C_6H_4OH$ (Tyrosine)

$-CH_2-C_8H_5N$ (Tryptophan)

Figure 4.1 *Structure and atom labeling for the studied amino acids. R denotes the side chain of the amino acid, while ϕ and ψ denote the torsional dihedral angles. Different values of these angles result in the different conformations of the amino acids.*[31]

4.1.1 Non-polar amino acids

Exploration of the conformational landscape of alanine, valine and isoleucine generated 23, 29, and 66 structures, respectively. All of them were optimized using DFT. The two most stable structures of each amino acid are shown in **Figure 4.2**, together with the value of their relative

energies, the distances of the most relevant interactions, and the NCI plots. The values of the distances and angles of the hydrogen bonds and the electron density of these interactions are collected in **Tables 4.1** and **4.2**.

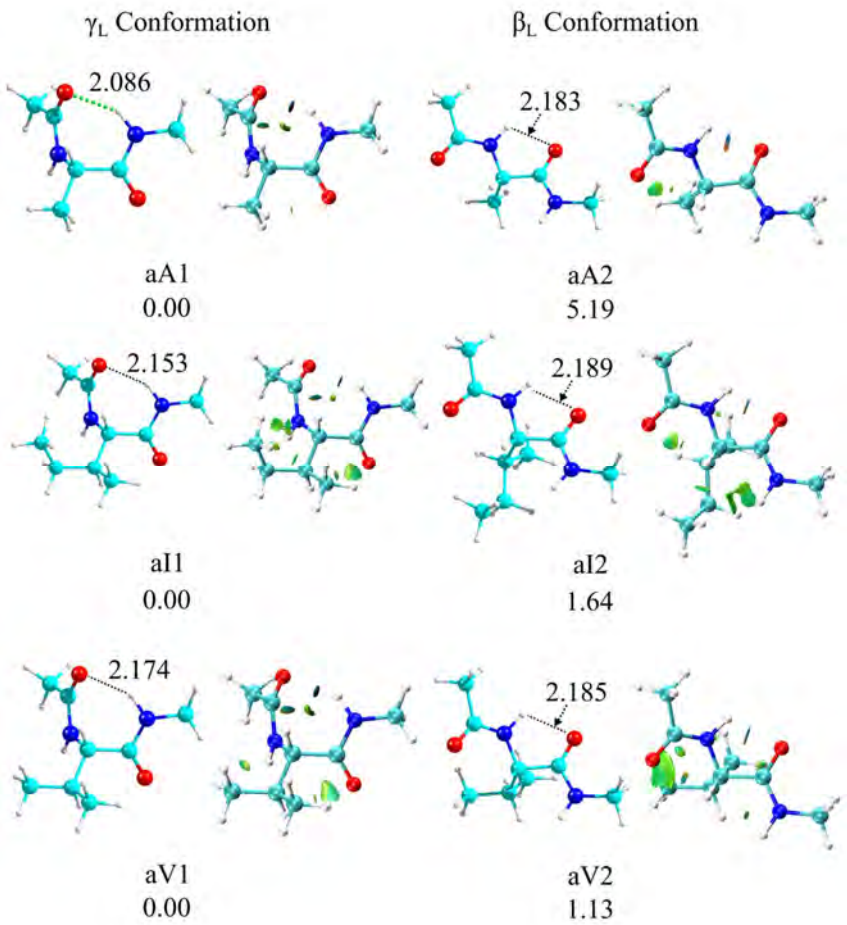


Figure 4.2 *The two most stable conformations, γ_L and β_L, for aA, aI and aV with distances (Å) of the hydrogen bonds (green) and the most relevant van der Waals interactions (black) and NCI plots. Relative energy values in kJ mol^{-1} are shown.*

The most stable structure of all three amino acids presents a C1=O···HN5 7-membered ring in what is known as an γ_L conformation.[4] Despite that C1=O···HN5 interaction is also present in γ_D conformations, γ_L ones are more stable due to an additional C1=O···HC3 interaction which stabilizes the final structure and become the global minimum.

Table 4.1. *Distances (Å) and angles (º) of the relevant intramolecular interactions observed for aA, aV and aI structures.*

		Interaction type		
		H bond	vdW	
		C1=O···HN5	C5=O···HN5	N2H···O=C4
γ_L	aA1	2.086/141.4		
	aV1		2.174/136.6	
	aI1		2.153/138.6	
β_L	aA2			2.183/106.0
	aV2			2.185/105.7
	aI2			2.189/105.7

According to the NCI plots, the C1=O···HN5 interaction is a hydrogen bond, as indicated by the dark blue disc in the NCI plot. However, the angle of the interaction indicates a non-optimal overlap between the orbitals of the interacting atoms, at least in the case of valine and isoleucine, due to steric reasons. In fact, the value of electron density confirms this assessment, which is slightly higher in alanine than in valine and isoleucine: 0.0192, 0.0164 and 0.0169 a.u. respectively. Furthermore, additional van der Waals interactions are also found, such as the above-mentioned C1=O···HC3 interaction, depicted with a bicolor blue/red isosurface due to the attractive contributions between the carbonyl and the hydrogen (blue) and the repulsive effect of the ring closure (red), together with some weak interactions with the hydrogens of the lateral chain depicted in green.

Thus, the final structures are a subtle balance of forces, where C1=O···HN5 is the leading interaction which helps to close the 7-membered ring. As the size of the side chain increases from valine to isoleucine, the interactions with the peptidic bond increase, opening the ring and reducing its relative contribution to the overall stability of the system. This is reflected in the energy difference between the global minimum and the second most stable structure, a β_L conformation, which decreases from alanine to valine and isoleucine in parallel with the increase in the C1=O···HN5 distance (5.19, 1.13 and 1.64 $kJ \cdot mol^{-1}$, respectively).

Table 4.2. *Electron density (a.u.) of the relevant interactions observed for aA, aV and aI, calculated from the NCI grid with 0.1x0.1x0.1 increments.*

		Interaction type		
		H bond	vdW	
		C1=O····HN5	C5=O····HN5	N2H···O=C4
γ_L	aA1	0.0192		
	aV1		0.0164	
	aI1		0.0169	
β_L	aA2			0.0197
	aV2			0.0196
	aI2			0.0194

Regarding the second most stable type of structures, β_L conformation for the three amino acids, the most relevant intermolecular interaction N2H···O=C4 forms an intramolecular 5-membered ring, which causes significantly stronger steric hindrances, as reflected by the red and blue bicolor isosurface in the NCI plots. The blue zone is related to the intramolecular attractive interaction (N2H···O=C4), while the red one is related to the steric effect due to the small number of members of the ring. There are also additional weak van der Waals interactions, mostly due to the contact between C1=O and N5H with the hydrogens of the lateral chain.

4.1.2 Polar amino acids

The size of the polar amino acids investigated, asparagine, glutamine and arginine, together with the presence of a carbonyl and an amino group at the end of their lateral chains, results in an increase in the number of structures found after the exploration and their later optimization: 51, 46 and 44 for aN, aQ and aR, respectively. The two most stable conformations found for each amino acid are displayed in **Figure 4.3**, together with the relative energies, the distance of the most relevant interactions, and the NCI plots. The values of the distances and angles for

the hydrogen bonds and the values of the electron density of the interactions are collected in **Tables 4.3** and **4.4**.

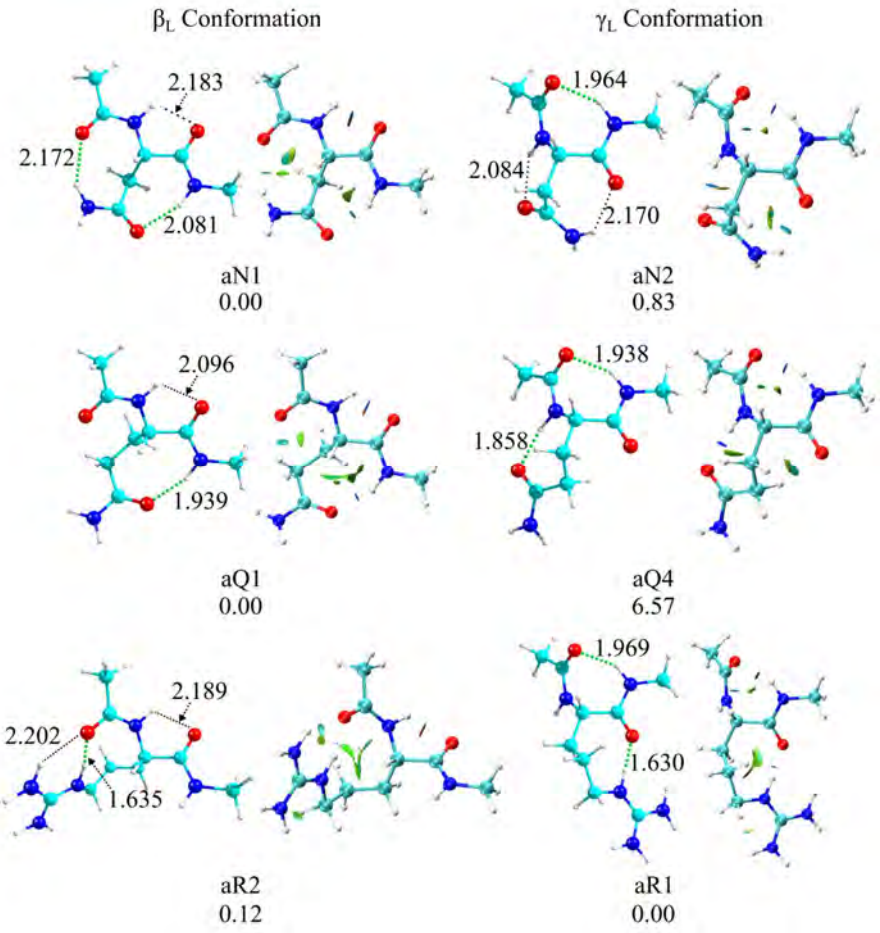


Figure 4.3 *The two most stable conformations, β_L and γ_L, for aN, aQ and aR with distances (Å) of the hydrogen bonds and the most relevant van der Waals interactions and NCI plots. Relative energy values in kJ mol^{-1} are shown.*

Leaving aside the structural differences with the non-polar amino acids, the two most stable structures of the three polar amino acids are also γ_L and β_L. However, the relative energetic order is reversed for asparagine and glutamine, for which the β_L conformer is more stable. Furthermore, the three most stable conformers for glutamine are β_L conformations and only differ in the rotation of the lateral chain. For this amino acid, the first structure with conformation γ_L is aQ4.

Table 4.3. *Distances (Å) and angles (º) of the relevant intramolecular interactions observed for aN, aQ and aR structures.*

		Interaction type								
		H bond					vdW			
		R: NH·· ··O=C1	R: C=O·· ··HN5	C_1=O·· ··HN5	R: C=O·· ··HN2	R: NH·· ··O=C4	N2H·· ··O=C4	R: NH·· ··O=C1	R: NH·· ··O=C4	R: C=O·· ··HN2
β_L	aN1	2.173/145.9	2.081/153.3				2.154/107.0			
	aQ1		1.939/160.5				2.096/109.2			
	aR2	1.635/166.4					2.138/105.7	2.202/136.3		
γ_L	aN2			1.964/147.4					2.170/127.7	2.084/134.7
	aQ4			1.938/148.3	1.858/158.5					
	aR1			1.969/143.3		1.630/174.4				

In addition to the N2H···O=C4 van der Waals interaction already observed in non-polar amino acids in β_L conformation, the presence of heteroatoms in the lateral chain of polar amino acids results in a rich collection of interactions, as it can be seen in **Figure 4.3**. The amide group of asparagine and glutamine side chains is prone to establish intramolecular interactions. So the aN1 conformer exhibits two intramolecular hydrogen bonds with the carbonyl and the amine moieties of the lateral chain. On the other hand, despite that glutamine's side chain only differs from asparagine in a methyl group, the orientation it adopts is very different: the carbonyl moiety presents a single hydrogen bond with the amine group of the peptidic backbone, although with shorter distance and higher electron density than the equivalent in asparagine. It seems that the lateral chain in glutamine somehow reduces the number of hydrogen bonds, preferentially maximizing one of them. The side chain of arginine seems also suitable to establish intramolecular interactions. In fact, its NH group forms a strong hydrogen bond with the C=O moiety of the peptidic chain, remarked by the high value of the electron density (0.0543 a.u) and the short binding distance (1.635 Å). Furthermore, an additional stabilizing interaction is observed between one of the NH_2 groups and the same carbonyl, depicted by a light blue zone in the NCI plot, but the angle is lower than 140º, far away from linearity. Some weaker van der Waals interactions are also present in the structure.

The main interaction in the γ_L conformation of polar amino acids exhibits a reduced interaction distance compared with the non-polar ones. Thus, the C1=O···HN5 distance varies from 2.086, 2.153 and 2.174 Å in aA1, aI1 and aV1 respectively to 1.964, 1.938 and 1.969 Å for aN2, aQ4 and aR1. The higher values of electron density in these interactions

(0.0264, 0.0259 and 0.0243 a.u., respectively) endorse the stronger character of this hydrogen bond in polar amino acids. Moreover, additional interactions with the electronegative moieties of the side chain seem to stabilize the final structures. One strong hydrogen bond is observed for glutamine (C=O···NH2, 0.0310 a.u.) and especially in arginine (H···O=C4, 0.0532 a.u.), which remarks the relevance of the intramolecular side chain interactions for the stabilization of these structures. For asparagine, two weaker stabilizing interactions with the amide group of the side chain are also observed. Furthermore, weaker van der Waals interactions related to the ring closure or to electronegative groups interacting with aliphatic hydrogen, are also found.

Table 4.4. *Electron density (a.u.) of the relevant interactions observed for aN, aQ and aR, calculated from the NCI grid with 0.1x0.1x0.1 increments.*

		Interaction type								
		H bond					vdW			
		R: NH·· ··O=C1	R: C=O·· ··HN5	C1=O·· ··HN5	R: C=O·· ··HN2	R: NH·· ··O=C4	N2H·· ··O=C4	R: NH·· ··O=C1	R: NH·· ··O=C4	R: C=O·· ··HN2
	aN1	0.0151	0.0188				0.0206			
β_L	aQ1		0.0242				0.0226			
	aR2	0.0543					0.0214	0.0142		
	aN2			0.0264					0.0171	0.0196
γ_L	aQ4			0.0259	0.0310					
	aR1			0.0243		0.0532				

4.1.3 Aromatic amino acids

Because the aromatic lateral chains are voluminous and lack the flexibility of the chains of previous systems, their conformational landscape is less complicated than for non-polar or polar systems. Consequently we obtained 29, 46 and 36 conformers for aF, aW and aY, respectively in the 30 kJ mol^{-1} window. All of them were optimized by DFT calculations.

As in previous systems, the most stable structures are γ_L or β_L conformations, as it can be seen in **Figure 4.4**; the relevant parameters of the most important interactions were collected in **Tables 4.5** and **4.6**.

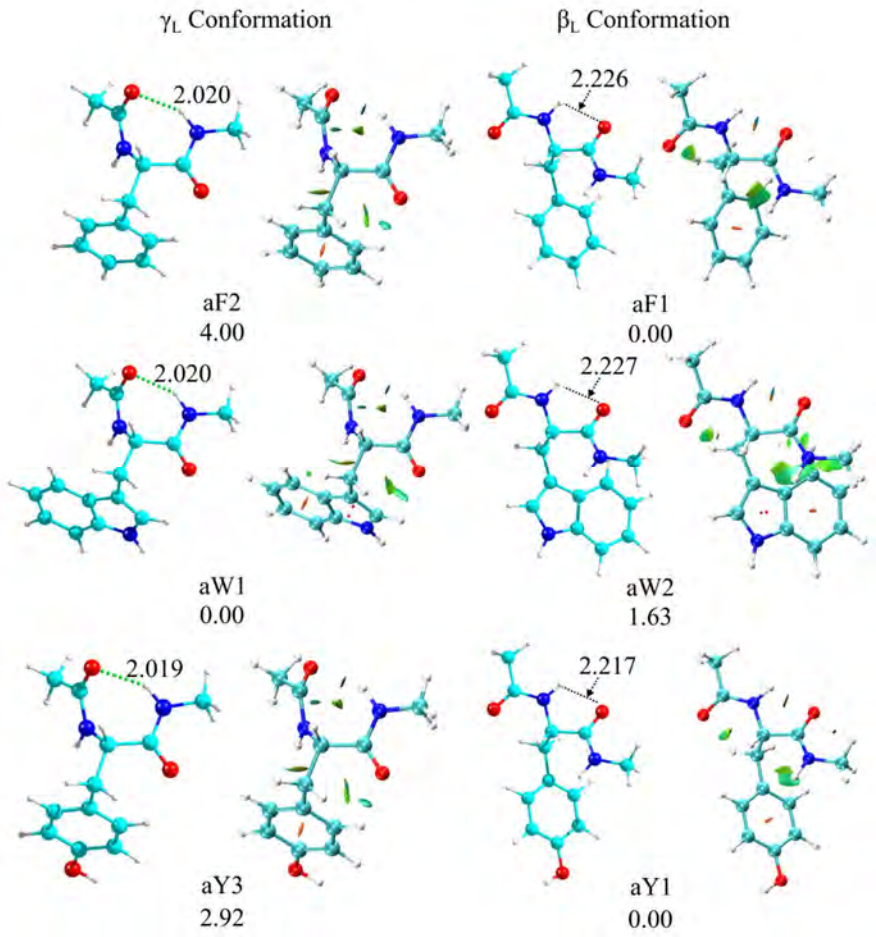


Figure 4.4 *The two most stable types, γ_L and β_L, for aF, aW and aY with distances (Å) of the hydrogen bonds and the most relevant van der Waals interactions and NCI plots. Relative energy values in kJ mol^{-1} are shown.*

The difference between γ_L and β_L conformations is mainly the relevance of the interaction between the peptidic bond and the aromatic ring. So, when the amino acid adopts an γ_L geometry, direct C=O···π and N-H···π interactions take place, as NCI plots display. However, β_L conformation allows the molecule to establish an N2-H···π interaction. Thus, the difference between both types of conformations seems very small - within the computation error – as indicated by the relative energy values and the interaction distances.

Table 4.5. *Distances (Å) and angles (°) of the relevant intramolecular interactions observed for aF, aW and aY structures.*

		Interaction type		
		H bond	vdW	
		C1=O····HN5	N2H···O=C4	R: N2H···π
β_L	aF1		2.226/104.8	~2.600/ -
	aW2		2.227/104.9	~2.700/ -
	aY1		2.217/105.0	~2.600/ -
γ_L	aF2	2.020/147.3		~2.500/ -
	aW1	2.019/147.1		~2.500/ -
	aY3	2.019/147.4		~2.500/ -

It is interesting to note that, despite the presence of a polar group in the lateral chain of aY, its behavior is different from that of previous polar amino acid. Thus, the rigidity of the lateral ring hampers formation of intramolecular interactions with the peptide backbone, as it occurs for aF and aW.

Table 4.6. *Electron density (a.u.) of the relevant interactions observed for aF, aW and aY, calculated from the NCI grid with 0.1x0.1x0.1 increments.*

		Interaction type	
		H bond	vdW
		C1=O····HN5	N2H···O=C4
β_L	aF1		0.0185
	aW2		0.0185
	aY1		0.0186
γ_L	aF2	0.0214	
	aW1	0.0215	
	aY3	0.0215	

4.1.4 Conformational distribution and Gibbs free energy analysis

The conformational distribution and the number of structures that each amino acid may adopt in the stability window investigated is displayed in **Figure 4.5**.

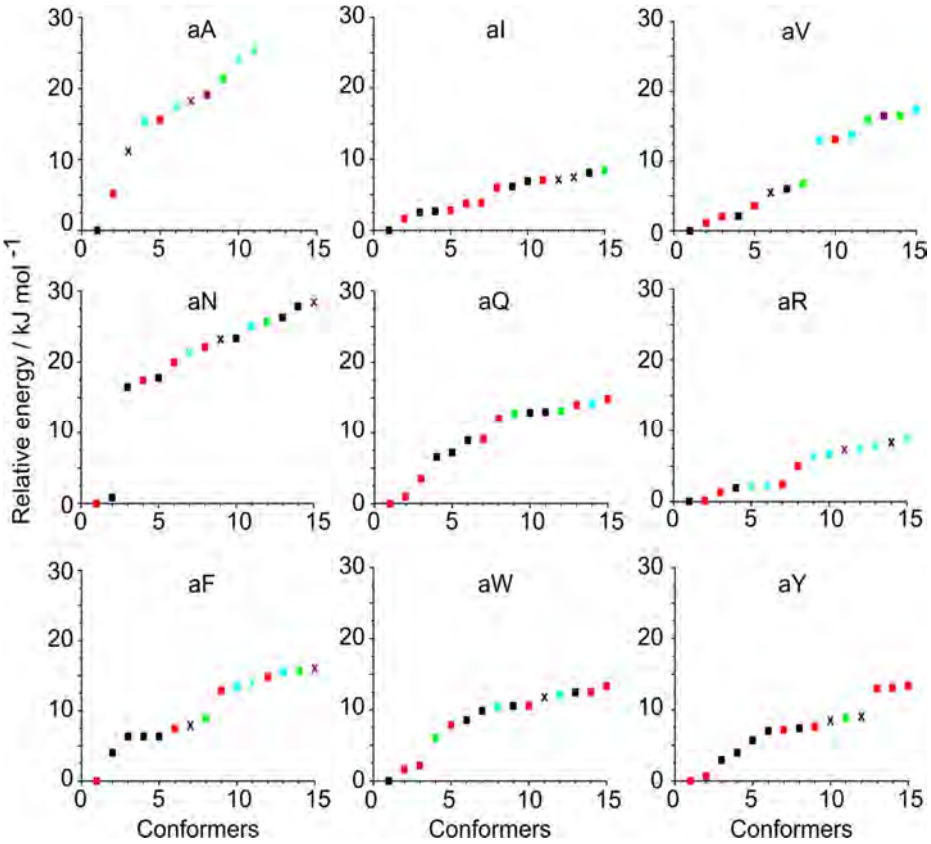


Figure 4.5 *Conformational distribution of the 15 most stable conformers of aA, aI, aV, aN, aQ, aR, aF, aW, and aY. Black: γ-conformation; red: β-conformation; blue: α-conformation; green: δ-conformation; purple: ε-conformation. (□: L-disposition; x: D-disposition).*

Although the calculation level of the study is reasonably high, changes in the relative stability of some structures can still occur if the calculation level is further increased. However, the tendency of all the amino acids investigated to adopt 7-member ring conformations (γ_L or γ_D,

represented as black squares and crosses, respectively) or 5-member conformations (β_L red squares) is clear and governs the region around the global minimum. For example, the third different conformation in non-polar amino acids appears above 6 kJ mol^{-1}; above 10 kJ mol^{-1} for polar and above 7 kJ mol^{-1} for aromatic. The only exception is aR, which presents several structures adopting α_D conformations. Despite the four most stable conformers are γ_L or β_L, the interactions with the lateral chain also favor this third type of structure (the first conformer of this family, aR5, is found at 2.17 kJ mol^{-1}).

It is worthy to mention that other conformations, such as α_L or α_D (blue symbols), δ_L or δ_D (green symbols) and ε_L or ε_D (purple symbols) are also found, but less frequently. Therefore, we have investigated a representative member of those isomers in order to analyze their structures. Structures and NCI plots of these conformations may be found in **Figure 4.6**. Hence, we found a weak N··NH interaction for α_D and δ_L with a distance too long to be taken as a 5 membered-ring. For the ε_D orientations, the N is not even pointing to the NH group, so it cannot form a ring. The rest of the non-covalent interactions found for α_D, δ_L and ε_D are not adequate to produce cyclic structures. Thus, we conclude that α, δ and ε do not give rise to ringed geometries. Moreover, they appear at high energy values (ranging from 6.05 kJ mol^{-1} for aW4 to 15.28 kJ mol^{-1} for aA4), remarking that the most stable conformers in gas phase are stabilized by formation of cyclic structures of γ or β types. The only exception is aR, as commented previously.

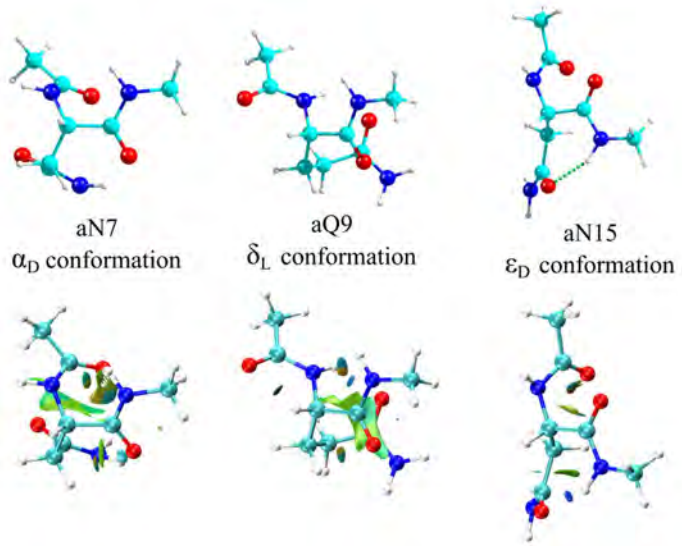


Figure 4.6 *Structure and NCI plots of a representative member for α_D, δ_L and ε_D conformations.*

In order to study the biological behavior of the amino acids, we calculated the variation of the Gibbs free energy from 0 to 700 K, the temperature at which organic matter decomposes.[32,33] The ΔG values are presented in **Figure 4.7**. As it can be seen, γ_L and β_L dispositions are still the most favored structures for all the amino acids investigated between 0-700 K. Focusing in the temperature range of life (300-310 K), it is clear that γ_L conformation is preferred in non-polar amino acids, being β_L less stable as temperature increases. For polar amino acids, aN and aR are clearly leaded by γ_L and β_L, which are separated by less than 2.5 kJ mol^{-1} in the whole interval of temperature. In fact, despite that β_L is the most stable orientation for aN, γ_L becomes more favorable between 150 and 700 K. On the other hand, for aR, γ_L and β_L conformers are isoenergetic. However, as temperature increases, β_L conformation evolves into the most favored structure. Furthermore, aQ clearly exhibits a β_L preference, being

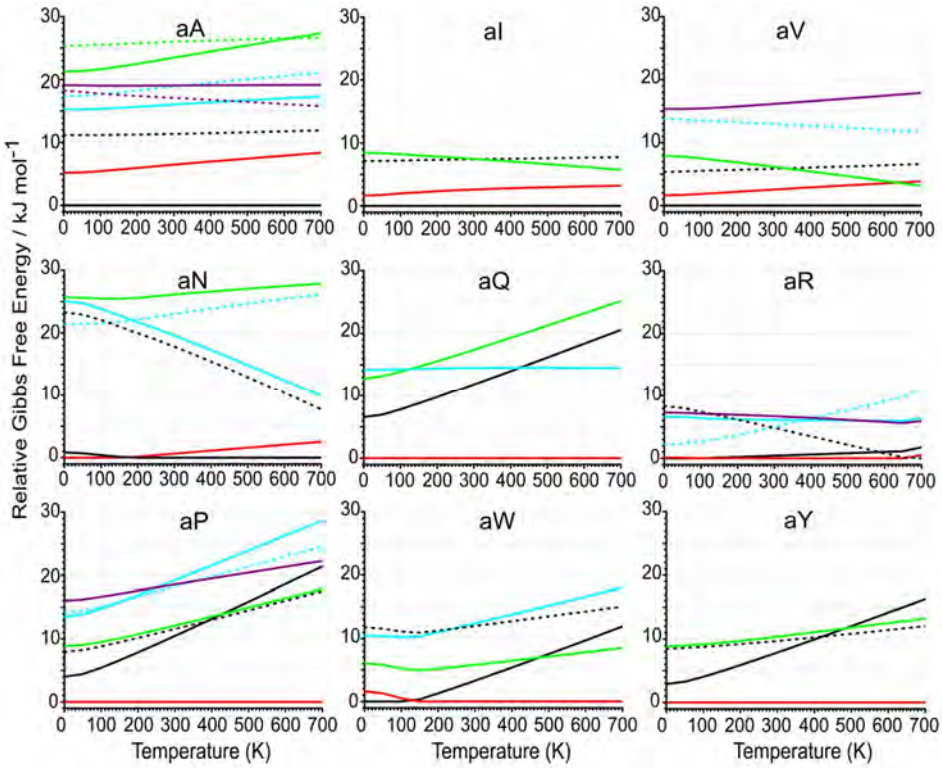


Figure 4.7 *Relative Gibbs free energy (kJ mol^{-1}) of aA, aI, aV, aN, aQ, aR, aF, aW, and aY calculated in the 0-700 K range. Black: γ-conformation; red: β-conformation; blue: α-conformation; green: δ-conformation; purple: ε-conformation. (continous line: L-disposition; dot line: D-disposition).*

even more separated from γ_L with the increase of the temperature. Finally, despite aromatic amino acids exhibited different global minimum conformations, when the temperature is taken into account, the preference of aP and aY for β_L conformations is clear in the whole range of temperature and above 150 K for aW.

Finally, comparison between different theoretical methods was made for three amino acids in order to select the most adequate computational method and to estimate the calculation time. Alanine, arginine and phenylalanine were selected due to the different nature of the side chain and size, with a number of atoms ranging from 22 (alanine) to 36 (arginine). Four different methods (B3LYP, B3LYP-D, MP2 and M06-2X) were tested; all of them with the triple-ζ basis set with polarization and diffuse functions. Furthermore, the time inverted in a re-optimization and frequency calculation for MP2 and B3LYP-D method, starting from a M06-2X optimized structure was also tested (**Figure 4.8**).

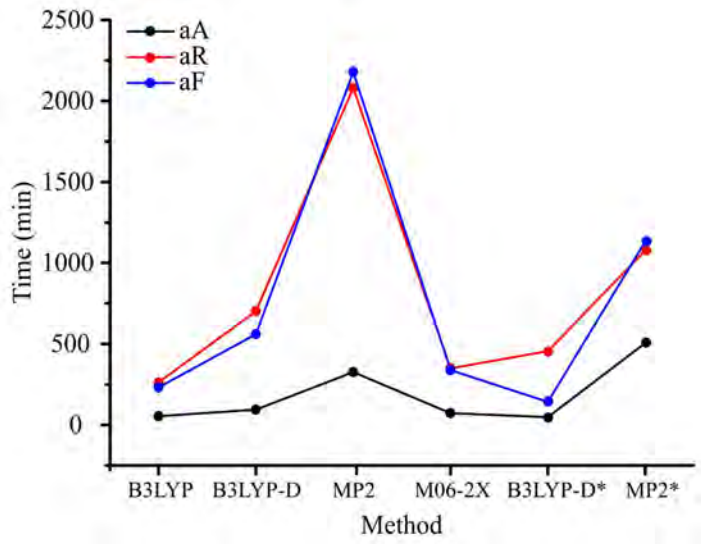


Figure 4.8 *Comparison of the calculation time of the geometry optimization and normal-mode analysis for different methodologies with 6-311++G(d,p) basis set. The asterisk (*) indicates that the starting structure comes from a previous M06-2X optimized geometry.*

As it can be seen, the calculation time increases with the size of the system. So, for alanine's most stable structure, the calculation time goes from 54 to 505 minutes for the different methodologies; whereas for phenylalanine, the time varies from 231 to 2179 minutes. For arginine,

computation times between 262 and 2080 minutes were needed. Thus, a substantial increase of the time was observed with the system size.

Analyzing the different methods, MP2 is significantly more CPU-expensive. Furthermore, it is evident that B3LYP functional requires the least calculation time although not significant differences were observed between the three functionals (B3LYP, B3LYP-D and M06-2X). Unfortunately B3LYP is not a suitable methodology to address systems in which dispersive forces are relevant. Hence, using either M06-2X or B3LYP-D results in significantly shorter computation times and therefore, the rest of the study will be done using those two functionals.

4.2 DNA bases

Due to their importance for life, there is a vast literature of theoretical and experimental studies on the DNA bases.[34–42] **Figure 4.9** collects the atom numbering of the puric and pyrimidinic DNA bases investigated.

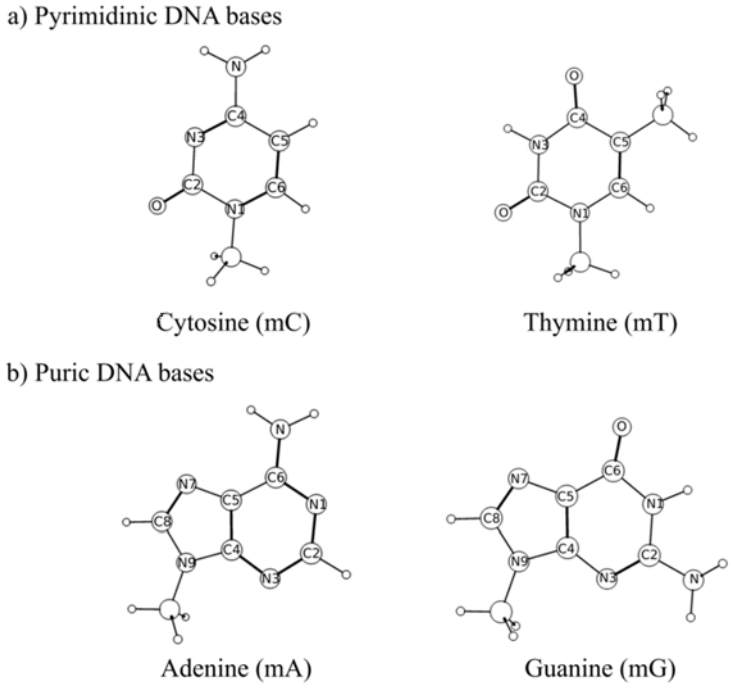


Figure 4.9 *Structure and atom labeling for the studied DNA bases. The position and the name of the methyl group has been omitted for a better understanding.*

The DNA bases present cyclic, rigid structures whose only conformational variability is the tautomeric form they may adopt, depending mainly on the pH of the surrounding medium. As previously explained, only the biologically relevant tautomer was considered here, as a previous step in the study of the DNA-protein interaction and a methyl group was added in the position where the sugar is attached to the base. The resulted structures may be found in **Figure 4.10**.

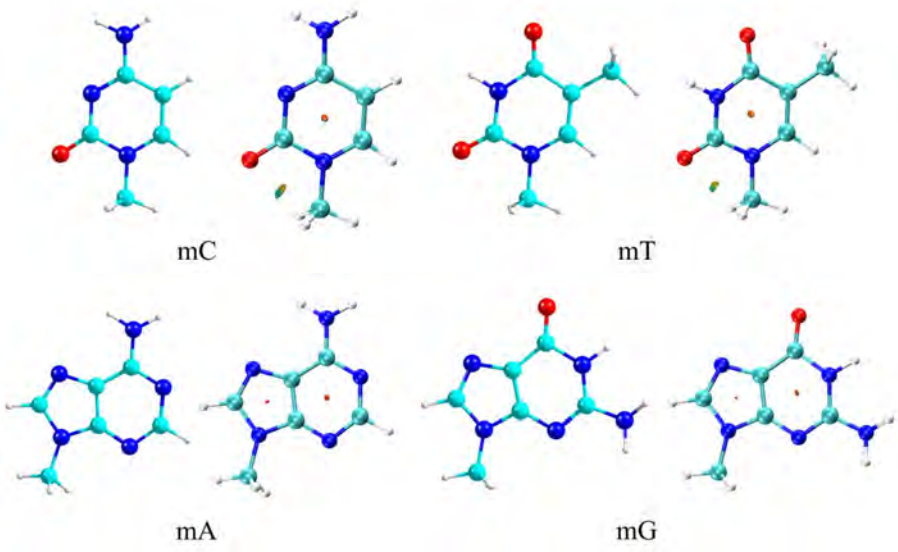


Figure 4.10 *Structure of the four DNA bases and the most relevant interactions depicted in the NCI plots.*

4.3 Conclusions

Several conclusions can be extracted from the study of the conformational behavior of the amino acids and the bases investigated.

We found that non-covalent intramolecular interactions, hydrogen bonds and van der Waals dispersive forces, control the orientation of the 15 selected conformers for each amino acid, for which we perform a detailed conformational map. However, they play a different role in the studied amino acids, according to the character of their lateral chain.

Thus, for non-polar amino acids (alanine, valine and isoleucine), the lateral chain stays away from the backbone for the two most stable conformers, whereas, for polar amino acids (asparagine, glutamine and

arginine), the lateral chain plays a main role in the stability of the system. The same applies to the aromatic amino acids (phenylalanine, tryptophan and tyrosine), which exhibit interactions between the backbone and the aromatic ring.

The analysis of the most relevant non-covalent intramolecular interactions found for the two most stable conformations of the nine amino acids showed that they are always cyclic structures, having a ring of 7 (γ_L or γ_D) or 5 (β_L) members. Both conformations are of similar stability, with differences in most cases within the computation error. Fort the rest of structures investigated, γ or β conformations are still the preferred dispositions in all the range of temperatures investigated. These configurations present characteristic intramolecular interactions; the most relevant is collected in **Figure 4.11**: γ_L conformation presents a stabilizing $C_1{=}O\cdots HN_5$ interaction, whereas, β_L conformation exhibits an interaction with double character, $N_2H\cdots O{=}C_4$. Furthermore, α, δ or ε conformations do not present interactions that help the formation of rings and appear less frequently and at high values of energy.

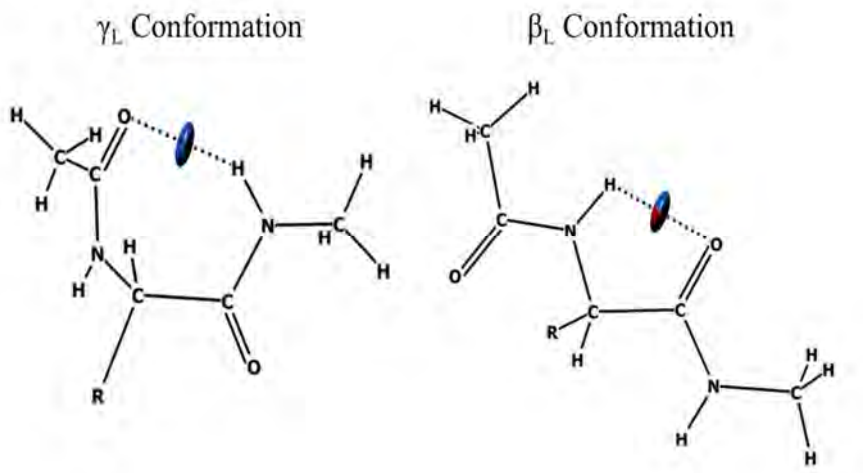


Figure 4.11 *The two most stable conformations for the studied amino acids, with the most relevant interaction depicted.*

So, we found the incipient dispositions that are recognized in larger systems (γ_L or β_L for the most stable conformers), as in polypeptides or protein structures. Therefore, the information about the conformations adopted by the isolated amino acids will be useful for their application to

larger systems because even for these small molecules and in absence of water, the biological structures are also present.

On the other hand, DNA bases exhibit only destabilizing interactions due to 5 and 6-member strains of the aromatic rings.

Finally, it was found that M06-2X or B3LYP-D are the suitable functionals to be used for the rest of the study.

4.4 References:

1 D. L. Nelson and M. M. Cox, *Principles of Biochemistry*, Omega, New York, 4th edn., 2000.

2 L. Stryer, J. M. Berg and J. L. Tymoczko, *Biochemistry*, Reverté, New York, 5th edn., 2002.

3 J. S. Richardson, in *Advances in Protein Chemistry*, Academic Press, New York, 1st edn., 1981, vol. 34, pp. 167–339.

4 G. A. Chass, M. A. Sahai, J. M. S. Law, S. Lovas, Ö. Farkas, A. Perczel, J.-L. Rivail and I. G. Csizmadia, *Int. J. Quantum Chem.*, 2002, **90**, 933–968.

5 G. A. Petsko and D. Ringe, *Protein Structure and Function*, New Science Press, London, 1st edn., 2004.

6 B. Alberts, A. Johnson, D. Morgan, M. Raff and J. Lewis, *Molecular Biology of the Cell*, Taylor & Francis Inc, New York, 2014.

7 J. R. Lane, J. Contreras-García, J. P. Piquemal, B. J. Miller and H. G. Kjaergaard, *J. Chem. Theory Comput.*, 2013, **9**, 3263–3266.

8 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625–632.

9 B. J. Miller, J. R. Lane and H. G. Kjaergaard, *Phys. Chem. Chem. Phys.*, 2011, **13**, 14183–14193.

10 S. K. Kim, T. Ha and J.-P. Schermann, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3334–3335.

11 R. Chaudret, B. de Courcy, J. Contreras-Garcia, E. Gloaguen, A. Zehnacker-Rentien, M. Mons and J.-P. Piquemal, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9876–9891.

12 T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 730–748.

13 A. G. Császár, *J. Phys. Chem.*, 1996, **100**, 3541–3551.

14 A. G. Császár, *J. Mol. Struct.*, 1995, **346**, 141–152.

15 M. Rassolian, G. A. Chass, D. H. Setiadi and I. G. Csizmadia, *J. Mol. Struct.-THEOCHEM*, 2003, **666–667**, 273–278.

16 A. Barth, *Prog. Biophys. Mol. Biol.*, 2000, **74**, 141–173.

17 M. N. Barroso, E. S. Cerutti, A. M. Rodríguez, E. A. Jáuregui, O. Farkas, A. Perczel and R. D. Enriz, *J. Mol. Struct.-THEOCHEM*, 2001, **548**, 21–37.

18 B. Boeckx and G. Maes, *Biophys. Chem.*, 2012, **165–166**, 62–73.

19 C. Cabezas, M. Varela, I. Pena, S. Mata, J. C. Lopez and J. L. Alonso, *Chem. Commun.*, 2012, **48**, 5934–5936.

20 F. C. Calaza, M. V. Rigo, A. N. Rinaldoni, M. F. Masman, J. C. P. Koo, A. M. Rodríguez and R. D. Enriz, *J. Mol. Struct.-THEOCHEM*, 2003, **634**, 201–213.

21 M. Chen, Z. Huang and Z. Lin, *J. Mol. Struct.-THEOCHEM*, 2005, **719**, 153–158.

22 M. A. da Silva, M. da Silva, A. F. Santos, M. V Roux, C. Foces-Foces, R. Notario, R. Guzman-Mejia and E. Juaristi, *J. Phys. Chem. B*, 2010, **114**, 16471–16480.

23 C. M. Deane, F. H. Allen, R. Taylor and T. L. Blundell, *Protein Eng.*, 1999, **12**, 1025–1028.

24 P. D. Godfrey, S. Firth, L. D. Hatherley, R. D. Brown and a. P. Pierlot, *J. Am. Chem. Soc.*, 1993, **115**, 9687–9691.

25 S. Gronert and R. A. J. O'Hair, *J. Am. Chem. Soc.*, 1995, **117**, 2071–2081.

26 B. Hernández, F. Pflüger, M. Nsangou and M. Ghomi, *J. Phys. Chem. B*, 2009, **113**, 3169–3178.

27 B. Lakard, *J. Mol. Struct.-THEOCHEM*, 2004, **681**, 183–189.

28 A. Lesarri, R. Sánchez, E. J. Cocinero, J. C. López and J. L. Alonso, *J. Am. Chem. Soc.*, 2005, **127**, 12952–12956.

29 S. Dokmaisrijan, V. S. Lee and P. Nimmanpipug, *J. Mol. Struct.-THEOCHEM*, 2010, **953**, 28–38.

30 I. Hudáky, R. Kiss and A. Perczel, *J. Mol. Struct.-THEOCHEM*, 2004, **675**, 177–183.

31 G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99.

32 I. Usabiaga, J. González, P. F. Arnáiz, I. León, E. J. Cocinero and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12457–12465.

33 J. Gonzalez, I. Usabiaga, P. F. Arnaiz, I. Leon, R. Martínez, J. Millán and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2017, (DOI: 10.1039/C6CP08476A).

34 E. Nir, C. Plützer, K. Kleinermanns and M. De Vries, *Eur. Phys. J. D*, 2002, **20**, 317–329.

35 N. J. Kim, G. Jeong, Y. S. Kim, J. Sung, S. K. Kim and Y. D. Park, *J. Chem. Phys.*, 2000, **113**, 10051–10055.

36 L. Sadr-Arani, P. Mignon, H. Chermette, H. Abdoul-Carime, B. Farizon and M. Farizon, *Phys. Chem. Chem. Phys.*, 2015, **17**, 11813–11826.

37 D. Mishra and S. Pal, *J. Mol. Struct.-THEOCHEM*, 2009, **902**, 96–102.

38 K. W. Choi, J. H. Lee and S. K. Kim, *J. Am. Chem. Soc.*, 2005, **127**, 15674–15675.

39 G. C. P. Van Zundert, S. Jaeqx, G. Berden, J. M. Bakker, K. Kleinermanns, J. Oomens and A. M. Rijs, *ChemPhysChem*, 2011, **12**, 1921–1927.

40 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. De Vries, *J. Chem. Phys.*, 2001, **115**, 4604–4611.

41 E. Nir, M. Müller, L. I. Grace and M. S. De Vries, *Chem. Phys. Lett.*, 2002, **355**, 59–64.

42 A. M. Rijs and J. Oomens, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Springer International Publishing, Cham, 1st edn., 2015.

CHAPTER 5

DNA-PROTEIN AGGREGATION

5. - DNA – PROTEIN AGGREGATION

In the previous chapter, we explored the conformational preferences of representative amino acids and DNA bases. Such studies will be complemented here with the exploration of the DNA-protein interactions, which are abundant in the cell and control some of the most important biological processes, including DNA/RNA-protein interactions, such as RNA transcription, DNA methylation, DNA expression and nucleosome formation.[1–10] In fact, the nucleosome formation is one of the most interesting examples of these contacts. It is known that interactions between histones and DNA result in formation of a superstructure composed of 146 base pairs wrapped around an histone octamer.[11] However, the leading mechanism that guides this process is still unknown. Is there a specific sequence of amino acids that forces the DNA into bending and surrounding the histone or is the DNA sequence that may contains certain repetitions of base pairs which guides the process?

In this chapter, we applied a reductionist approach to shed some light on the subject. We divided the interaction between the two supermolecules into DNA base – amino acid pairs, which may be analyzed employing high-level quantum chemistry calculations. The results were compared with those from previous studies on the contacts between the 20 natural amino acids and the four DNA bases. We try here to address if the nature of the amino acid (polar, charged, non-polar or aromatic) is a determinant factor in the interactions with the DNA base, if there is an interaction pattern in the DNA bases and amino acids interactions and if such interactions can be divided in pairs.

The model systems built to study the DNA base-amino acid interactions take into account that the DNA bases are attached to a deoxyribose and that the amino acids form part of a protein. To simulate these facts at an affordable computational cost, we used the methylated version of the bases and amino acid derivatives with peptide bonds at both sides of the lateral chain. The structures generated in this way and their aggregates were named as aXmYn, where aX represents the amino acid: the a denotes that it is a modified amino acid and the X is the IUPAC nomenclature for amino acids; mY represents the DNA base in the aggregate: m refers to the presence of the methyl group and the Y is the IUPAC letter for the DNA base. Finally, n refers to the relative stability of the isomer, starting from 1 for the most stable one. So a structure named aQmT2 is the second most stable structure formed by the combination of

the acetylated derivative of glutamine (aQ) and the methylated derivative of thymine (mT).

As described in section 4.1, the study was carried out in three steps: exhaustive exploration of the PES for the intermolecular interactions, full optimization using DFT methods and analysis of the non-covalent interactions using NCIplot. The large amount of calculations carried out during this work makes impractical to present all the structures within the text. So, only the most stable structure of each cluster will be presented here. The twenty most stable structures of each aggregate may be found in Appendix 8.2, together with their relative energy.

The force field selected for the Molecular Mechanics in this chapter was Merck Molecular Mechanics Force Field (MMFFs)[12] with an energy window of 30 kJ mol^{-1}. The number of structures obtained with this force field ranged from 108 structures for the smallest system to 429 structures for the largest one. These structures were gathered into families as explained in chapter 3 in order to reduce the problem to a reasonable number of structures for quantum mechanics calculations.

DFT calculations were conducted using M06-2X, which has been applied to countless systems, due to its accurate description of dispersive interactions,[13–15] together with the 6-311++G(d,p) basis set. However, some of the systems were also re-optimized using B3LYP-D to estimate the contribution of the dispersive forces.[16] All the binding energy values include the BSSE correction estimated using the Counterpoise method.[17] NCIplots of each system will also be presented.

Formation of the cluster induced sometimes a distortion in the geometry of the amino acids, and the dissociation energy was affected of a difference at least equal to the deformation energy which had to be taking into account in the calculation of the dissociation energy. This energy was calculated as the difference between the stability of the amino acid in the same geometry of the cluster and in its closest minimum.

5.1 Amino acid – DNA base aggregates

Statistical studies on the contacts between the amino acids of the histones and the DNA bases display large disparity in the number of contacts,[18–28] that seems to be related to the nature of the amino acid's residue: those amino acids with polar or charged side chains show higher number of contacts with the nucleobase than the ones with a non-polar side chain. Furthermore, the aromatic amino acids do not fit very well within this sequence.

In order to have representative examples of each kind of interaction, we selected the nine amino acids previously studied: the three polar and charged amino acids with the highest number of contacts: arginine (R), asparagine (N) and glutamine (Q); the three non-polar amino acid with the lowest ones: alanine (A), valine (V) and isoleucine (I); and the three amino acids with aromatic side chain: phenylalanine (F), tryptophan (W) and tyrosine (Y).

5.1.1 Non-polar amino acids – DNA base

5.1.1.1 Non-polar amino acids – pyrimidinic DNA bases

Exploration of the conformational landscape of the amino acids presented in chapter 4 resulted in 23, 29 and 66 structures for aA, aV and aI respectively in an energy window of 30 kJ mol^{-1}. On the other hand, DNA bases can present several different tautomers. However, only the biologically-relevant ones were considered.

Non-polar amino acids-mC:

The large number of structures found for the isolated amino acids translated into a complex conformational landscape for the aggregates. Since the differences between some of the structures were not significant, they were gathered into families using Xcluster tool.[29] Next, at least two structures of each family, the energetically most favorable and the one with the lowest mean differences within the family, were subjected to full optimization. This process resulted in 32 structures for aAmC, 42 for aVmC and 45 geometries for aImC. The most stable structure found for each system using this procedure may be found in **Figure 5.1**, together with the corresponding binding energy value. The length of the most

relevant intermolecular interactions are also shown in the left panel, while the right panels show the NCI plots with all the non-covalent interactions, classified by the color code defined in Section 3.2.6. Clearly, mC clusters show the formation of two hydrogen bonds (HB) between the peptidic bond members of the amino acid and the electronegative groups of the DNA bases. This is further confirmed by the NCI plots, which show deep blue zones between these groups. There is an additional interaction, indicated with a black line in **Figure 5.1**, between the hydrogen of the anomeric carbon of the amino acid and the N3 of the cytosine. This interaction is not a hydrogen bond, as highlighted by the NCI plot, but is

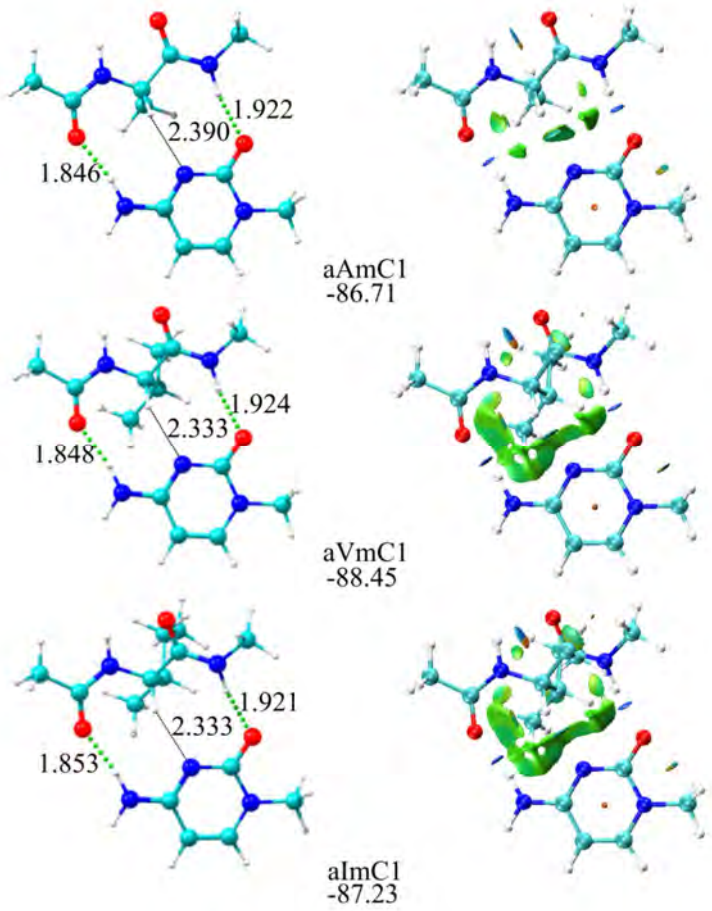


Figure 5.1 *The most stable structure found for aAmC, aVmC and aImC complexes, together with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are also shown.*

produced by dispersive forces. The values of the distance and the angles of the different interactions were collected in **Table 5.1.** The length of these hydrogen bonds is very similar, and therefore, of comparable strength. A deeper analysis of the NCI peaks shows that both C1=O···H(NH) hydrogen bonds exhibit similar electron density. Some differences appear in the extension of the dispersive interaction (aAmC1 surfaces are smaller than those of aVmC1 and aImC1), but such differences do not significantly alter the binding energy of the system, mainly determined by the hydrogen bond framework. The values of the electron density of the HB are collected in **Table 5.2**.

Table 5.1 *Distances (Å) and angles (degrees) of the intermolecular interactions calculated for aAmC1, aVmC1 and aImC1 dimers*

	Interaction type		
	H bond		**vdW**
	C1=O···H(NH)	**N5H···O(C)**	**C3H···N3**
aAmC1	1.846/168.3	1.922/168.9	2.390/129.6
aVmC1	1.848/169.13	1.924/168.4	2.333/138.9
aImC1	1.853/169.5	1.921/169.5	2.333/138.7

Table 5.2 *Electron density (a.u.) of the hydrogen bonds was calculated for the clusters aAmC1, aVmC1and aImC1, from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type	
	H bond	
	C1=O···H(NH)	**N5H···O(C)**
aAmC1	0.0281	0.0253
aVmC1	0.0279	0.0252
aImC1	0.0276	0.0255

As it can be seen in **Figure 5.1,** the main aggregation forces between non-polar amino acids and DNA base are HB. In fact, the side chain of the amino acid stays away from the mC (at almost 3 Å), playing a secondary role in the aggregation process. There are also some van der Waals (vdW)

interactions, but due to their weakness do not play a leading role in the aggregation processes.

The analysis of the dispersive contribution to the binding energy interaction by Grimme's D3 approach is shown in **Figure 5.2** and it confirms the low influence of the dispersive force in the binding energy. The contribution is very similar in all three systems, although slightly smaller in aAmC1. Consequently, the dissociation energy values are very similar.

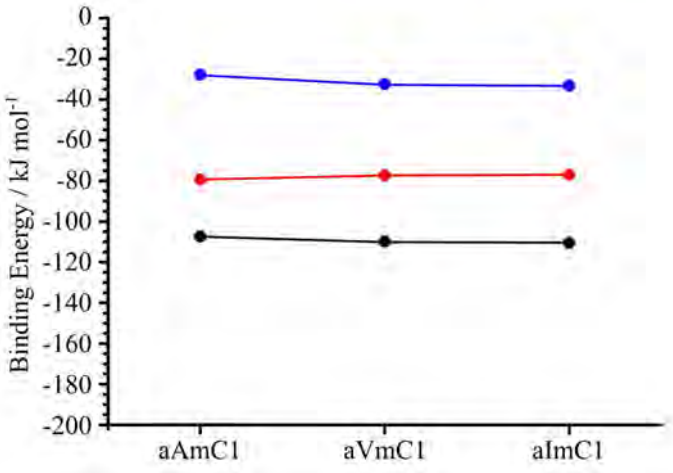


Figure 5.2 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aAmC, aVmC and aImC.*

So far we analyzed the structure and the leading forces behind the formation of the most stable structure in each of the three dimers containing non-polar amino acids and a DNA base. Regarding the rest of the structures calculated for each dimer, some changes in the number of intermolecular hydrogen bonds and in the contribution of the dispersive forces to the total binding energy was observed, depending on the type of structure adopted by the system. **Figure 5.3** shows a summary of the 20 most stable structures, sorted according to their relative stability. The symbols follow a color-and-shape code to help identification of the leading interactions in each of the conformers.

As it can be seen, the preferred interactions are two hydrogen bonds between the electronegative groups of the backbone and the base, as

already-described for the global minimum. In fact, the four most stable structures (three in the case of the aAmC) belong to the same family and the first structure with a different aggregation trend appears above 10 kJ mol^{-1}, confirming the prevalence of the observations made for the global minimum. In summary, the study in **Figure 5.3** also confirms that the hydrogen bond between the peptide bond members and the base is the favored leading force in the aggregation of non-polar amino acids and mC.

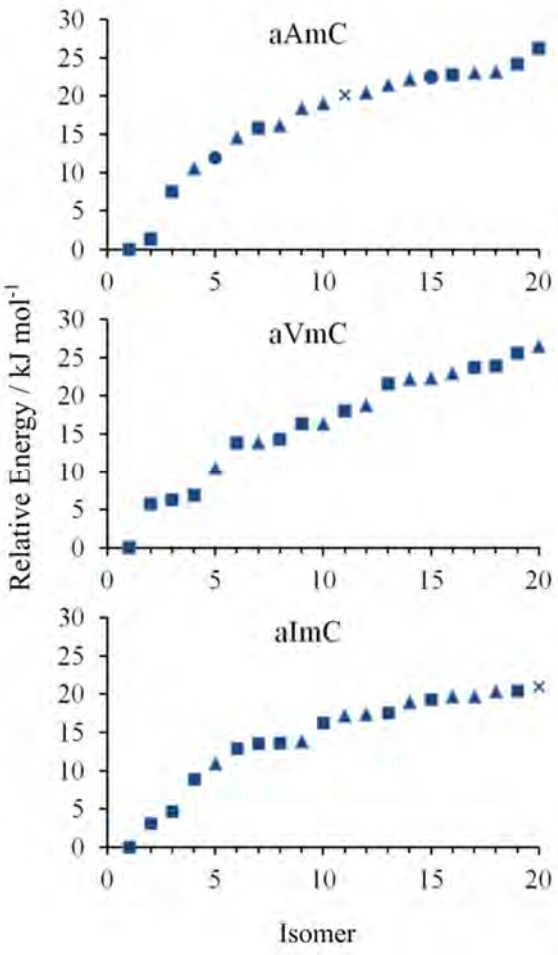


Figure 5.3 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aAmC, aVmC and aImC. Blue color depicts the interaction between the peptidic bond and the base. Types of interactions: HB: hydrogen bond (○: three, □: two, and Δ: one HB), vdW: van der Waals interactions (×).*

Non-polar amino acids-mT

Following the clustering process, the number of structures submitted to full optimization was 44 for aAmT, 58 for aVmT and 61 structures for aImT. The most stable structure of each aggregate may be found in **Figure 5.4**, with their respective binding energy. The most relevant interactions with their distances were added to the panels on the left for comparison and the NCI plots in BGR code may be found in right panels.

Clearly, mT clusters present a hydrogen bond between the amine group of the peptide bond and one of the carbonyl groups of mT. This is confirmed by the NCI plots, which show a deep blue area between those groups. The values of the distance and the angles of the different interactions are displayed in **Table 5.3** and the values of the electron density of these hydrogen bonds were collected in **Table 5.4**. The length of the hydrogen bond is very similar for aVmT and aImT and slightly longer for aAmT. This is translated into a marginally lower binding energy

Table 5.3 *Distances (Å) and angles (degrees) of the intermolecular interactions observed in aAmT1, aVmT1 and aImT1.*

	Interaction type				
	H Bond			**vdW**	
	N2H··· ···O=C2	**N5H··· ···O=C2**	**N2H··· ···O=C4**	**C4=O··· ···π**	**R:H···π**
aAmT1	2.042/142.7			~2.900/-	
aVmT1		1.999/167.0			~2.700/-
aImT1			1.984/156.9	~2.800/-	

Table 5.4 *Electron density (a.u.) of the hydrogen bonds observed for aAmT1, aVmT1and aImT1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type		
	H bond		
	N2H···O=C2	**N5H···O=C2**	**N2H···O=C4**
aAmT1	0.0207		
aVmT1		0.0201	
aImT1			0.0223

value for aAmT. Compared with mC complexes, mT aggregates present longer hydrogen bonds distances and therefore, smaller binding energy values. In good agreement, the electron density values for the hydrogen bonds in mT aggregates are smaller than for mC complexes.

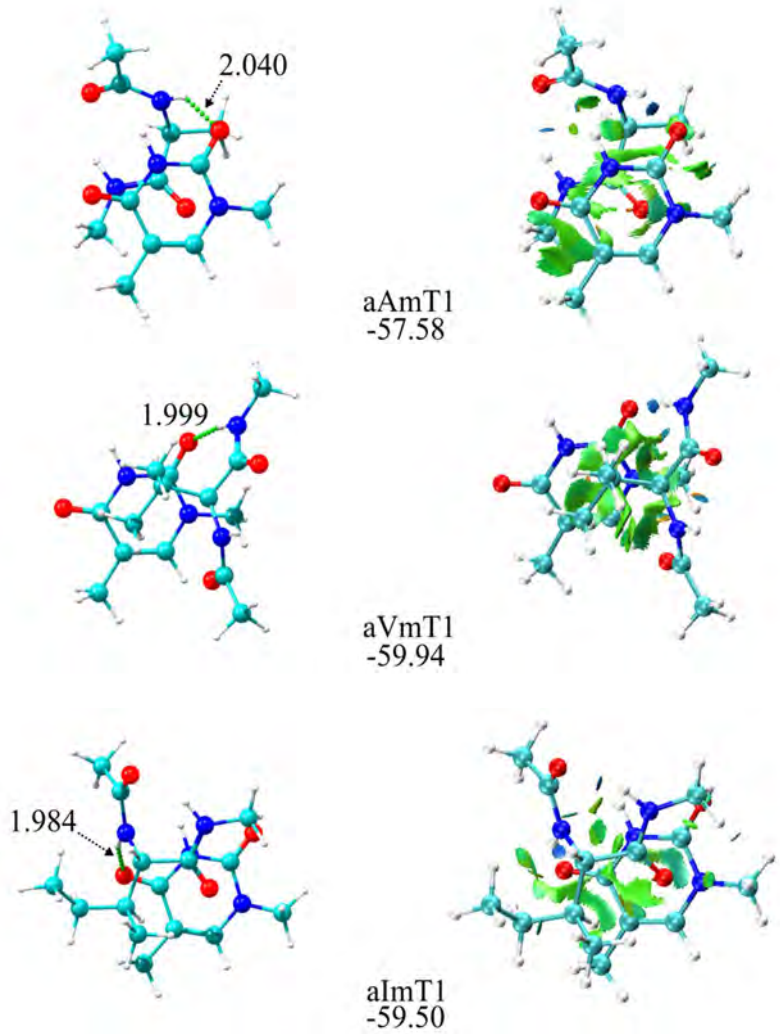


Figure 5.4 *The most stable structure found for aAmT, aVmT and aImT complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are also shown.*

NCI plots also highlight the increase in the van der Waals interactions due to the contact between the aromatic ring of the base and the electronegative groups of the amino acids. The methyl group in C5 in mT also increases this effect, in comparison with the mC analogues. In order to quantify the dispersive contribution to the binding energy, **Figure 5.5** shows the analysis by Grimme's D3 approach. There is no significant variation between the mT aggregates; nevertheless, the higher impact of the dispersive forces in comparison with the mC aggregates it is clearly observed (26-30% of the total interaction energy in mC aggregates compared to 68-78% in mT aggregates).

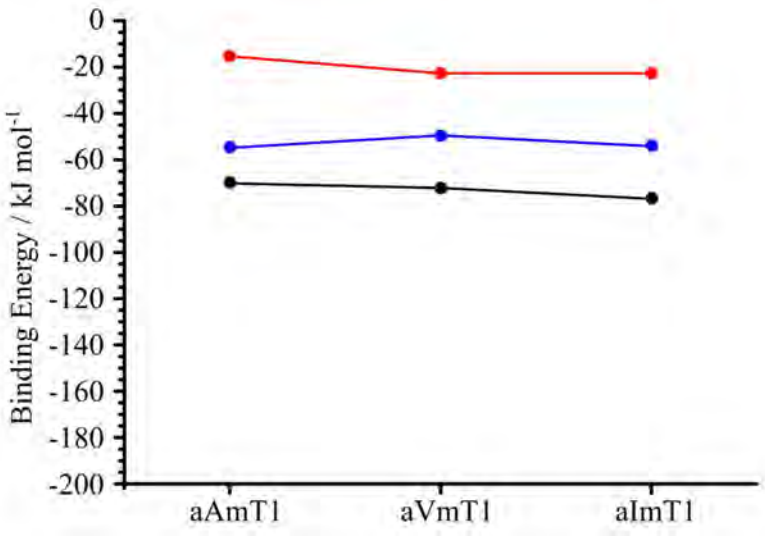


Figure 5.5 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aAmT, aVmT and aImT.*

An analysis of the leading forces for the 20 most stable structures of each aggregate may be found in **Figure 5.6**. The study confirms that ~98% of the structures present at least one hydrogen bond between the backbone of the amino acid and the base. Furthermore, analyzing the region nearby the global minimum, the two most stable structures display the same interaction pattern: one hydrogen bond between the groups of the peptide bond and mT.

In summary, this study corroborates the dispersive interactions as the preferred leading force in the aggregation between non-polar amino acids and mT, as in the mC aggregates, although some hydrogen bonds between the electronegative groups of the backbone and the base are also formed. However, the moderate growth of the dispersive forces and the important decrease in the strength of the hydrogen bonds resulted in lower values of the binding energy, in comparison with the mC complexes.

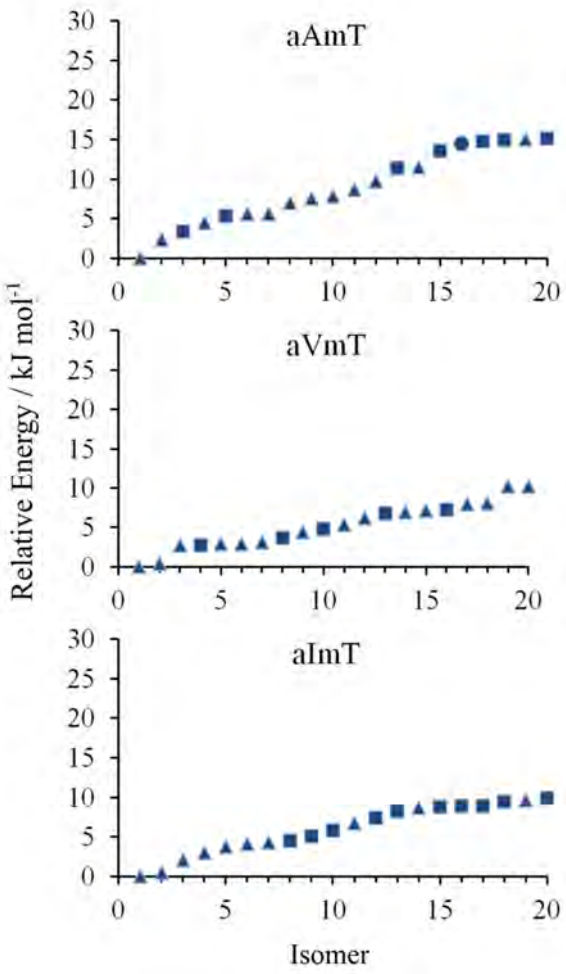


Figure 5.6 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aAmT, aVmT and aImT. Blue color depicts the interaction between the peptidic bond and the base. Types of interactions: HB (○: three, □: two, and Δ: one HB), vdW interactions (×).*

5.1.1.2 Non-polar amino acids – puric DNA bases

A similar study to that presented in sections 5.1.1 was carried out, replacing mC and mT by the puric methylated DNA bases: mG and mA. Once more, only the biologically-relevant tautomer of the DNA base was considered.

Non-polar amino acids-mA:

A set of 39 structures for aAmA, 43 for aVmA and 46 for aImA were subjected to full optimization. The structures of the most stable non-polar amino acid-mA aggregates were collected in **Figure 5.7**, together with the value of their binding energy and the distance of the most relevant interactions. The NCI plots of these structures are presented in the right panels.

Formation of one hydrogen bond between the members of the peptide bond and the mA is clearly observed. NCI plots highlight as well the presence of this hydrogen bond, depicted by the blue zones. The distance of these interaction, **Table 5.5,** and the values of the electron density, **Table 5.6,** are similar for all the aggregates studied. An important observation is that in the three structures in **Figure 5.7** the lateral chain of the amino acid stays away from the DNA base and therefore it does not contribute to the intermolecular interaction.

Table 5.5 *Distances (Å) and angles (º) of the intermolecular interactions observed for aAmA1, aVmA1 and aImA1.*

	Interactiontype		
	H bond	**vdW**	
	C1=O···H(NH)	**O···π**	**N···π**
aAmA1	2.033/151.4	~ 3.100/-	~3.200/ -
aVmA1	2.016/151.7	~3.100/-	~3.300/ -
aImA1	2.046/151.1	~3.100/-	~3.100/-

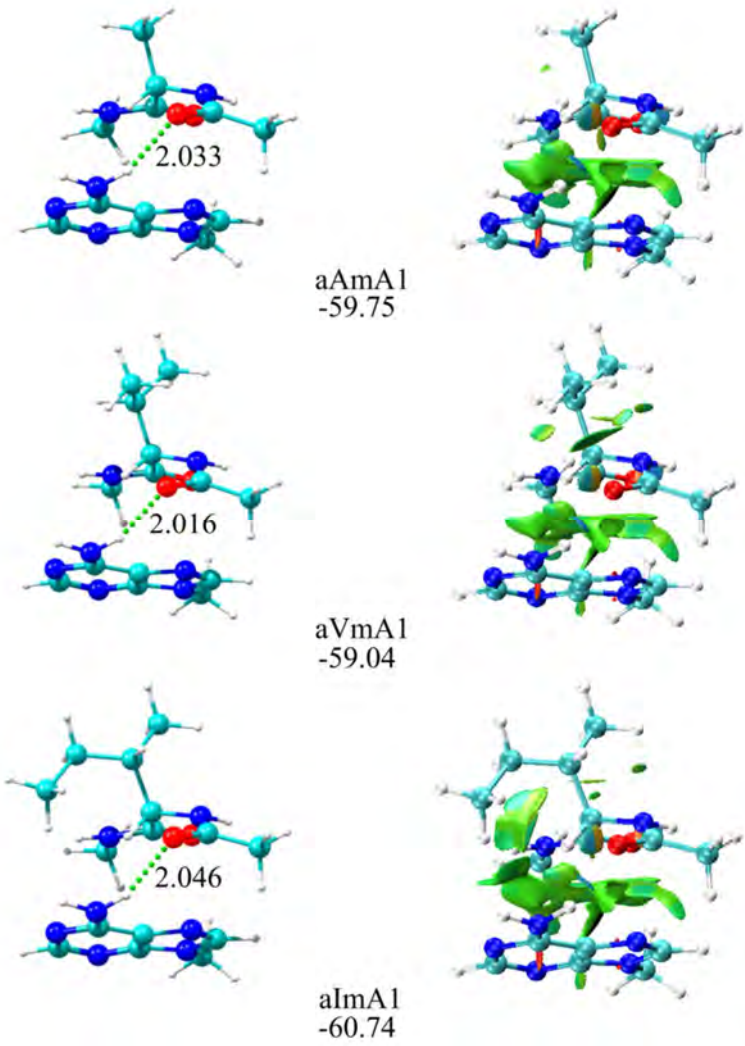


Figure 5.7 *The most stable structure found for aAmA, aVmA and aImA complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are also shown.*

NCI representation highlights the dispersive nature of the stacking interactions for mA aggregates, represented by a broad green surface in the plot. This dispersive character was already observed for mT aggregates. In fact, mA aggregates present an interaction pattern visibly similar to that of mT complexes.

Table 5.6 *Electron density (a.u.) of the hydrogen bonds observed for aAmA1, aVmA1and aImA1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interactiontype
	H bond
	C1=O···H(NH)
aAmA1	0.0189
aVmA1	0.0198
aImA1	0.0187

The analysis of the dispersive contribution is shown in **Figure 5.8**. This dispersive contribution is similar for all the mA complexes and accounts for the ~86-92% of the total interaction contribution. NCI plots also remark the importance of this effect with wide green zones.

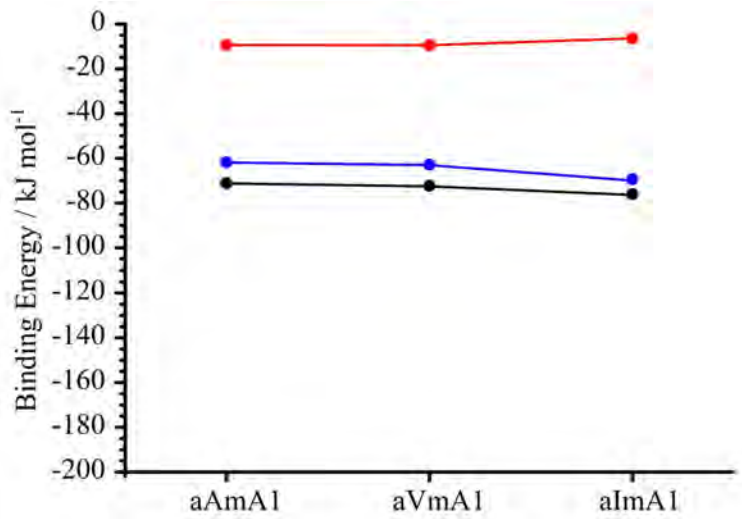


Figure 5.8 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aAmA, aVmA and aImA.*

Figure 5.9 shows the study of the non-covalent leading interactions found in the 20 most stable structures of each complex. As in previous systems, formation of hydrogen bonds between the amines and carbonyls

of the backbone and the DNA base take place in ~77% of the aggregates. However, a significant number of structures present no formation of intermolecular hydrogen bonds, especially in aAmA. Probably the increased size of the π electronic cloud is now able to stablish

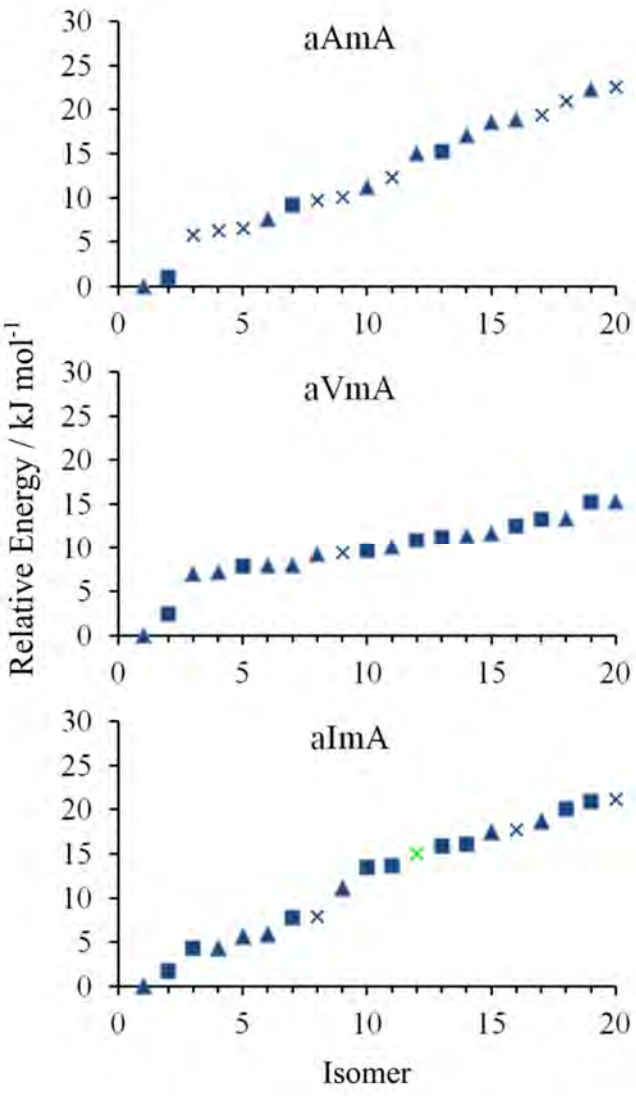


Figure 5.9 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aAmA, aVmA and aImA. Blue: interaction between the peptidic bond and the base. Green: interaction between the lateral chain and the base. Types of interactions: HB (□: two, Δ: one HB), vdW interactions (×).*

strong-enough intermolecular interactions, that compete favourably with the formation of hydrogen bonds. Therefore, those species only exhibit vdW interactions with the π cloud.

It is worthy to mention that in one of the structures, aImA12, there is a direct interaction between the side chain and the π cloud of the DNA base, depicted in green in **Figure 5.9**. However, this is not a relevant structure as it lies at ~15 kJ mol^{-1} above the global minimum. Focusing in the region close to the global minimum, two different families were observed, showed in **Figure 5.10**. They stablish one or two hydrogen bonds with the peptide bond members, emphasizing the importance of the backbone electronegative members in the interaction with the mA.

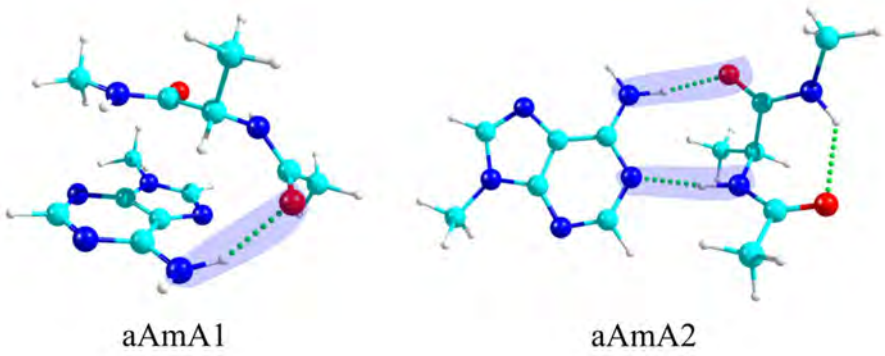


Figure 5.10 *Structure of aAmA1 and aAmA2. The intermolecular hydrogen bonds with the peptidic bond was highlighted in blue.*

In summary, the hydrogen bond between the peptide bond and the base are the favored leading force in the aggregation of non-polar amino acids with mA, but the vdW interactions have increased in importance in comparison with mC and mT aggregates.

Non-polar amino acids-mG

The conformational analysis of the interaction between non-polar amino acids and mG resulted in 21 structures for aAmG and for aVmG and 32 conformers for aImG, which were subjected to full optimization using DFT. The most stable structure of each aggregate may be found in **Figure 5.11**, with the corresponding value of the binding energy. In this figure is shown as well the most relevant interactions with their distances and the NCI plots of these structures with BGR color code.

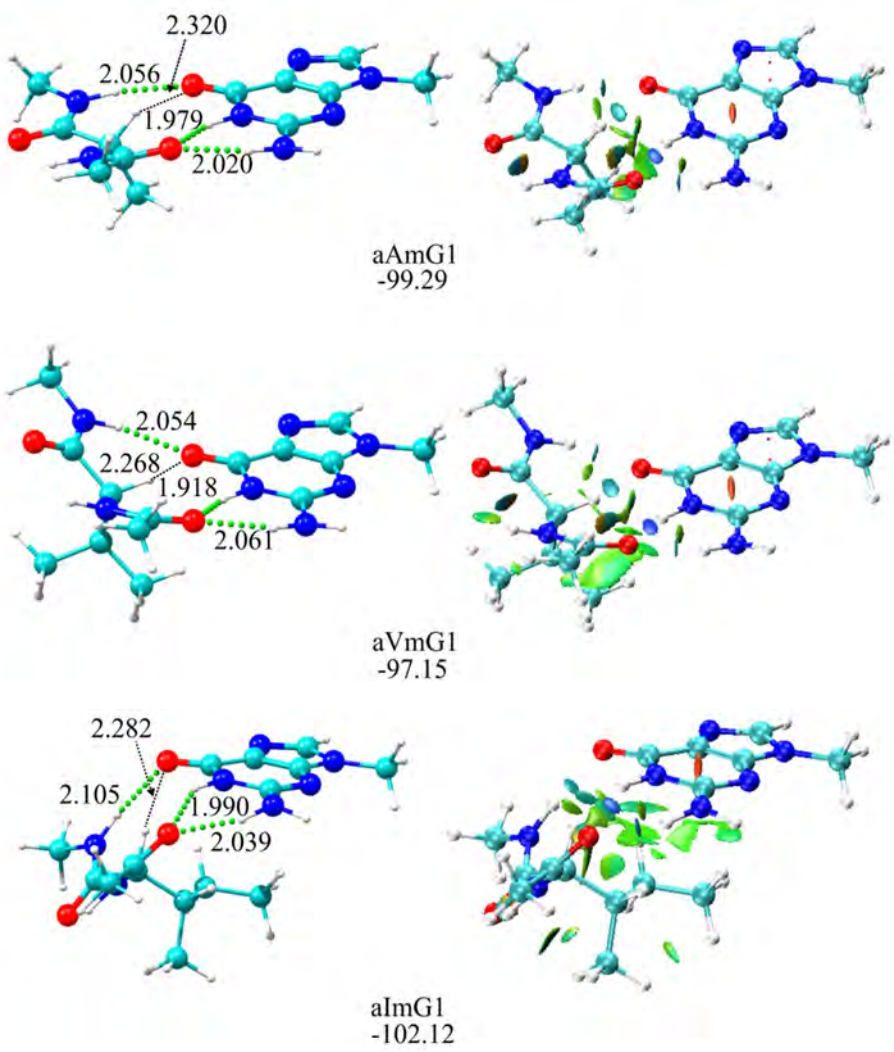


Figure 5.11 *The most stable structure found for aAmG, aVmG and aImG complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

The O=C-NH- sequence present in mG favors the formation of intermolecular hydrogen bonds. Indeed, such sequence matches very well the distribution of the hydrogen bond donor and acceptor groups in the amino acids. Consequently, all three systems presented three hydrogen bonds with those sites of mG, confirmed by blue zones in the NCI plot. Moreover, an additional attractive interaction with a lower importance was

found between the hydrogen of the asymmetric carbon and the carbonyl of the base, indicated by a light blue area in the NCI plots. The values of the distance and the angles of the different interactions are displayed in **Table 5.7**. The lengths of these hydrogen bonds present very close values, remarking the similar strength of these interactions.

Therefore, the interaction pattern changes absolutely in comparison with the other puric base. In this case, the presence of the C6=O leads to the formation of a hydrogen bond in a similar way as observed in mC aggregates. In good agreement, the analysis of the NCI peaks shows that C1=O···HN1 is the strongest interaction. The values of the electron density can be found in **Table 5.8**.

Table 5.7. *Distances (Å) and angles (º) of the intermolecular interactions observed in aAmG1, aVmG1 and aImG1.*

	Interactiontype			
	H bond			**vdW**
	C1=O·· ··H(NH)	**C1=O·· ··HN1**	**N5H·· ··O=C6**	**C3H·· ··O=C6**
aAmG1	2.020/147.2	1.979/153.4	2.056/157.3	2.320/124.7
aVmG1	2.061/145.7	1.918/155.1	2.054/157.7	2.268/133.1
aImG1	2.039/147.2	1.990/146.9	2.105/160.4	2.282/146.8

Table 5.8 *Electron density (a.u.) of the hydrogen bonds observed for the clusters aAmG1, aVmG1and aImG1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type		
	H bond		
	C1=O···H(NH)	**C1=O···HN1**	**N5H···O=C6**
aAmG1	0.0187	0.0233	0.0169
aVmG1	0.0174	0.0261	0.0171
aImG1	0.0185	0.0227	0.0159

It is worthy to note that the non-polar amino acid - mG interaction takes place exclusively with the peptidic chain, while the lateral chain stays away from the interaction site. In fact, it is placed in the opposite side, except for aImG, where there are secondary weak interactions between the side chain and the DNA.

The analysis of the dispersive contribution to the binding energy by Grimme's D3 approach may be found in **Figure 5.12** and confirms the low influence of the dispersive forces in the interaction energy, as expected by the change in the HB pattern. aImG aggregate presents a slightly higher contribution of the dispersive forces (41.6% vs ~26% for aAmG and aVmG) due to the non-significant interaction of the lateral chain with the base, already described above.

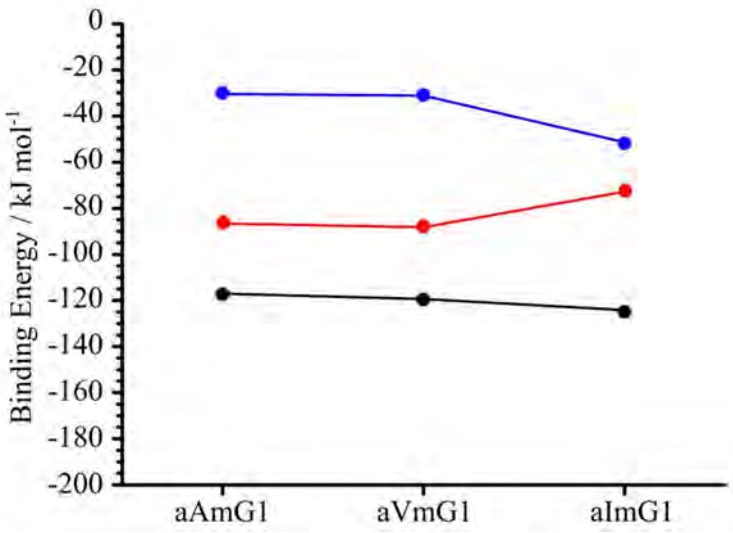


Figure 5.12 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aAmG, aVmG and aImG.*

When the relative stability of the calculated structures are represented against their energetic position, **Figure 5.13**, the color code highlights that in all three systems the neighborhoods of the global minima are dominated by structures with three hydrogen bonds between the peptidic bond and the DNA base. Indeed, this interaction is observed in ~72% of the studied structures, dominating the conformational landscape.

On the other hand, only a few structures, high in energy, interact throughout van der Waals interactions, in contrast with mA, the other puric base. Thus, despite having the same π cloud, the O=C-NH- sequence of mG tips the balance towards formation of intermolecular hydrogen bonds in non-polar amino acid - mG aggregates.

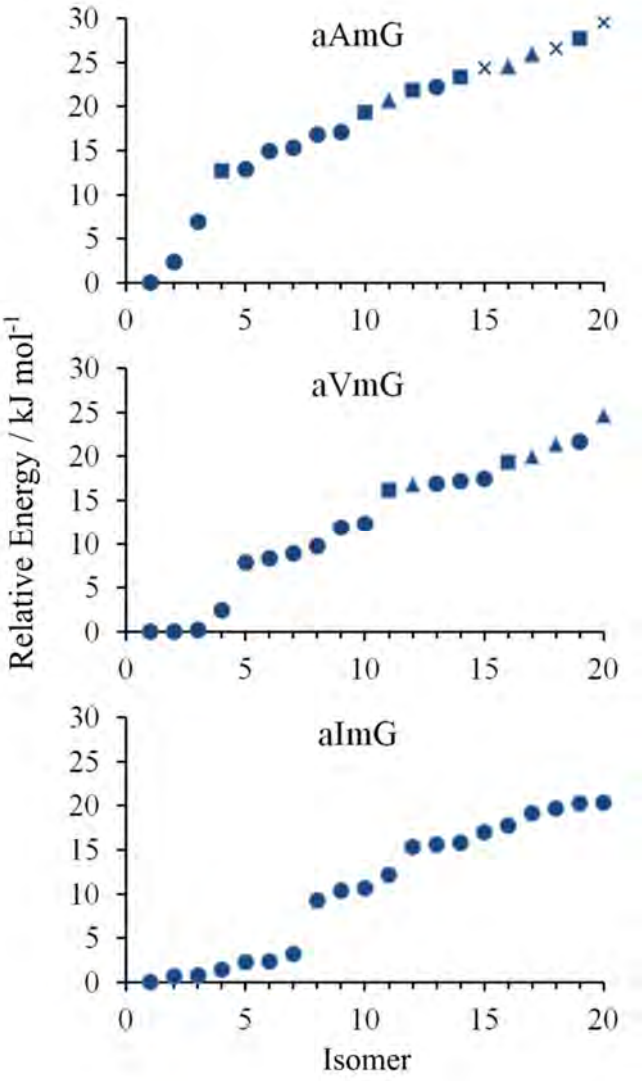


Figure 5.13 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aAmG, aVmG and aImG. Blue: interaction between the peptidic bond and the base. Types of interactions: HB (○: three, □: two, Δ: one HB), vdW interactions (×).*

5.1.2 Polar amino acids – DNA base

5.1.2.1 Polar amino acids – pyrimidinic DNA bases

As in previous sections, the polar amino acid will be presented first to the pyrimidinic DNA bases. The results obtained in chapter 4 for polar amino acids generated a set of 51, 46 and 44 structures for aN, aQ and aR respectively in an energy window of 30 kJ mol^{-1}.

Polar amino acids-mC

Taking into account the conformational complexity of the isolated monomer it is not surprising that a large number of structures were found for each complex. Clustering them into families using XCluster, the number of structures was reduced to 46 aggregates for aNmC, 56 for aQmC and 40 for aRmC.

The most stable structure of each aggregate may be found in **Figure 5.14**, together with the value of their binding energy. The most relevant interactions for each conformer and their distances are also displayed and the NCI plots highlight the character of the interactions. The values of the distances and angles for each conformer may be found in **Table 5.9**.

Table 5.9. *Distances (Å) and angles (º) of the intermolecular interactions observed in aNmC1, aQmC1 and aRmC1.*

	Interaction type					
	H bond					vdW
	N5H·· ··O(C)	C4=O·· ··H(NH)	N2H·· ··N3	R:H·· ··O(C)	R:O(C)·· ··H(NH)	R:H···N3
aNmC1	2.018/148.3		1.979/165.8		1.826/178.6	
aQmC1	2.110/140.1		1.953/170.6		1.836/173.2	
aRmC1		1.872/168.5		1.928/149.0 1.883/154.5		

The substitution of the non-polar side chain by polar substituents changes the interaction landscape. mC aggregates exhibit three intermolecular hydrogen bonds, as in the previous non-polar mC aggregates. However, at least one of the hydrogen bonds (two hydrogen

bonds in aRmC) is formed between the side chain and the mC base. This is further demonstrated by the NCI plots, which display these zones with a deep blue surface. In fact, the analysis reveals that the strongest hydrogen bonds in aNmC and aQmC are those formed between the lateral chain and the base (1.826 and 1.836 Å respectively) and a double hydrogen bond in aRmC (1.883 and 1.928 Å). NCI peaks analysis underlines the high electron density and the strength of these hydrogen bonds, as **Table 5.10** shows.

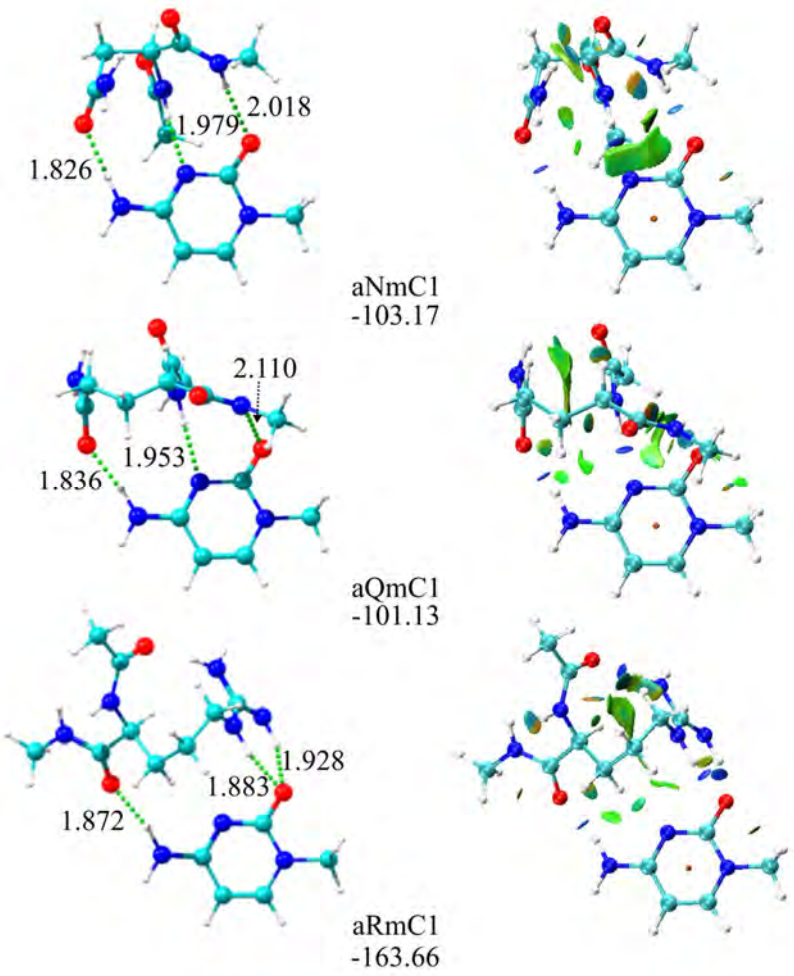


Figure 5.14 *The most stable structure found for aNmC, aQmC and aRmC complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

Table 5.10. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aNmC1, aQmC1and aRmC1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type				
	H bond				
	N5H·· ··O(C)	C4=O·· ··H(NH)	N2H·· ··N3	R:H·· ··O(C)	R:O(C)·· ··H(NH)
aNmC1	0.0209		0.0269		0.0288
aQmC1	0.0175		0.0286		0.0281
aRmC1		0.0297		0.0265 0.0258	

The binding energy values displayed in **Figure 5.14** are higher than those found for non-polar amino acids mC aggregates, due to formation of an additional hydrogen bond in polar – mC aggregates. The remarkably high binding energy of aRmC, may be due to the strong interaction between the lateral chain and the cytosine derivative.

Figure 5.15 shows the analysis of the dispersive contribution by Grimme's D3 approach. This dispersive contribution is similar for aNmC and aQmC complexes (29-31% of the binding energy). The contribution of the dispersive forces is noticeably smaller for aRmC aggregates, around 15% to the total binding energy. Such change is not due to a reduction in the strength of the dispersive forces compared with the other polar complexes, but to an important increase in the binding energy, related to the double hydrogen bond with the side chain that the electron deficiency of the charged side chain makes particularly strong.

Figure 5.16 shows the non-covalent leading interactions in the 20 most stable structures of each polar amino acid mC aggregate. This study highlights the presence of a type of contact not observed in previous cases. This interaction has been colored in red and represents simultaneous interactions of the base with the backbone and side chain (H-bonds or vdW). All the conformers that present such contacts are depicted in red, independently of which contact is more intense or takes places at shorter distance. Furthermore, when the hydrogen bond is formed only with the backbone, but a relevant van der Waals interaction also appears with the side chain, the contact appears in black.

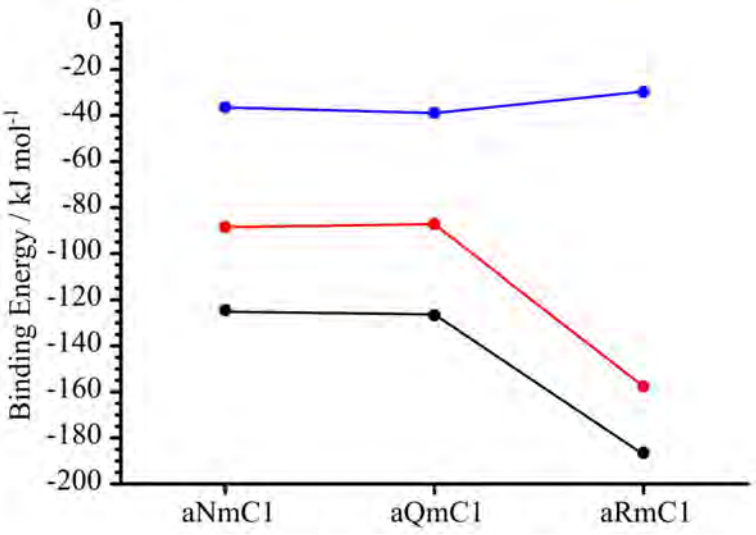


Figure 5.15 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aNmC, aQmC and aRmC.*

Altogether, the analysis emphasizes that hydrogen bonds are by far the leading force in the aggregation of mC with polar amino acids. The complexes formed always present at least one HB in all the structures with the exception of aNmC19, in which the main interaction is due to van der Waals forces with the electronegative moieties of the backbone. Moreover, the study also highlights the importance of the side chain in the aggregation, whose HB interactions with the DNA base are present in most of the studied structures. In fact, this tendency contrasts with the interaction pattern observed in non-polar amino acids DNA bases, in which its lateral chain was always away from the base.

Furthermore, the same trend is observed in the region near the global minimum. The energy distance between the most stable conformer and the next conformer with a different interaction trend is ~10 kJ mol^{-1}, pointing to a relatively deep potential energy well for the global minimum.

The conclusion of this analysis is the confirmation of the hydrogen bond as the leading force in the aggregation between polar amino acids and mC and the importance of the side chain in the process of aggregation.

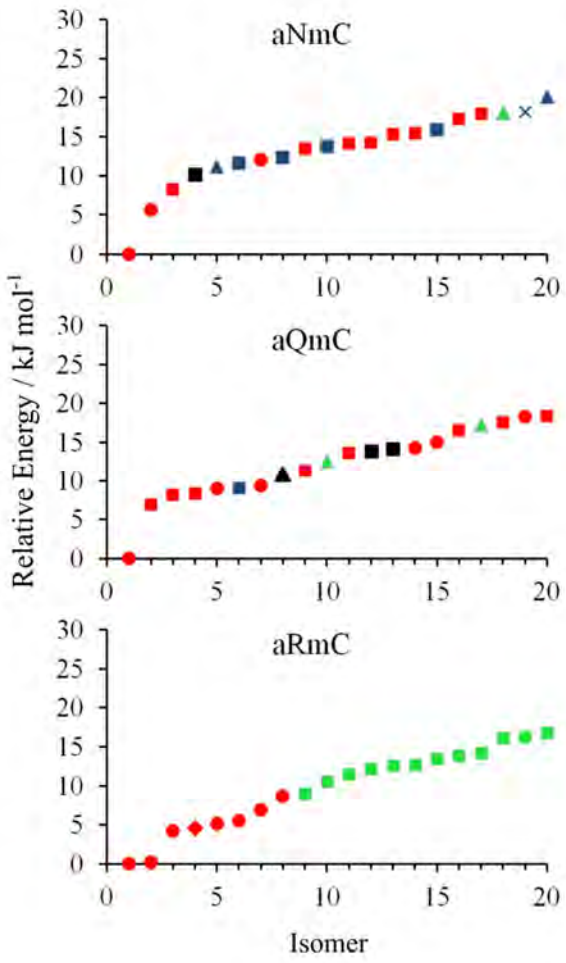


Figure 5.16 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aNmC, aQmC and aRmC. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

Polar amino acids-mT

The number of structures subjected to full optimization for polar-mT aggregates was 55 for aNmT, 48 for aQmT and 44 structures for aRmT. The most stable structure of each aggregate may be found in **Figure 5.17**, with the corresponding value of the binding energy. The most relevant

interactions with their distances and the NCI plots of these structures are also shown.

Once more, hydrogen bonds are the main interaction between the molecules. Their distance and angles are collected in **Table 5.11** and the values of the corresponding electron density are collected in **Table 5.12**. The two hydrogen bonds in aNmT are established between the lateral chain and the base, as exhibit in **Figure 5.17**. An interaction in aNmT between the N5H of the backbone and O=C4 of the base is clearly observed, but the distance is slightly large to be considered as a hydrogen bond. Accordingly, NCIplot colored that region of the space in light blue, which is an indicator of a vdW interaction. aQmT aggregate presents two intermolecular hydrogen bonds, one between the peptide bond and the base and the other between the polar side chain and mT.

Table 5.11. *Distances (Å) and angles (º) of the intermolecular interactions observed in aNmT1, aQmT1 and aRmT1.*

	Interaction type					
	H bond			**vdW**		
	C1=O‥ ‥HN_3	**R:O(C)‥ ‥HN3**	**R:H(NH)‥ ‥O=C2**	**N5H‥ ‥O=C4**	**C3H‥ ‥N1H**	**R:C=O‥‥π**
aNmT1		1.816/168.6	2.040/155.6	2.210/140.6		
aQmT1	2.032/144.7		1.987/151.0		2.479/159.0	~3.000/-
aRmT1	1.809/165.6		2.015/145.0	2.130/132.6		
			2.107/146.9			

Table 5.12. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aNmT1, aQmT1and aRmT1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type		
	H bond		
	C1=O‥ ‥HN_3	**R:O(C)‥ ‥HN3**	**R:H(NH)‥ ‥O=C2**
aNmT1		0.0324	0.0197
aQmT1	0.0194		0.0216
aRmT1	0.0313		0.0217
			0.0174

In the case of the aRmT complexes, two hydrogen bonds between the side chain and the base, as in the case of mC complexes, and between the peptide bond and the base are readily observed. An important attractive van der Waals interaction was also found between the H5N of the backbone and the O=C4 of the mT, which cannot be considered as a hydrogen bond due to the relative small angle, although it presents a short interaction distance. In fact, NCI represents this interaction with a mid-blue color, and therefore it may be mid-way between a true hydrogen bond and a vdW interaction.

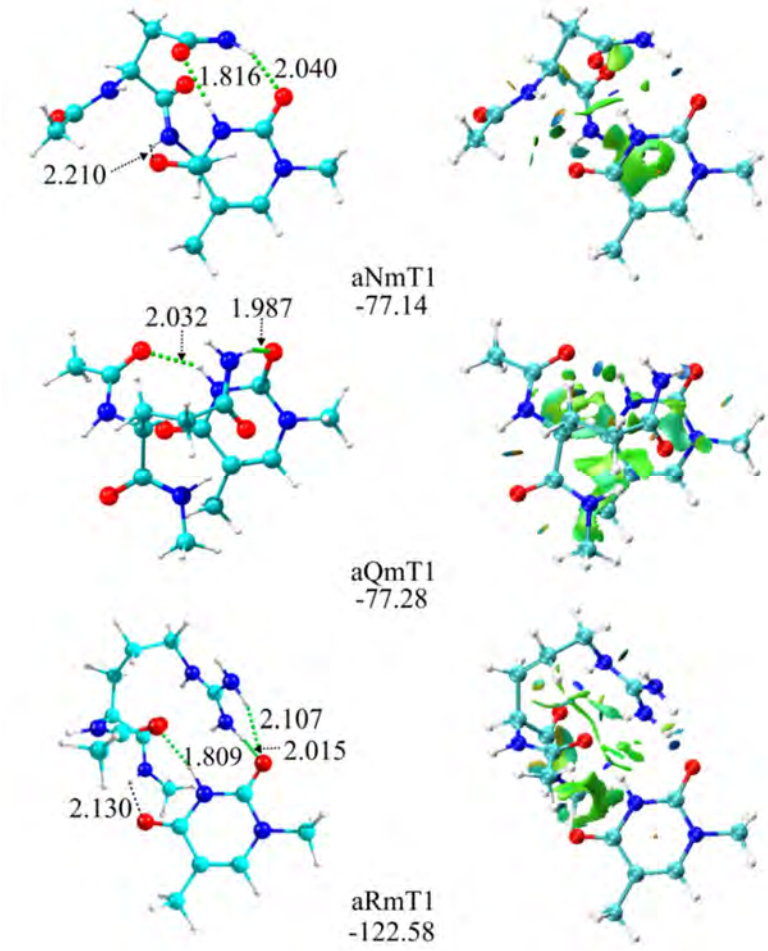


Figure 5.17 *The most stable structure found for aNmT, aQmT and aRmT complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

A deeper analysis of the NCI peaks shows that the hydrogen bonds between the lateral chain and the base are stronger than those formed between the peptidic bond and the base, with electron density values of 0.0324 and 0.216 a.u. Or put in another words, the side chain–base interaction contributes to a significant extend to the total binding energy. Likewise, the double NH···O=C hydrogen bond in aRmT accounts for a significant part of the total interaction energy. Therefore, the importance of the lateral chain in the aggregation of the polar amino acids and pyridiminic bases is undoubtedly proven.

Figure 5.18 illustrates the analysis of the dispersive contribution to the total binding energy, which varies significantly between the three systems. aNmT exhibits lower contribution than aQmT, due to the interactions in the latter between the C=O group of the side chain and the π cloud of the base, but higher than in aRmT, in which the presence of an additional hydrogen bond substantially increases the binding energy. Comparison with the results from previous mT aggregates shows that the change in the hydrogen bonding pattern (from one hydrogen bond and a dispersive interaction of the ring in non-polar aggregates to at least two hydrogen bonds) is traduced in a lower importance of the contribution of dispersive forces, as it can be observed in the **Figure 5.18**, compared to **Figure 5.5**, and in stronger binding energies. Furthermore, the

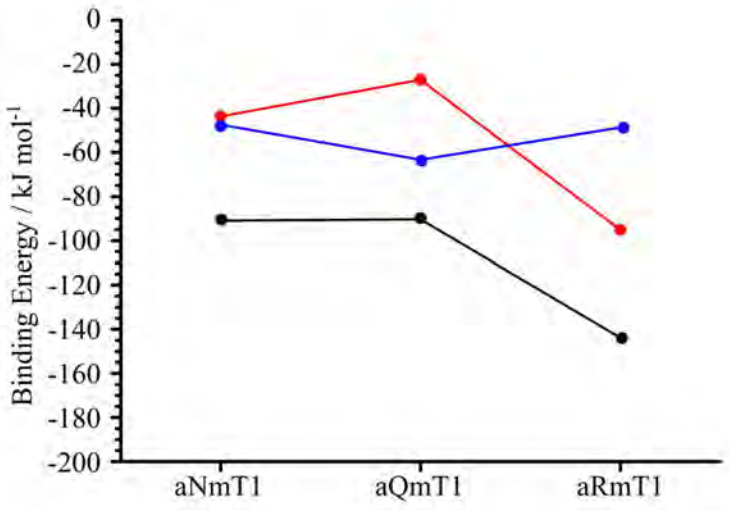


Figure 5.18 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aNmT, aQmT and aRmT.*

contribution of dispersive forces is bigger than in polar-mC aggregates, due to the reduction from three to two hydrogen bonds in aN and aQ and from four to three in the case of aR.

Finally, **Figure 5.19** displays the non-covalent leading interactions in the 20 most stable structures of each aggregate. The study reflects, as in the case of the polar - mC aggregates, the same trends of hydrogen bonding contacts: the base with the peptide bonds, with the lateral chain

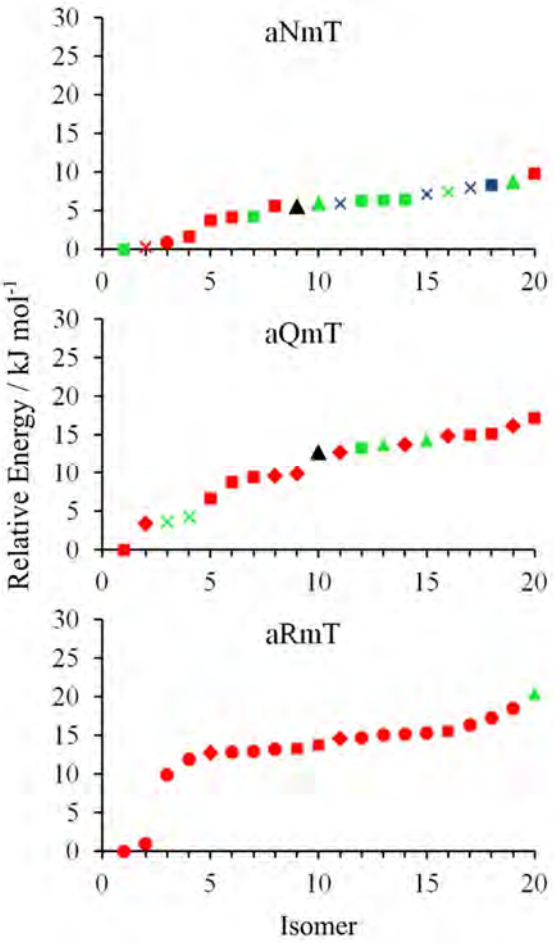


Figure 5.19 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aNmC, aQmC and aRmC. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

and the combination of both. In fact, ~88% of the structures exhibits at least one hydrogen bond between the interacting molecules, a percentage slightly lower than in mC aggregates (due to the presence of the methyl group, which hampers formation of the hydrogen bond). However, this high value reinforces the idea that hydrogen bonds are the leading force in the polar amino acid - DNA base aggregation.

Furthermore, focusing in the region near the energy minimum, the side chain plays a main role in the interaction with the base. aQmT and aRmT are governed by hydrogen bonds with the side chain, combined with additional ones with the peptidic bonds. In the case of the aNmT, the minimum exhibit only hydrogen bonds with the side chain. It is worthy to mention that, despite that the second most stable structure is governed by vdW interactions with the peptidic bond (although a vdW interaction with the side chain is also observed), the panorama in this region is similar to that in the other polar amino acids – mT aggregates, exhibiting hydrogen bonds with the peptidic bond and the side chain. Altogether, the importance of the lateral chain in the aggregation process is clearly underlined, making it the key of the interaction between polar amino acids and mT.

5.1.2.2 Polar amino acids – puric DNA bases

Polar amino acids-mA

The dimers of polar amino acids – mA present similar complexity as those analyzed in previous sections. Hence, after clustering the conformers obtained by MMs into families, a total of 42, 45 and 47 structures of aNmA, aQmA and aRmA were selected for full optimization by QM. **Figure 5.20** shows a ball & stick representation of the most stable structure found for each aggregate, together with the value of their binding energy, the most relevant interactions with their distance and the NCI plots. The distance and angles of these interactions were collected in **Table 5.13**.

Conversely to what was observed for non-polar amino acids – mA, the systems in **Figure 5.20** do not present a common trend. Aggregation of aNmA is governed by the two interactions highlighted in **Figure 5.20**, C4=O···H-NH and R-NH···N7, which accordingly to the NCI plots are not hydrogen bonds but due to dispersive forces. It is worthy to note that the side chain plays an important role in the intermolecular interactions, as highlighted by its close contact with the base. On the other hand, the most

stable structure of aQmA exhibits one hydrogen bond and a strong vdW interaction with the side chain, as observed in the previous polar amino acids aggregates. The values of the electron densities may be found in **Table 5.14**.

Table 5.13. *Distances (Å) and angles (º) of the intermolecular interactions observed in aNmA1, aQmA1 and aRmA1.*

	Interaction type							
	H bond			vdW				
	N5H·· ··N7	C4=O·· ··H(NH)	R:H···N1	N5·· ··H(NH)	C4=O·· ··H(NH)	R:H···N7	O···π	N···π
aNmA1					2.374 /117.9	2.275 /143.7	~2.700 /-	
aQmA1	1.920 /163.6			2.350 /159.3				~3.300 /-
aRmA1		1.970 /167.9	2.112 /146.1					
			2.073 /150.7					

Table 5.14. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aNmA1, aQmA1 and aRmA1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type		
	H bond		
	N5H···N7	C4=O···H(NH)	R: H···N1
aNmA1			
aQmA1	0.0312		
aRmA1		0.0219	0.0236
			0.0211

Finally, the two molecules forming aRmA are held together by three hydrogen bonds, exhibiting the characteristic double hydrogen with the side chain, which results in a significantly higher binding energy. Although it could seem that the interaction pattern was different, the aggregation always tries to maximize the contacts with the side chain,

even when no H-bonds are established. Furthermore, aQmA and aRmA exhibited higher values of electron density (0.0312 and 0.0219 a.u.) than the previous mA aggregates, due to the formation of stronger hydrogen bonds.

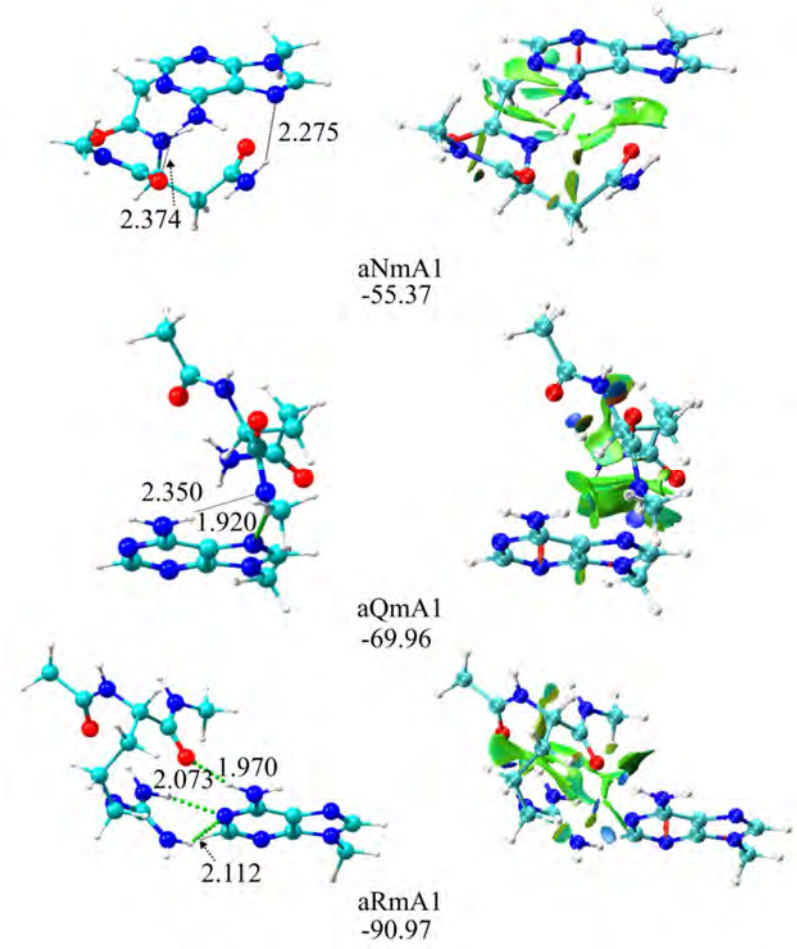


Figure 5.20 *The most stable structure found for aNmA, aQmA and aRmA complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

NCI plots also highlight the variable importance of the π···π interactions along the series of three amino acids, as the wide green zones and the analysis of the dispersive contribution displayed in **Figure 5.21** show: the contribution of the dispersive force decreases in importance

from aNmA1 to aRmA1, as the number of intermolecular hydrogen bonds increases from zero in aNmA1 to one in aQmA1 and three in aRmA1. More precisely, the increase in hydrogen bonds formation results in a decrease of the contribution of the dispersive forces to the total binding energy, from 92% in aNmA to 62% in aQmA and to a mere 29% in aRmA. The general trend in the previous sections was an increase in binding energy from systems containing non-polar amino acids to those formed by polar amino acids. However, such a general trend is lost in aNmA, which presents similar binding energy values to those formed by non-polar amino acids and mA,

Regarding the general tendency of the systems studied, **Figure 5.22** depicts the energetic order of the first 20 computed structures, with the leading interactions highlighted by the symbols and the color code.

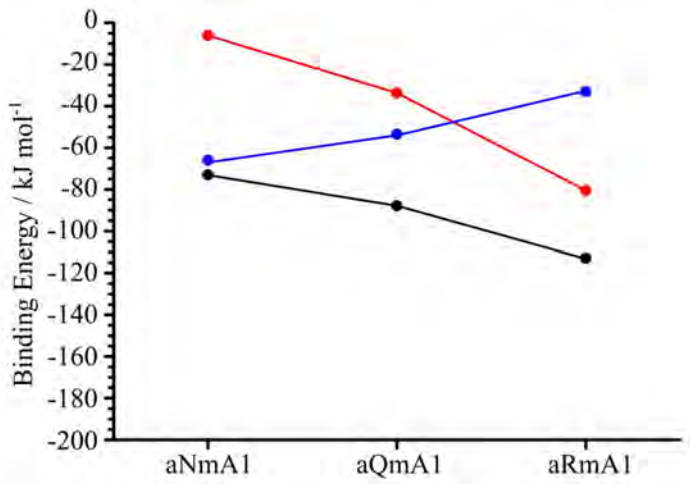


Figure 5.21 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aNmA, aQmA and aRmA.*

All the structures calculated for aNmA up to 12 kJ mol^{-1} follow two tendencies: absence of hydrogen bonds, which includes the global minimum, or presence of one hydrogen bond with the side chain. Both trends are governed by dispersive forces, which account for most of the interaction energy. Furthermore, the energy diference between aNmA1 and aNmA2 is ~5 kJ mol^{-1} and points to a relatively deep potential energy well for the global minimum. A similar conclusion may be reached for

aQmA, where only four structures were found in ~17 kJ mol^{-1}. Despite that the leading force is one hydrogen bond with the peptidic chain for aQmA1, aQmA2 and aQmA4, the side chain is positioned in such a way that it can interact with the electron density of the mA. In the case of aQmA3, three hydrogen bonds are observed, two of them with the side chain. The absence of competition with other families of structures reinforces the conclusions reached for the dominant families: the importance of the side chain in the interaction with mA.

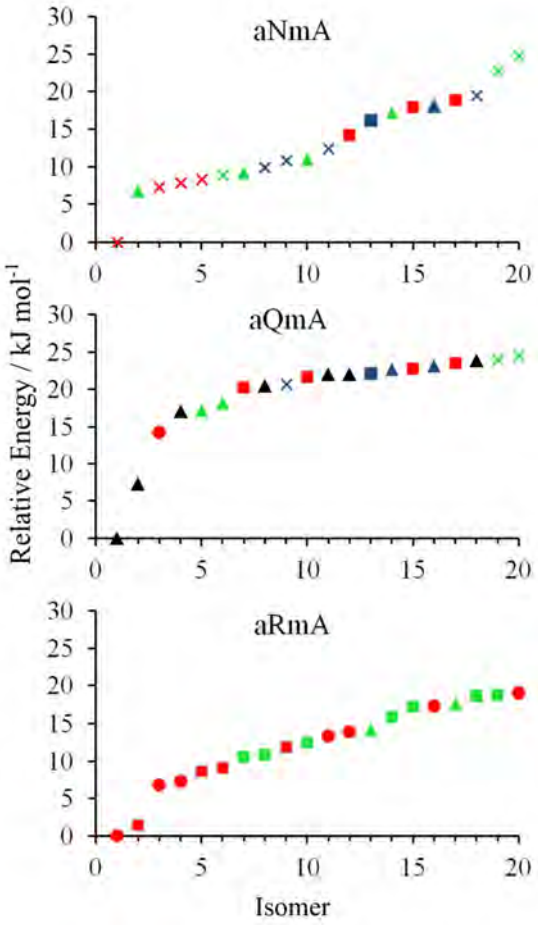


Figure 5.22 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aNmA, aQmA and aRmA. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

Finally, a competition between two kinds of structures seems to be present in aRmA, as the second most stable conformer is almost isoenergetic with the global minimum. Nevertheless, the side chain takes part in the interaction in both cases. The structure of both conformers may be found in **Figure 5.23**, with the intermolecular hydrogen bonds highlighted in blue or green depending on the part of the amino acid involved in the interaction. Thus, they present one hydrogen bond with the peptidic chain and one/two hydrogen bonds with the side chain, of the type NH···N···NH or with a single NH···N.

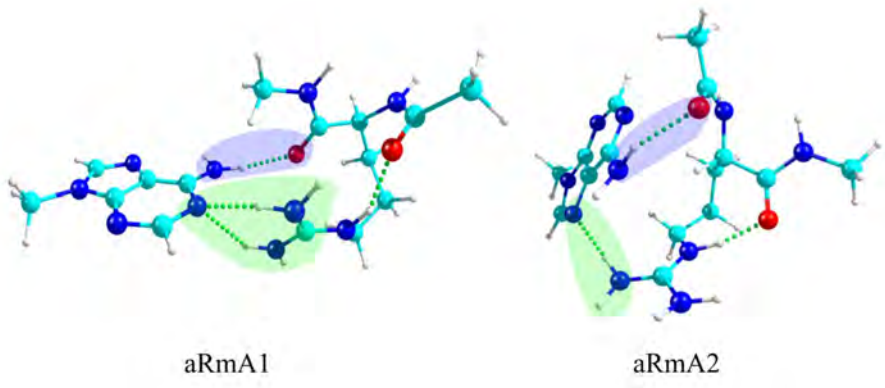


Figure 5.23 *Structure of aRmA1 and aRmA2. The intermolecular hydrogen bonds with the peptidic bond was highlighted in blue and with the side chain was highlighted in green.*

In summary, the central role played by the side chain as an essential component in the aggregation of polar amino acids and mA was confirmed. On the other hand, the contribution of dispersive forces is more important than expected, but hydrogen bond is still the leading force in two of the three systems studied.

Polar amino acids-mG

Once the structures found by MM were grouped into families using XCluster, a set of 26 structures of aNmG, 48 of aQmG and 52 of aRmG were considered for the study using DFT. The structure of the most stable polar amino acid-mG aggregate may be found in **Figure 5.24**, together with the value of their binding energy. The most relevant intermolecular interactions with their distance were also included. Additionally, the NCI plots of these structures were included. Distances and angles and electron density values may be found in **Tables 5.15** and **5.16**.

Table 5.15. *Distances (Å) and angles (º) of the intermolecular interactions observed in aNmG1, aQmG1 and aRmG1.*

	Interaction type						
	H bond						
	C4=O·· ··H(NH)	N5H·· ··O=C6	C4=O·· ··HN1	R:H·· ··O=C6	R:O(C)·· ··H(NH)	R:O(C)·· ··HN1	R:H(NH)·· ··N7
aNmG1	2.014 /145.9		1.930 /147.2	1.883 /164.1			
aQmG1		1.888 /159.8			2.202 /141.1	1.847 /147.2	
aRmG1	1.867 /155.7						2.016 /150.3

Table 5.16. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aNmG1, aQmG1 and aRmG1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type						
	H bond						
	C4=O·· ··H(NH)	N5H·· ··O=C6	C4=O·· ··HN1	R:H·· ··O=C6	R:O(C)·· ··H(NH)	R:O(C)·· ··HN1	R:H(NH)·· ··N7
aNmG1	0.0206		0.0245	0.0263			
aQmG1		0.0275			0.013	0.0307	
aRmG1	0.0247						0.0253

As it was already observed for non-polar amino acids – mG aggregates, the aggregation process is strongly conditioned by the -NH-CO- group of the mG structure, which is an excellent anchoring point. In non-polar amino acids, the side chain was sitting away from the DNA base. Conversely, in this case, the side chain interacts directly with the mG through hydrogen bonds, as **Figure 5.24** shows. aNmG presents three hydrogen bonds, being the shortest and strongest (1.883 Å and 0.0263 a.u.) the one formed between the side chain and the base. aQmG presents three hydrogen bonds too, but in this case two of them are established with

the lateral chain. The shortest one presents an electron density of 0.0307 a.u., an indication of the high strength of this interaction. Finally, aRmG exhibits only two hydrogen bonds, but the strongest one is formed between the side chain and the base, as highlighted also by the NCI study.

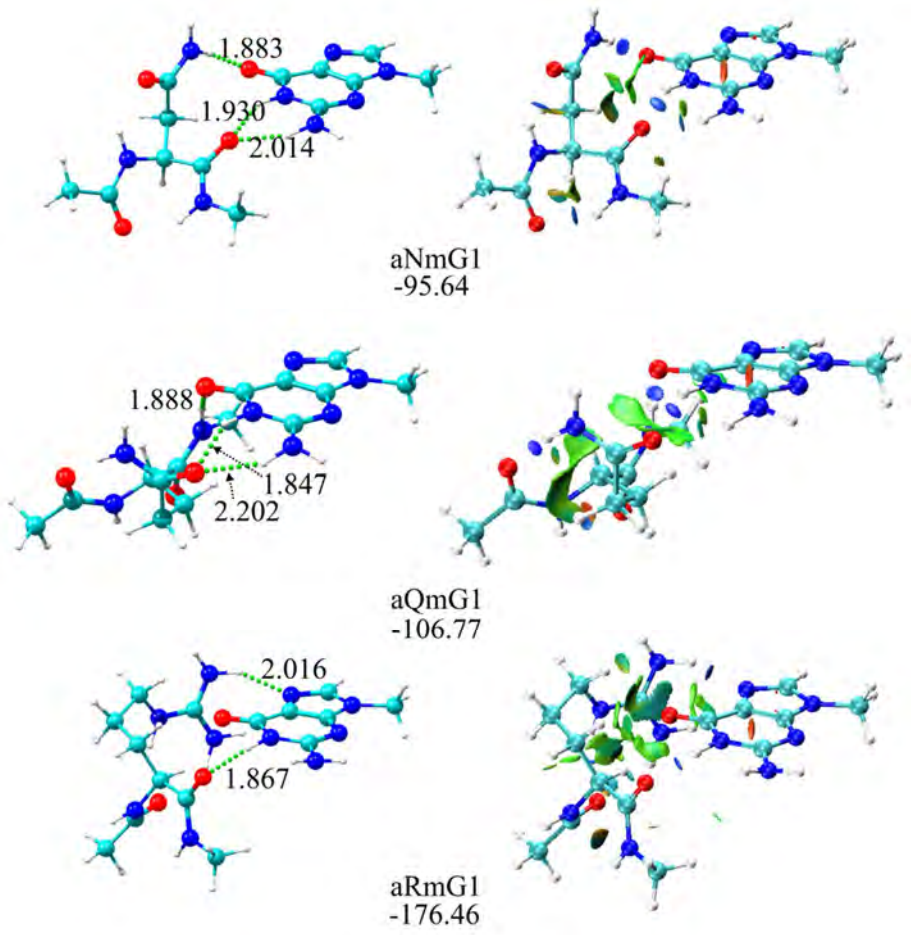


Figure 5.24 *The most stable structure found for aNmG, aQmG and aRmG complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

The analysis of the contribution due to dispersive forces is displayed in **Figure 5.25**. As in the case of mG aggregates, the contribution of these forces is small and almost constant and around 24-28% for the three aggregates analyzed. On the other hand, the binding energies values are

similar to those for polar amino acid – mC systems, due to the decisive role played by the hydrogen bonds. In fact, a similar rise as in aRmC in the binding energy was observed in aRmG due to the interaction between the charge in the side chain and the electronegative C=O moiety of mG.

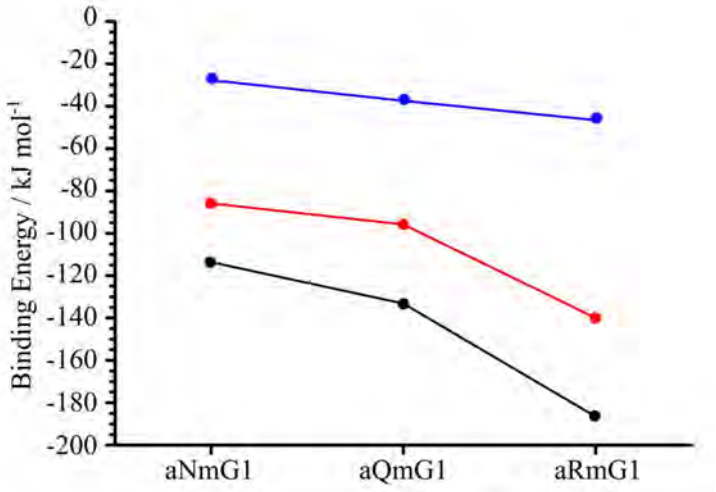


Figure 5.25 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aNmG, aQmG and aRmG.*

Figure 5.26 shows the study of the non-covalent leading interactions found in the 20 most stable structures of each complex. This study stresses the importance of the hydrogen bond as the prominent aggregation force, as in the case of mC complexes. In fact, all the studied structures present at least one hydrogen bond, and ~99% of the structures presents two or more hydrogen bonds.

Analysis of the region close to the minimum shows several differences between the systems. The three most stable structures for aNmG exhibit the same trend: three hydrogen bonds with the side chain and the backbone, while the first structure with a different trend appears around ~9 kJ mol^{-1}. For aQmG aggregates, the panorama is more complex. At least four different aggregates are observed in ~2 kJ mol^{-1}, as depicted in **Figure 5.27**, presenting between one and four hydrogen bonds. On the one hand, aQmG1, aQmG3 and aQmG4 exhibit hydrogen bonds between the electronegative groups of the peptide bonds and the side chain with mG. On the other hand, aQmG2 exhibits a different hydrogen bond pattern

with the backbone members. However, vdW interactions with the side chain are also observed in this structure and their contribution is relevant. Thus, despite the leading interaction for aQmG2 takes place with the peptide bond, the side chain of the glutamine also plays a relevant role in the interaction.

Finally, most of aRmG exhibits two or three hydrogen bonds with the side chain and the peptide bond in all the structures below ~18 kJ mol^{-1}. Hence, the observations made for the aRmG1 and aRmG2 are also valid for the rest of conformers in a wide energy window, confirming the importance of the side chain in the interaction with the mG.

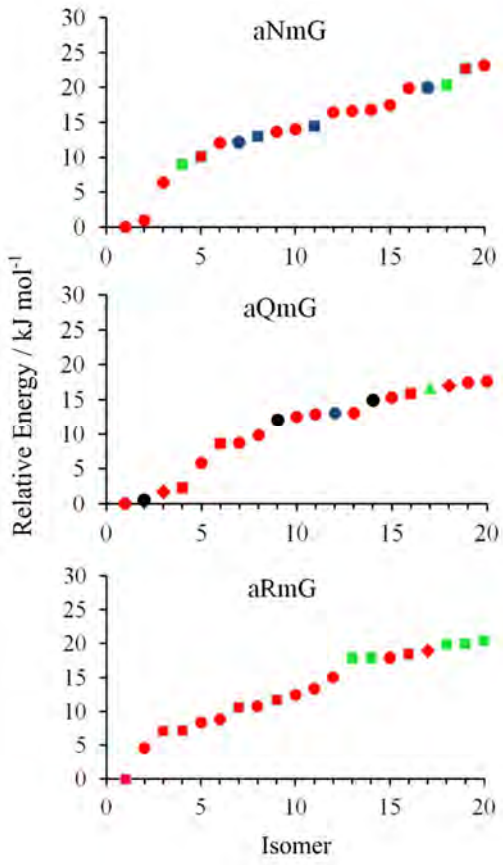


Figure 5.26 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aNmG, aQmG and aRmG. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

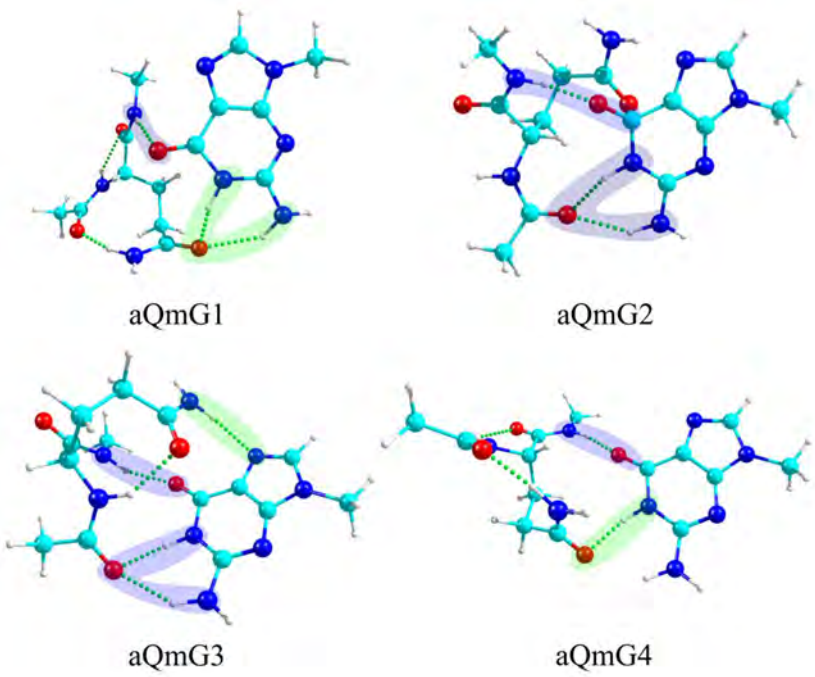


Figure 5.27 *Structure of the four lowest conformers of aQmG aggregates. The intermolecular hydrogen bonds with the peptidic bond was highlighted in blue and with the side chain was highlighted in green.*

In summary, the importance of the hydrogen bond as the leading force in the aggregation between polar amino acids and mG is evident, as in the case of the mC aggregates. The contribution of dispersive interactions is small as a result of the hydrogen bond pattern, and the importance of the side chain in the process of aggregation has been undoubtedly proven.

5.1.3. Aromatic amino acids – DNA base

5.1.3.1 Aromatic amino acids – pyrimidinic DNA bases

Preliminary exploration of the conformational landscape for aromatic amino acids in chapter 4 resulted in 29, 46 and 36 structures for aF, aW and aY respectively, in 30 kJ mol^{-1} of energy window.

Aromatic amino acids-mC

The main difference between these systems and those studied in previous sections is that the aromatic ring increases significantly the importance of the dispersive forces and new π···π interactions are possible, whose potential energy surfaces are shallow and result in a large number of conformers of very close stability. To take into account a representative number of structures, 54 conformers of aFmC, 82 of aWmC and 92 of aYmC were fully optimized. The most stable structure of each aggregate may be found in **Figure 5.28**, and the most relevant parameters were collected in **Table 5.17**.

Table 5.17. *Distances (Å) and angles (º) of the intermolecular interactions observed in aFmC1, aWmC1 and aYmC1.*

	Interaction type					
	H bond				vdW	
	C1=O·· ··H(NH)	N5H·· ··O(C)	C4=O·· ··H(NH)	N2H· ··O(C)	C3H···N3	R: π···π
aFmC1			1.916/163.7	1.928/171.5		~3.200/ -
aWmC1	1.874/166.4	1.864/165.6			2.307/158.4	~3.200/ -
aYmC1			1.909/160.8	1.946/171.0		~3.200/ -

As **Figure 5.28** shows, the backbone of the three aromatic amino acids establishes two hydrogen bonds with the electronegative groups (C2=O and NH_2) of mC, as the deep blue zones highlight in NCI plots. **Table 5.18** displays the electron density values (a. u.) of the hydrogen bond observed for each aggregates. Clearly, aWmC structures present shorter and stronger hydrogen bonds than the other aggregates, also in agreement with the electron density values.

However, the main observation is the appearance of an additional interaction of π···π character between the lateral chain and the DNA base, as NCI plots display with wide green surfaces between the electron density cloud of the aromatic lateral chain and mC. The same interaction can be observed in the three global minima. Such interaction requires the peptidic chain to approach the base, enabling the formation two hydrogen bonds.

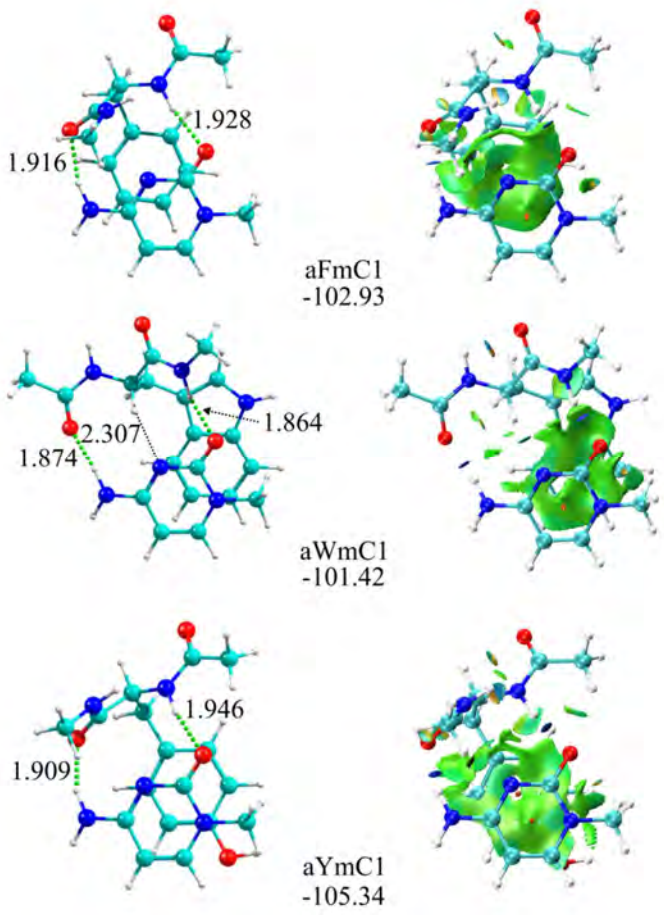


Figure 5.28 *The most stable structure found for aFmC, aWmC and aYmC complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

The analysis of the dispersive contribution by Grimme's D3 approach is shown in the **Figure 5.29**. As it can be seen, it is similar for all three systems (51-62% of the total binding energy) and more important than in polar and non-polar amino acids – mC aggregates, due to the stacking interaction between the lateral chain and the base. Thus, the small differences in dissociation energy values between the three systems are mainly due to the substituents. Taking phenylalanine as the base, the addition of a hydroxyl group to form tyrosine results in the strongest binding energy, due to an extra interaction between the hydroxyl moiety

and the mC. On the other hand, the double ring of tryptophan expands the electronic cloud, which also results in slightly stronger π···π interactions.

Table 5.18. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aFmC1, aWmC1 and aYmC1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type			
	H bond			
	C1=O···H(NH)	N5H···O(C)	C4=O···H(NH)	N2H···O(C)
aFmC1			0.0249	0.0257
aWmC1	0.0260	0.0294		
aYmC1			0.0256	0.0246

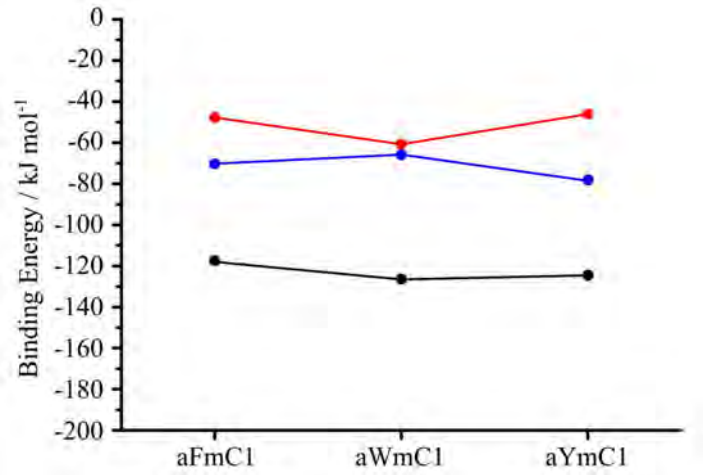


Figure 5.29 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aFmC, aWmC and aYmC.*

Comparison with previous mC aggregates highlights the influence of the aromatic side chain in the binding energy. Despite having a similar hydrogen bond pattern than non-polar amino acids – mC, the aromatic amino acids exhibit significantly higher binding energy, of the order of

those found in polar aNmC and aQmC complexes. The reason has to be the extra stabilization introduced by the π···π interactions. In order to quantify it, a comparison of the binding energies of the global minima with the first structure of each aggregate that presents the same interaction pattern, but without these π···π interactions, conformers aFmC6, aWmC5 and aYmC7, may be found in **Figure 5.30**. The values of the binding energy point to a contribution of the π···π interactions around 12.5-16 kJ mol^{-1} for each conformer.

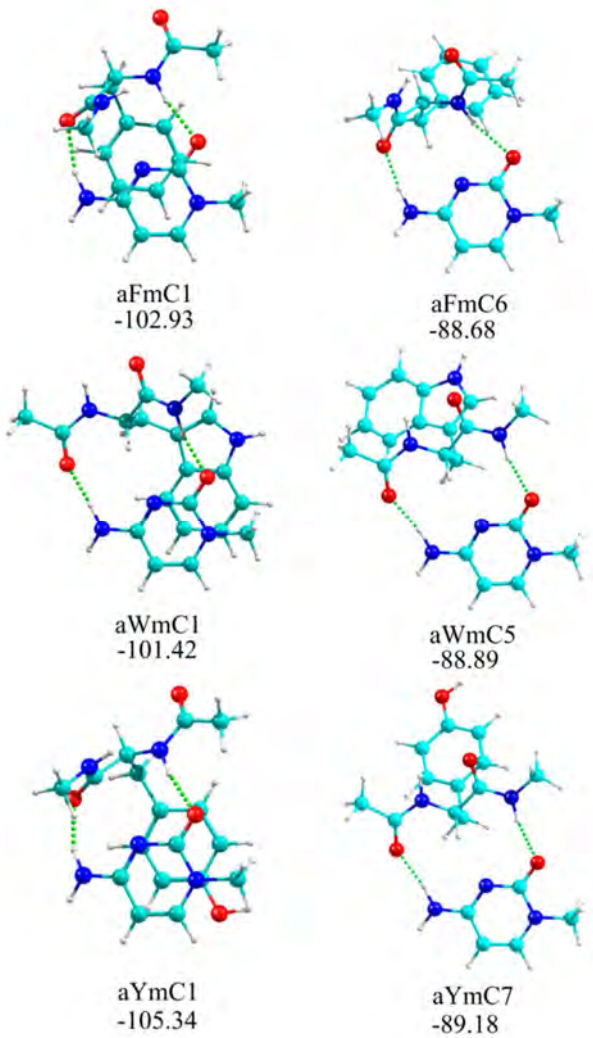


Figure 5.30 *The most stable structure and the first structure without π···π interactions of aFmC, aWmC and aYmC aggregates, with the values of their binding energies in kJ* mol^{-1}.

The analysis of the leading forces was applied to the 20 most stable structures. The results may be found in **Figure 5.31**. The figure highlights the trend already observed: π···π interactions plus hydrogen bond formation with the backbone, depicted in black. The study also remarks the co-dominance of both types of interactions: ~99% of the structures present at least one hydrogen bond. However, a high number of these structures present π···π interactions between the side chain and mC as well.

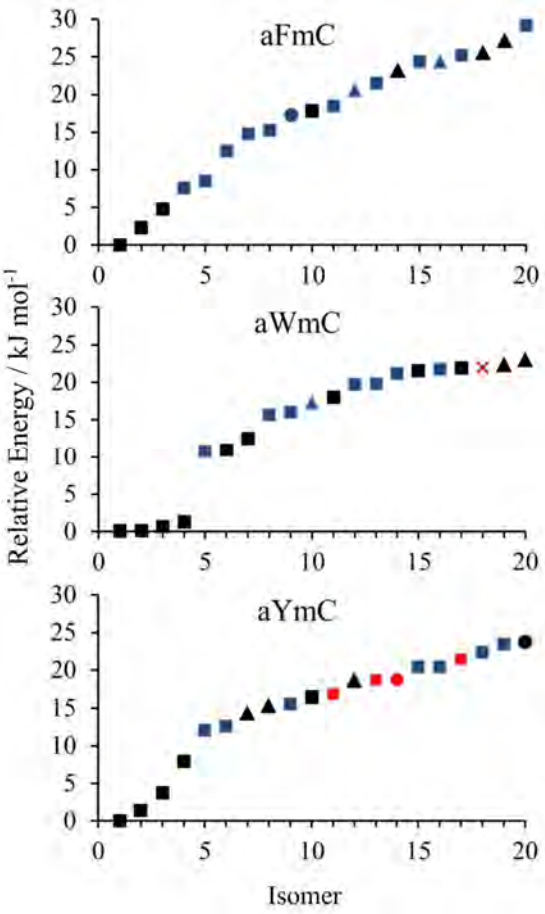


Figure 5.31 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aFmC, aWmC and aYmC. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

It is also worthy to note that all the structures in the low-energy region present the same trend. Thus, the first structure with a different hydrogen bond pattern appears above 15 kJ mol^{-1} in all three systems. Furthermore, if the position of the side chain is analyzed, the trend of π···π interactions with the side chain is observed in the global minimum of each group and in the conformers near them. The first structure without these interactions appears around 7 kJ mol^{-1} for aFmC and above 10 kJ mol^{-1} for aWmC and aYmC, pointing to a dominance of the family of structures of the global minimum.

In summary, the conformational landscape of these systems is dominated by π···π interactions with the aromatic lateral chain and hydrogen bonds between the peptidic skeleton and mC, in a situation of co-dominance of the two interactions.

Aromatic amino acids-mT

The initial stages of conformational landscape exploration, clustering into families and selection of structures resulted in 66 structures for aFmT, 79 for aWmT and 100 for aYmT to be subjected to full optimization using DFT. The most stable structure of aFmT1, aWmT1 and aYmT1 may be found in **Figure 5.32**, together with their binding energy, while distances and the angles are collected in **Table 5.19**.

Table 5.19. *Distances (Å) and angles (°) of the intermolecular interactions observed in aFmT1, aWmT1 and aYmT1.*

	Interaction type				
	H bond			vdW	
	N2H·· ··O=C2	N2H·· ··O=C4	C4=O·· ··HN3	R: π···π	R: OH···π
aFmT1		2.001/153.1	1.811/170.7	~3.100/ -	
aWmT1	1.963/153.4		1.848/166.6	~3.000/ -	
aYmT1	2.070/151.9		1.800/164.9	~3.100/ -	~2.800/ -

Figure 5.32 shows that these three systems also follow the same tendency towards formation of π···π interactions with the side chain, which guide the peptidic bond members to establish hydrogen bonds with mT. In fact, they present the same interaction pattern than aromatic - mC

aggregates, in contrast with non-polar and polar amino acids, which exhibited more contacts with mC than with the mT (except aR). NCI plots confirm these two different and simultaneous interactions: deep blue surfaces mark the hydrogen bond and a wide green area between the aromatic rings. The values of the electron density are collected in **Table 5.20**. They display the stronger character of the HB formed by C4=O···HN3 with values from 0.0276 to 0.0309 a.u. for the hydrogen bonds.

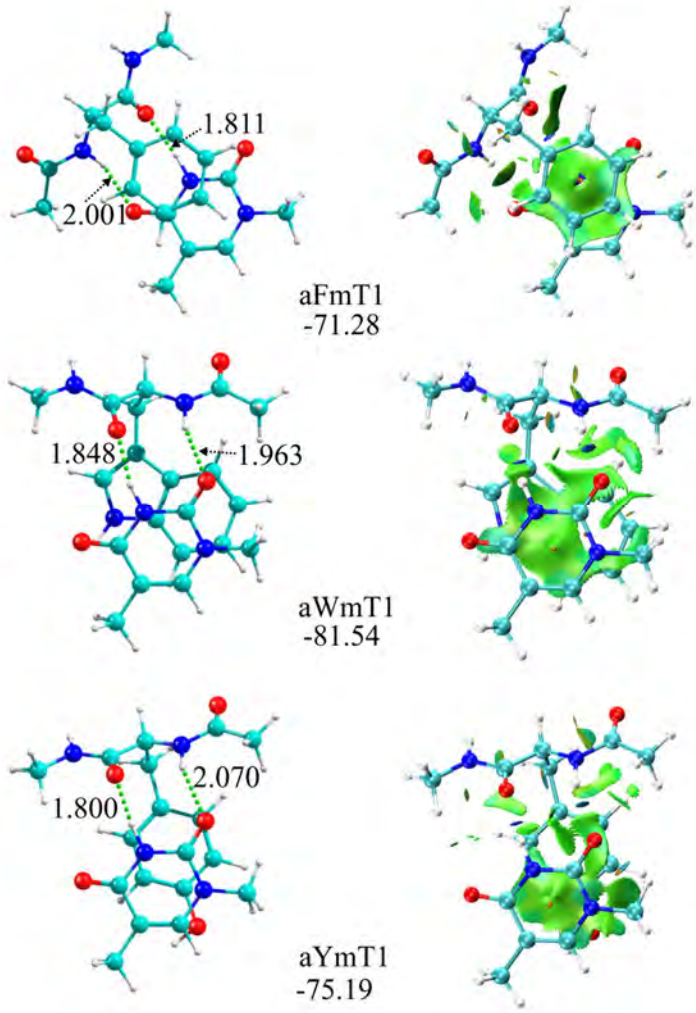


Figure 5.32 *The most stable structure found for aFmT, aWmT and aYmT complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

Table 5.20. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aFmT1, aWmT1 and aYmT1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type		
	H bond		
	N2H···O=C2	N2H···O=C4	C4=O···HN3
aFmT1		0.0216	0.0301
aWmT1	0.0230		0.0276
aYmT1	0.0184		0.0309

Figure 5.33 shows the analysis of the dispersive contribution to the interaction energy. Clearly the contribution of the dispersive force to the binding energy is similar for the three aggregates and ~71% of the total interaction energy. aWmT exhibits higher contribution than aWmC aggregate due to the larger surface of interaction.

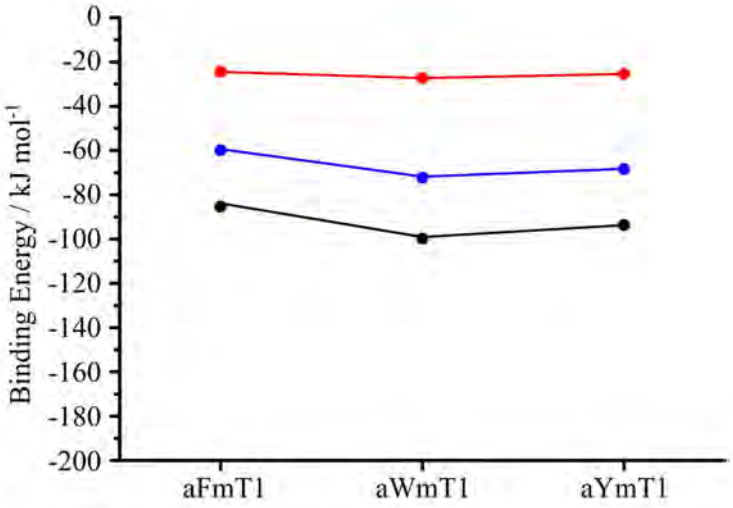


Figure 5.33 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aFmT, aWmT and aYmT.*

Comparison of these values with those from previous mT aggregates shows that aromatic mT complexes exhibit similar dispersive contribution than non-polar mT aggregates and higher than polar mT aggregates. This is due to the direct interaction between the aromatic group and the DNA base, which is due to pure dispersive forces, as NCI plots remark. Finally, the higher dispersive contribution tendency observed for mT complexes with respect to mC aggregates is maintained in comparison with the rest of pyrimidinic aggregates.

The binding energies values collected in **Figure 5.32** are higher for aWmT and aYmT than for aFmT. This is due to the stronger π···π interaction between the bicyclic ring of aW and the interaction of the OH substituent of aY with the electron density of the mT base. Moreover, as in aromatic mC complexes, the π···π interactions significantly contribute to the binding energies. The quantification of this effect is displayed in **Figure 5.34**, where a comparison between the structure and binding energy values of the global minimum and aFmT15, aWmT15 and aYmT23, the first structures with similar interaction pattern but without stacking, is offered. Small values of ~3-12 kJ mol^{-1} are found for the over-stabilization of mT aggregates, which is significantly smaller than in mC aggregates.

The study of the non-covalent leading interactions for the 20 most stable structures of each polar amino acid mT aggregates may be found in **Figure 5.35**. The study highlights the importance of the hydrogen bonds in the formation of the aggregates, which are present in all the structures. However, a number of these structures, larger than in mC aggregates, also present π···π interactions between the side chain and mT, due to the high contribution of the dispersive forces to the complex' dissociation energy.

Focusing in the region close to the global minimum, one concludes that the aggregation of aromatic amino acids and mT is guided by formation of two hydrogen bonds with the electronegative members of the backbone, because all the structures below 5 kJ mol^{-1} exhibit this pattern. Nevertheless, there are some exceptions, such as aYmT, which presents an additional hydrogen bond with the hydroxyl moiety of the aromatic ring. Altogether, it is clear that all the structures around the minimum present a similar tendency, reinforcing the conclusions already extracted.

The conclusion of the study of aromatic mT aggregates is the relevance of the π···π interactions between the aromatic side chain and the DNA base, which permits the system to establish two additional hydrogen bonds with the peptide bond members. The combination of both

interactions is crucial in the aggregation of the aromatic amino acids and mT.

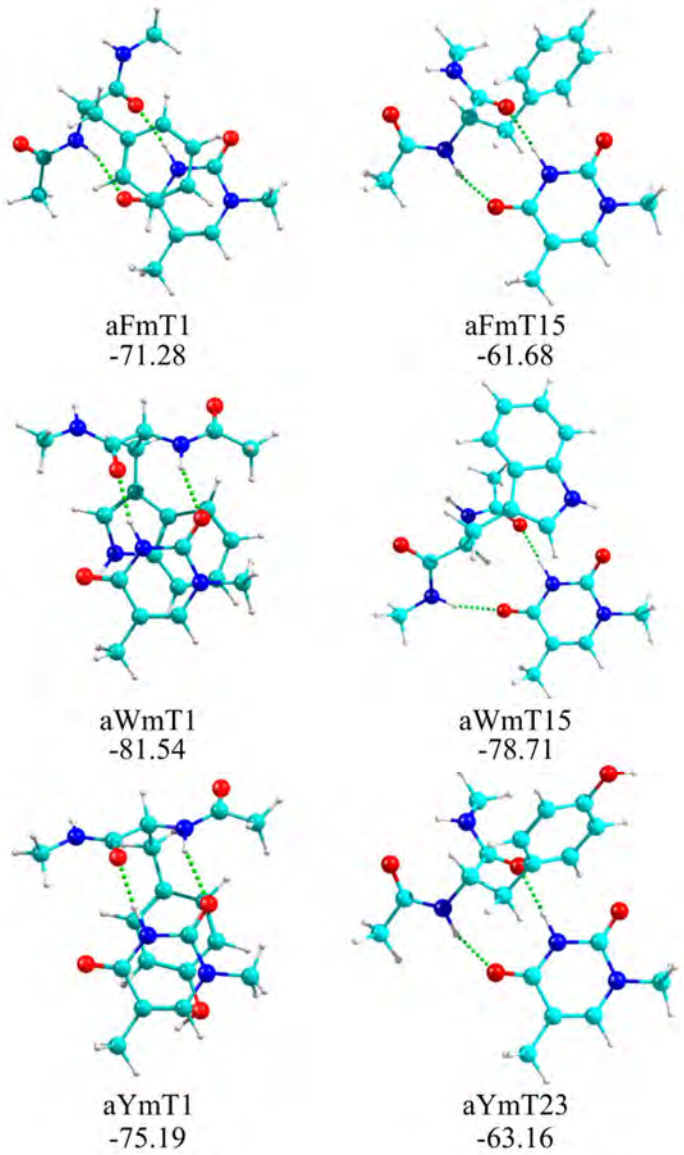


Figure 5.34 *The most stable structure and the first structure without π···π interactions of for aFmT, aWmT and aYmT aggregates, with the values of their binding energies in kJ mol^{-1}.*

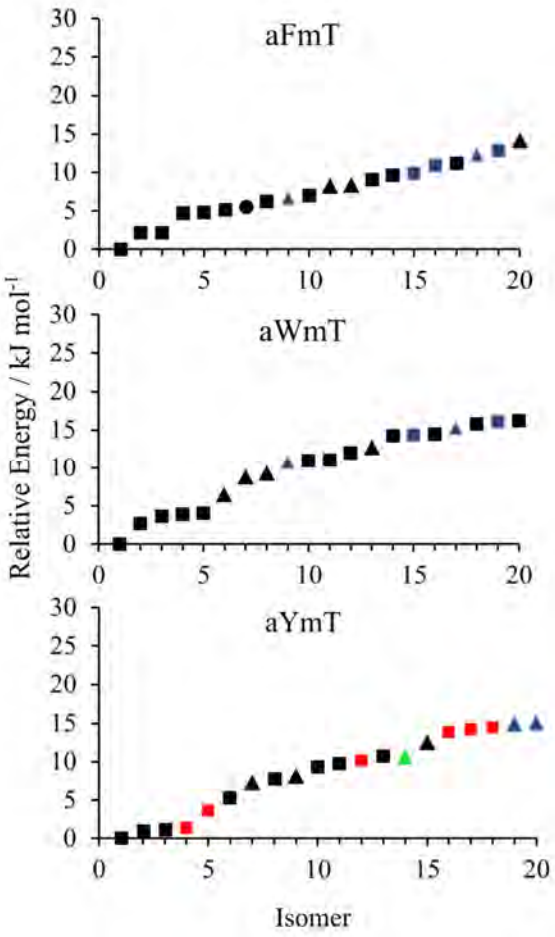


Figure 5.35 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aFmT, aWmT and aYmT. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

5.1.3.2 Aromatic amino acids – puric DNA bases

Aromatic amino acids-mA

A set of 93 conformers for aFmA, 123 for aWmA and 144 for aYmA were full optimized by DFT. **Figure 5.36** collects the most stable

structures of each system and **Table 5.21** summarizes the most important structural parameters.

As it can be seen, the interaction pattern is different from the other aggregates. aFmA exhibits only one hydrogen bond between the peptide bond and the base and relevant vdW interactions with the backbone. Furthermore, its side chain is placed perpendicularly to mA, and therefore it does not establish π···π interactions. It is a different tendency than that previously observed in aromatic amino acid aggregates. However, aFmA2, almost isoenergetic to the global minimum and within the computational error, exhibits the "canonical" interaction pattern (two hydrogen bonds and the side chain placed parallel to adenine). On the other hand, aWmA and aYmA return to the previous aromatic pattern, placing the side chain parallel to the DNA base and interacting throughout two hydrogen bonds with the amine and carbonyl groups of the backbone. NCI plots confirm this observation: aFmA presents smaller green surfaces around the aromatic side chain than the wide green zones observed for aWmA/aYmA. The values of the electron density were collected in **Table 5.22**.

Table 5.21. *Distances (Å) and angles (º) of the intermolecular interactions observed in aFmA1, aWmA1 and aYmA1.*

	Interaction type				
	H bond		**vdW**		
	C1=O·· ··H(NH)	**N5H·· ··N7**	**R: OH···N3**	**R: π···π**	**R: O···π**
aFmA1	2.055/148.4				~3.100/ -
aWmA1	2.058/140.5	2.016/161.0		~3.100/ -	
aYmA1	2.037/142.9	2.031/160.3	2.285/132.2	~3.000/ -	

The analysis of the dispersive contribution to the total binding energy is shown in **Figure 5.37** and shows similar values than in the three previous aggregates. Furthermore, the additional hydrogen bond of the aWmA1 and aYmA1 results in a smaller contribution of the dispersive forces to the total binding energy than in aFmA1.

All these evidences correlate with the binding energies values collected in **Figure 5.37.** aFmA has the lowest binding energies values of all the aromatic-mA aggregates, in the range of the non-polar amino acids

– mA aggregates, where the side chain interactions are almost absent. Nevertheless, the π···π interactions in aWmA and aYmA add some extra stabilization, as observed in the previous aromatic aggregates. The quantification of this effect may be found in **Figure 5.38**, where the structure and binding energy of the global minimum are presented together with the first structures with similar interaction pattern but without π···π side chain interactions: aWmA43 and aYmA34. The comparison allows one to extimate the contribution of the π···π interaction in ~ 24-28 kJ mol^{-1} respectively. On the other hand, despite aFmA1 does not present

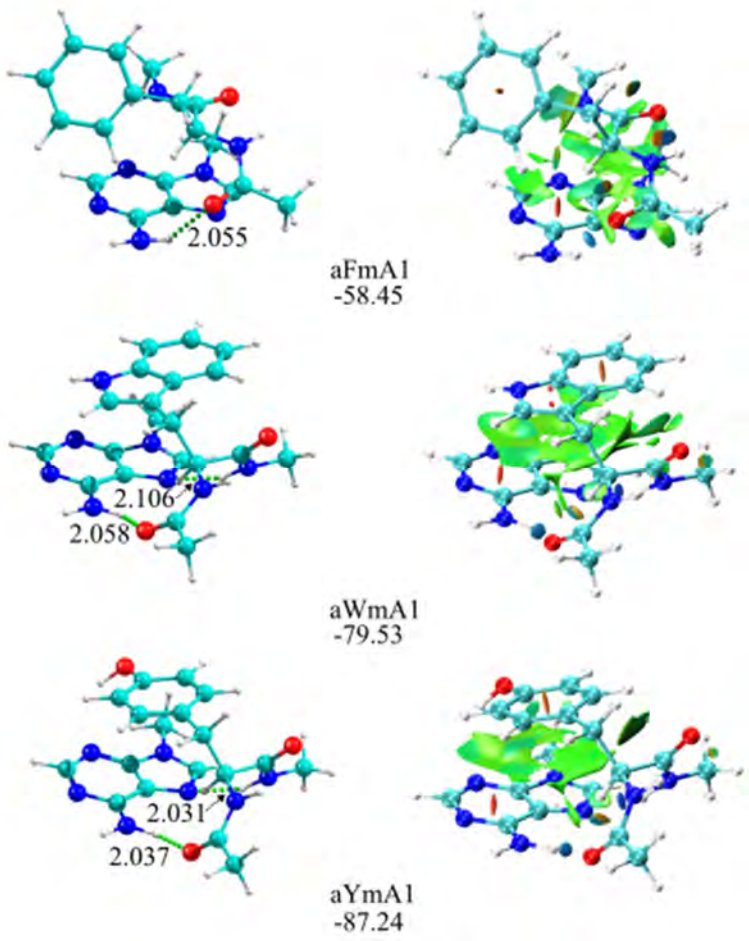


Figure 5.36 *The most stable structure found for aFmA, aWmA and aYmA complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ* mol^{-1}*) are shown.*

π···π interactions with the side chain, vdW contacts with the backbone are present, compensating the stabilization due to the π···π interaction, and resulting in similar binding energy value than aFmA11.

Table 5.22. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aFmA1, aWmA1 and aYmA1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type	
	H bond	
	C1=O···H(NH)	N5H···N7
aFmA1	0.0184	
aWmA1	0.0188	0.0240
aYmA1	0.0195	0.0233

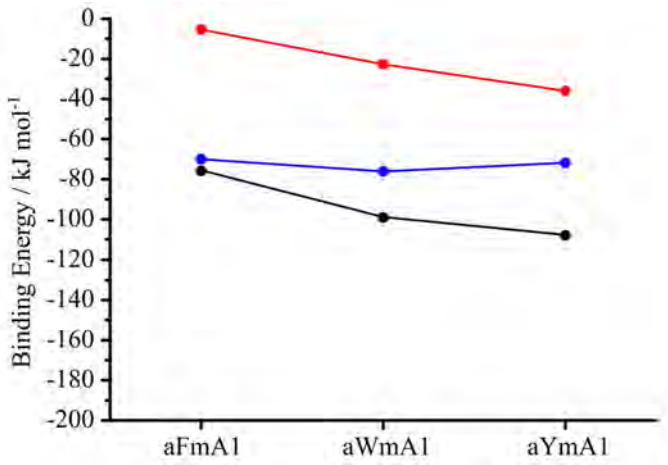


Figure 5.37 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aFmA, aWmA and aYmA.*

Regarding the general tendency of the systems studied, **Figure 5.39** depicts the energetic order of the first 20 computed structures, with the leading interactions represented using symbols and a color code.

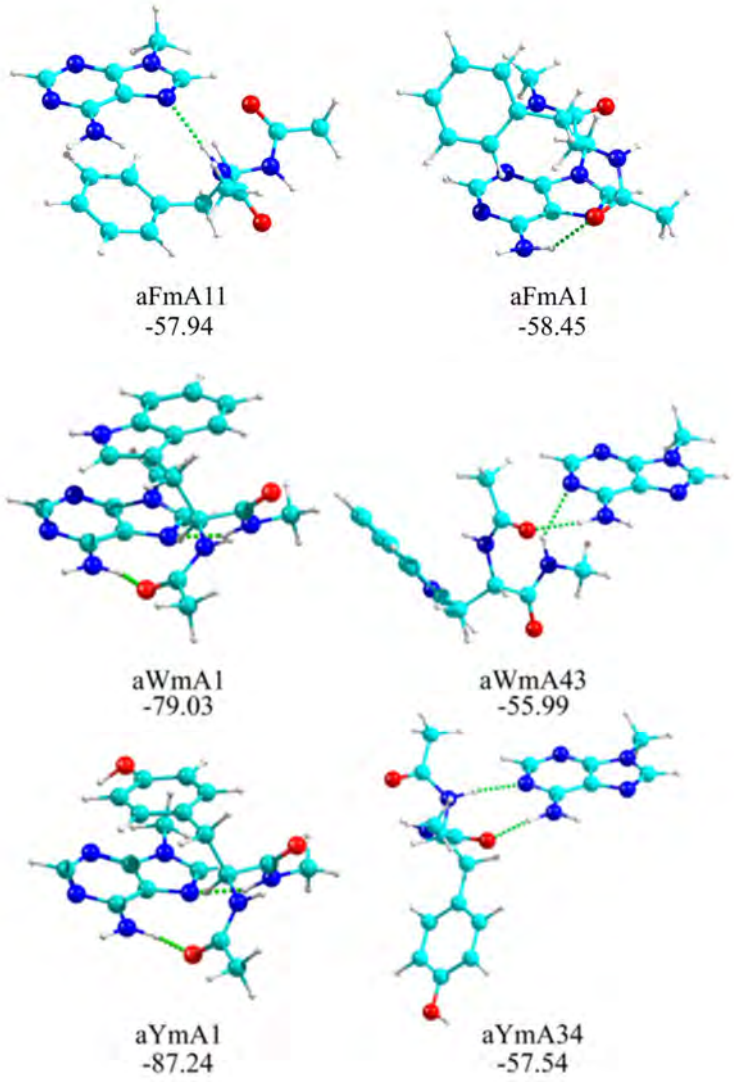


Figure 5.38 *The most stable structure and the first structure without π···π interactions of for aWmA and aYmA aggregates, with the values of their binding energies in kJ mol^{-1}.*

As it can be seen, most of the structures presents at least one HB with the DNA base, remarking the relevance of the hydrogen bonds in the aggregation process. However, as in the case of the previous aromatic clusters, the aggregation is guided by π···π interactions which bings the two molecules in close contact and permits the formation of hydrogen bonds between the two molecules.

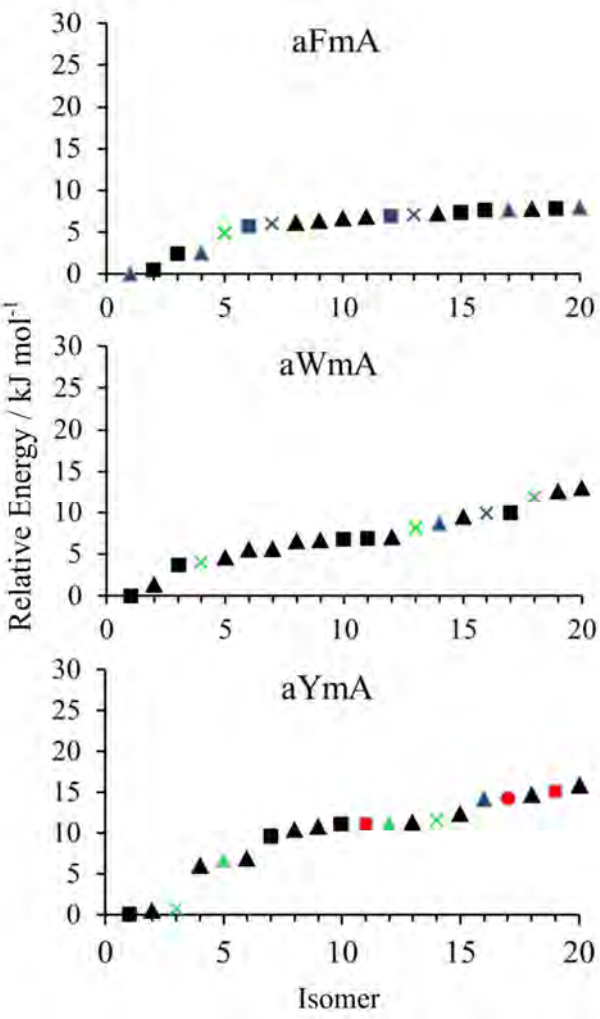


Figure 5.39 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aFmA, aWmA and aYmA. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

Moreover, the region near the minimum shows that structures of different families present very similar stability and therefore there seems to be a competition between several conformers for the final shape of the aggregates. Thus, aFmA1 and aFmA2 exhibit hydrogen bonds with the

backbone, but the side chain is oriented in a different way: perpendicularly to the DNA base or in parallel to the adenine establishing π···π interactions, respectively. **Figure 5.40** collects the structure of both trends, together with the values of their relative stability and binding energy. Despite that they are almost isoenergetic, the binding energy value of aFmA2 points that this trend is clearly favored in the aggregation compared to aFmA1. On the other hand, the global minimum and its closer conformers of aWmA and aYmA all present the aromatic ring parallel to adenine and the subsequent establishment of one or two hydrogen bonds with the backbone, except aYmA3, which only stablishes π···π interactions with the side chain.

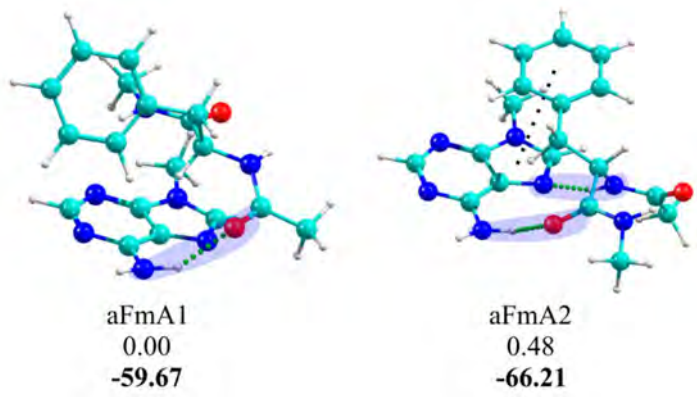


Figure 5.40 *Structure of the two lowest conformers of aFmA aggregates, together with the values of their relative energy and binding energy (bold) in kJ mol*$^{-1}$*. The intermolecular hydrogen bonds with the peptidic bond was highlighted in blue and the π···π interactions with side chain was highlighted with a black dot line.*

In summary, aromatic amino acids - mA aggregates maintain the tendency observed by the previous aromatic base aggregates: the establishment of hydrogen bonds with the peptide members due to the positioning of the aromatic lateral chain parallel to mA. However, no single family dominates the conformational landscape of aFmA.

Aromatic amino acids-mG

The exploration of the conformational landscape carried out for aFmG, aWmG and aYmG resulted in a set of 52, 71 and 75 conformers

respectively that were fully optimized by DFT. The most stable structure of each aggregate and their relevant parameters may be found in **Figure 5.41** and **Tables 5.23** and **5.24**. The structure of mG presents a C=O group that favors formation of hydrogen bonds. Accordingly, three hydrogen bonds were found in these aggregates between the electronegative moieties of the backbone and mG, while at the same time, the aromatic side chain establishes π···π interactions. NCI plots confirm the nature of these interactions, with blue discs and wide green zones respectively.

Table 5.23. *Distances (Å) and angles (º) of the intermolecular interactions observed in aFmA1, aWmA1 and aYmA1.*

	Interaction type							
	H bond						vdW	
	C1=O·· ··H(NH)	C1=O· ··HN1	N5H·· ··O=C6	N2H·· ··O=C6	C4=O·· ··H(NH)	C4=O·· ··HN1	R: π···π	R: OH···π
aFmG1	2.067 /146.0	1.915 /152.7	2.018 /161.9				~3.100 / -	
aWmG1				1.992 /156.5	2.029 /146.0	1.929 /152.5	~3.100 / -	
aYmG1				1.993 /158.8	2.006 /150.5	2.056 /148.7	~3.000 /-	~2.700 /-

Table 5.24. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aFmG1, aWmG1 and aYmG1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type					
	H bond					
	C1=O·· ··H(NH)	C1=O· ··HN1	N5H·· ··O=C6	N2H·· ··O=C6	C4=O·· ··H(NH)	C4=O·· ··HN1
aFmG1	0.0175	0.0262	0.0184			
aWmG1				0.0228	0.0189	0.0258
aYmG1				0.0223	0.0191	0.0173

The analysis of the dispersive contribution to the binding energy in aFmG, aQmG and aYmG aggregates is displayed in **Figure 5.42**. They exhibit a similar value than for the rest of the aromatic aggregates studied, however the relative contribution is significantly smaller, between 42-51% of the total interaction energy, due to an important increase in the strength of the hydrogen bonds. As in previous cases, the main dispersive contribution comes from the interaction of the electron density of the aromatic ring with the base. However, the increase in the number of the hydrogen bond reduces the importance of the dispersive contribution in the final binding energy. Furthermore, comparison between structures with and without π···π interactions (**Figure 5.43**) reveals an energy difference of ~19-22 kJ mol^{-1}, which is of the order of the previous aromatic – mA aggregates.

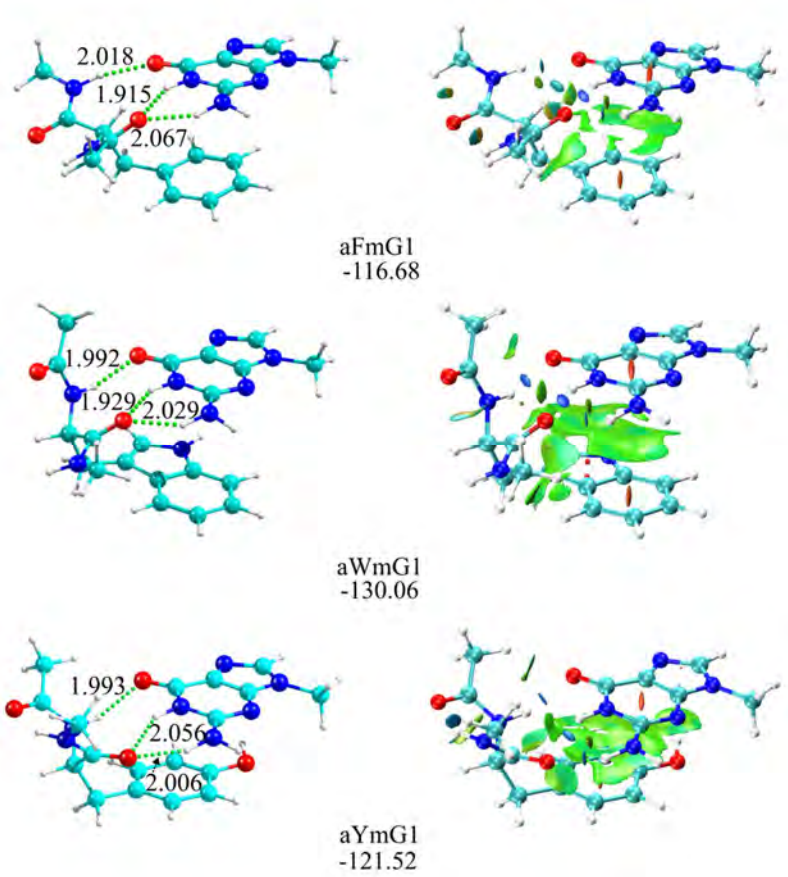


Figure 5.41 *The most stable structure found for aFmG, aWmG and aYmG complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

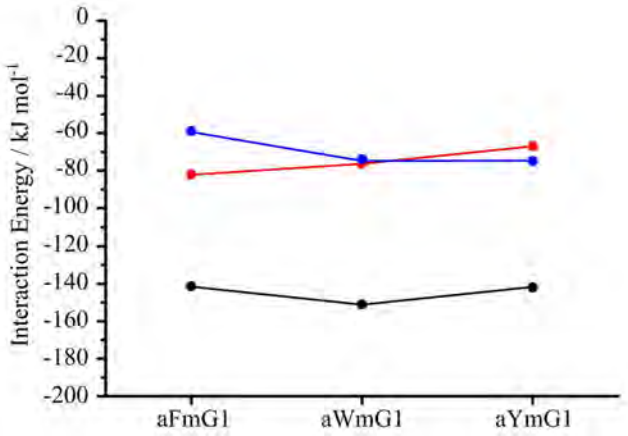


Figure 5.42 *Comparison between binding energies calculated at B3LYP-D/6-311++G(d,p) (●) and B3LYP/6-311++G(d,p) (●) levels and an estimation of the contribution of the dispersive forces to the binding energy (●) of aFmG, aWmG and aYmG.*

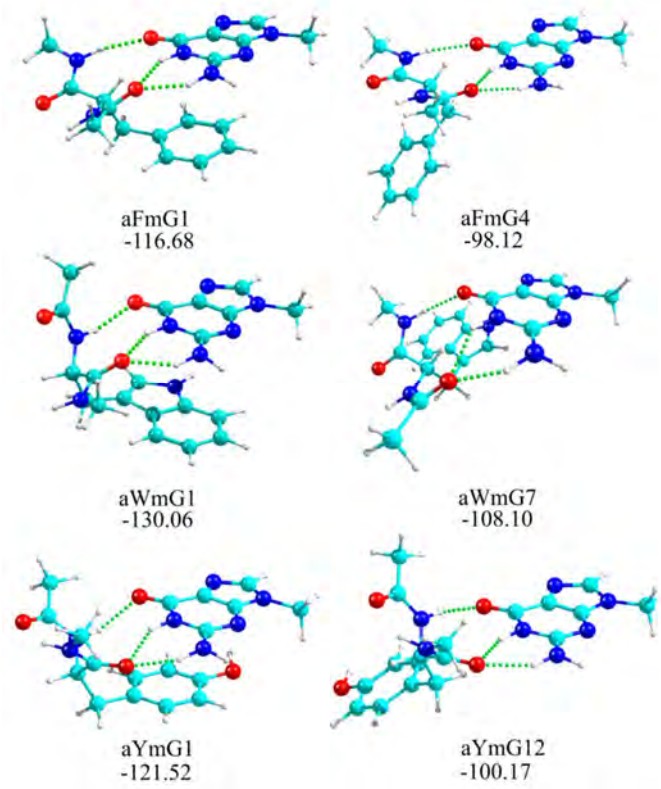


Figure 5.43 *The most stable structure and the first structure without π···π interactions for aFmG, aWmG and aYmA aggregates, with the values of their binding energies in kJ mol^{-1}.*

The analysis of the conformational preferences of the first 20 aggregates, **Figure 5.44**, shows that the neighborhood of the global minimum is dominated by a single family of conformers, where the aromatic ring interacts with the base and the backbone is close-enough to establish three HB. Therefore, the structures analyzed in this section are expected to dominate the conformational landscape.

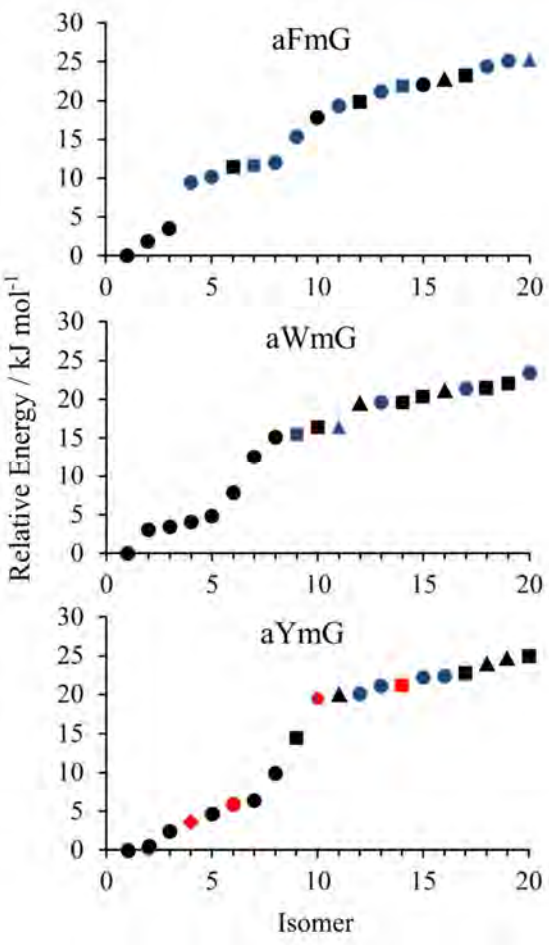


Figure 5.44 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aFmG, aWmG and aYmG. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

5.1.4. Discussion

Along this chapter, an exhaustive study of the conformational map and the properties of the aggregates of polar, non-polar and aromatic amino acids with DNA bases was carried out. Two kinds of intermolecular interactions, moderate/strong hydrogen bonds and van der Waals dispersive forces, control the aggregation properties of these clusters. It was demonstrated that hydrogen bond controls formation of the clusters, as the statistical analysis of 720 structures shows.

Three different interaction patterns were observed depending on the nature of the amino acid involved in the aggregates: non-polar amino acids establish hydrogen bonds between the carbonyl and amine groups of the backbone and the DNA bases, while the lateral non-polar chain lies away from the DNA bases. In contrast, the lateral chain of polar amino acids plays a central role in the aggregation process, adding extra stability to the system which also presents hydrogen bonds with the peptide bond. Aromatic amino acids show somehow a similar trend, although in this case the interaction between the side chain and the DNA base is of $\pi\cdots\pi$ character.

Interestingly, complementary bases exhibit similar behavior: aggregates of mC and mG present lower percentage of dispersive contribution than mT and mA aggregates, although it is usually due to an important increase in the number and/or strength of the hydrogen bonds.

Comparison of the binding energy values obtained with two different functionals, B3LYP-D and M06-2X for pyrimidinic and puric aggregates may be found in **Figure 5.45**. Clearly, there is a good agreement between the results obtained with both methods, what underlines that M06-2X is a good method to investigate the systems where dispersive interactions are present, as DNA - proteins systems.

Finally, one of the goals of this study is to test if a correlation between the structures of individual interactions and the statistical observations exists. The number of the contacts between the studied amino acids and the DNA bases obtained from Sathyapriya et al,[28] may be found in **Figure 5.46a**, while **Figure 5.46b** shows the average binding energy values of the amino acid – DNA base aggregates studied along this chapter. There is a clear correlation between both magnitudes: amino acids with aliphatic side chain exhibit both a lower number of contacts and smaller dissociation values.

On the other hand, amino acids with hydrophilic side chain exhibit an increasing number of contacts with the base. This tendency is preserved in the study of the dissociation energy of each individual pair. However, aN presents a lower binding energy value compared to its propensity. This may be due to the unusual binding energy found in aNmA, which significantly reduces the value of the mean binding energy.

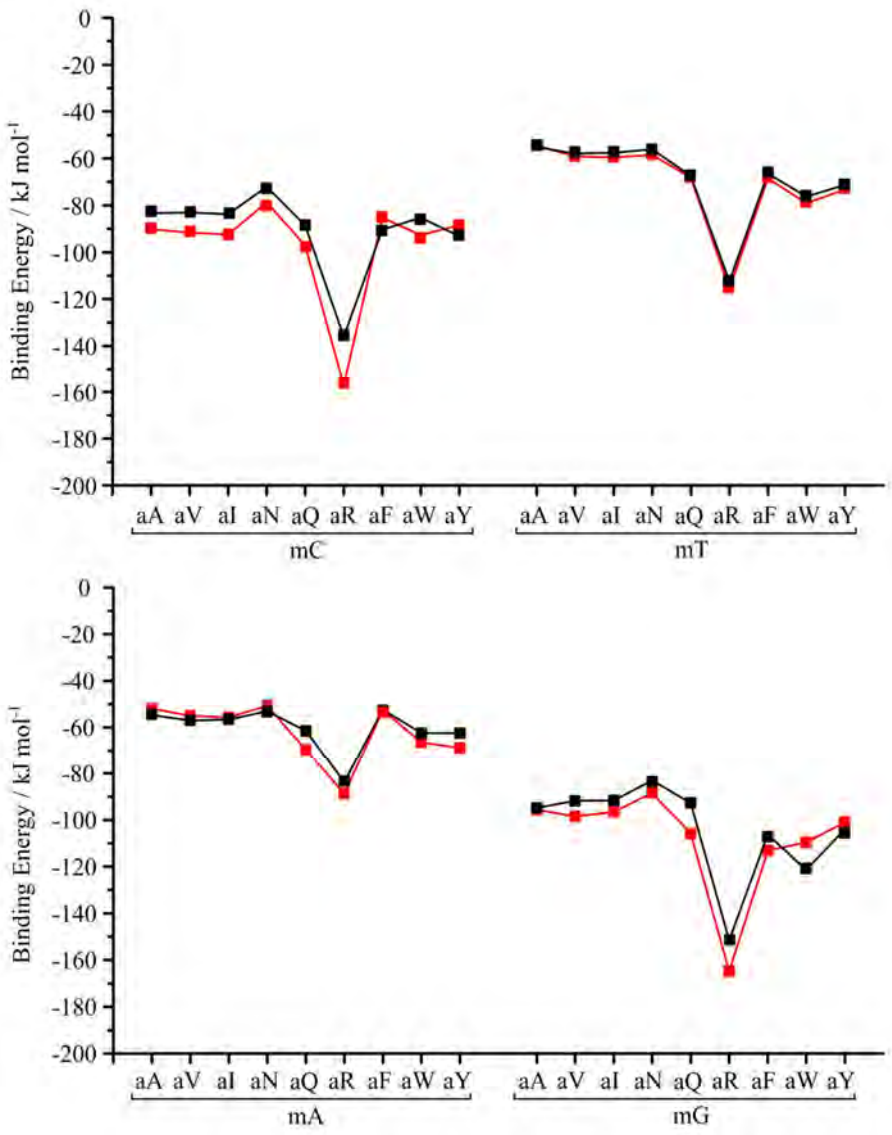


Figure 5.45 *Comparison between binding energies calculated at M06-2X/6-311++G(d,p) (■) and B3LYP-D/6-311++G(d,p) (■) levels for pyrimidinic (upper) and puric (lower) aggregates.*

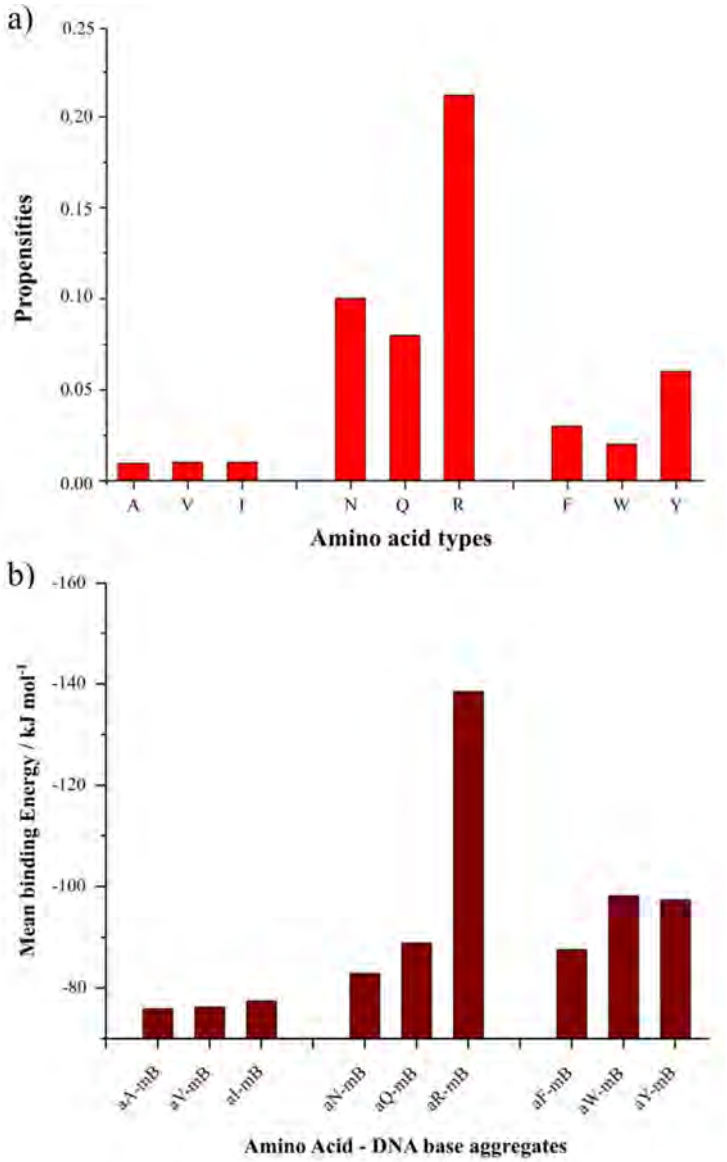


Figure 5.46 (a) *Representation of the propensities of the different amino acids with the DNA bases.*[28] ***(b)*** *Representation of the calculated mean binding energy of amino acid-DNA bases (mB) aggregates for the most stable structures.*

The correlation seems not to be so clear for aromatic amino acids, which present higher binding energy values than expected from the propensity of contacts. This is probably due to steric effects: while in the isolated pair the aromatic amino acid is free to establish $\pi\cdots\pi$ interactions,

those cannot take place in the real histone-DNA system, because the aromatic side chain cannot fit between two consecutive pairs of DNA bases.

Altogether, the results point to a control of the interaction between the superstructures of histones and DNA by individual amino acid-DNA base pair interactions. Moreover, it may be argued that the sequence of amino acids in the histones is designed to reinforce the interaction with some sections of DNA, and therefore that each histone is designed to bind to a specific section of DNA. Such specificity may well be related to their role in their regulation of DNA expression.

At this stage a question remains still open: several amino acids interact with the NH and CO moieties that are usually involved in formation of hydrogen bonds with the complementary base. Therefore, those places will not be available because the base pairing in the DNA double strand. Thus, this model is valid to explain the interaction between proteins and single-stranded DNA. For double-stranded DNA, the model must be improved.

5.2 Amino acids – DNA base pair

We improved here the theoretical model, by considering the interaction between the amino acids and Watson and Crick DNA base pairs. **Figure 5.47** displays the hydrogen-bonding patterns in base pairs defined by Watson and Crick,[2,30] with the X-ray diffraction data.[31]

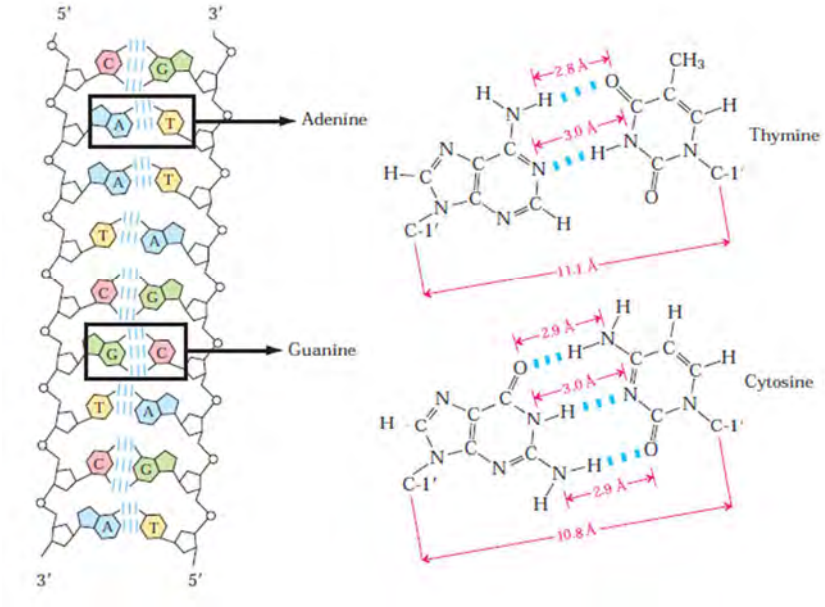


Figure 5.47 *Representation of the hydrogen-bonding patterns in the base pairs defined by Watson and Crick.*[2] *Hydrogen bonds are represented by three blue lines.*

The structure of both dimers, after optimization and vibrational analysis at M06-2X/6-311++G(d,p) level may be found in **Figure 5.48**, together with the distances depicted in the Watson Crick representation. Once the DNA base pairs were optimized, they were presented to the nine selected amino acids. In this study, several structural restrictions were introduced. Otherwise, structure optimization resulted in most cases in an unacceptable distortion of the Watson-Crick pair, as illustrated in **Figure 5.49**. One must keep in mind the main goal of this study, which is to characterize the interaction between the amino acids and the base pairs as they are found in the DNA double strand. Therefore, we froze the structure of DNA base pair during the whole process and only the position of the amino acid relative to the base pair was allowed to change.

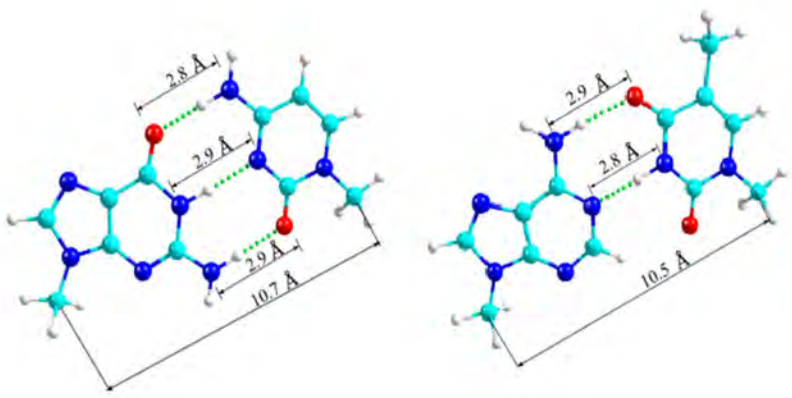


Figure 5.48 *Optimized structure of Watson-Crick DNA base pairs mC-mG (left) and mA-mT (right) obtained in the conformational study. Some relevant distances are also depicted with black lines.*

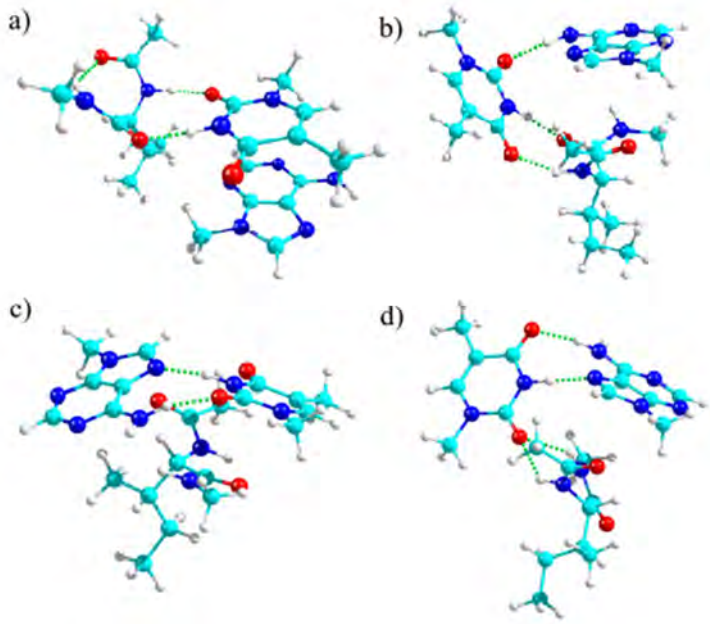


Figure 5.49 *Some of the structures observed in the conformational search of aImAT when no constraint were imposed.*

Due to the negative frequencies obtained in several aggregates, ZPE correction was not taken into account in the calculation of the final binding energies, and only BSSE correction was applied.

5.2.1 Non-polar amino acids – DNA base pairs

5.2.1.1 Non-polar amino acids – mCmG

The conformational analysis of the non-polar amino acids with mCmG base pairs was applied to 30, 44 and 41 aggregates of aAmCmG, aVmCmG and aImCmG respectively and the structure of the most stable non-polar amino acid-mCmG aggregate, together with the value of their binding energy, the most relevant interactions and the NCI plots may be found in **Figure 5.50**. The values of the distances and the angles of the most relevant interactions and the electron density may be found in **Table 5.25** and **5.26**.

Table 5.25. *Distances (Å) and angles (º) of the intermolecular interactions observed in aAmCmG1, aVmCmG1 and aImCmG1.*

	Interaction type	
	H bond	
	C1=O···H(NH)	N5H···O=C6
aAmCmG1	1.876/170.6	2.051/163.1
aVmCmG1	1.886/165.0	2.082/160.0
aImCmG1	1.870/172.8	2.065/167.9

Table 5.26. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aAmCmG1, aVmCmG1 and aImCmG1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type	
	H bond	
	C1=O···H(NH)	N_5H···O=C6
aAmCmG1	0.0274	0.0197
aVmCmG1	0.0264	0.0185
aImCmG1	0.0276	0.0187

The main consequence of replacing a single DNA base by a base pair is the reduction in the number of points of contact and the increase on the size of the aromatic cloud. However, the interaction pattern previously observed for non-polar amino acids is maintained in mCmG complexes. Thus, they form hydrogen bonds with the members of the backbone, with similar distances and strength. Meanwhile, the side chain lies away from the base, as a mere spectator. In good agreement, the binding energy values of these aggregates are very similar for all three systems, being slightly more negative in aAmCmG1.

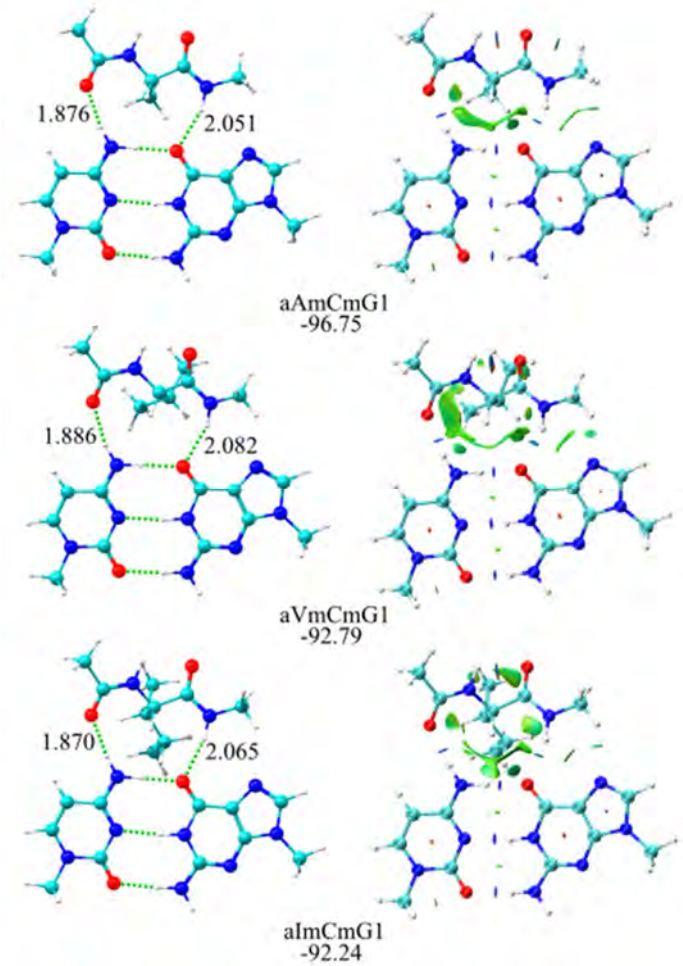


Figure 5.50 *The most stable structures found for aAmCmG, aVmCmG and aImCmG complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are also shown.*

The comparative study of the 20 most stable structures of each aggregate, **Figure 5.51**, shows that formation of hydrogen bond patterns is still the preferred trend of aggregation and most of the structures exhibited at least one hydrogen bond with the peptide bond members. Furthermore, structures with side chain interactions were not found in the low-energy region, governed by two HB with the backbone.

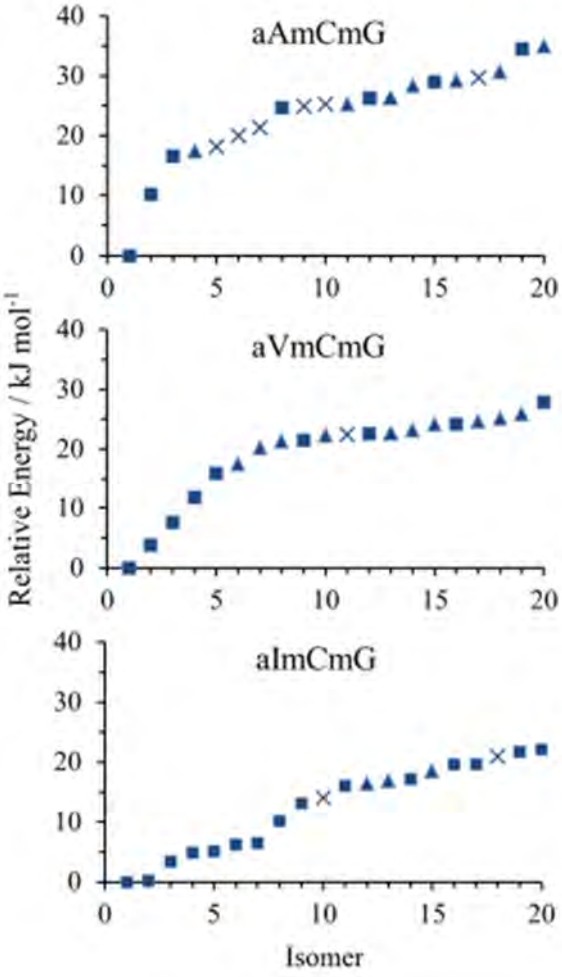


Figure 5.51 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aAmCmG, aVmCmG and aImCmG. Blue color: interaction between the peptidic bond and the base. Types of interactions: HB: hydrogen bond (○: three, □: two, and Δ: one HB); vdW: van der Waals interactions (×).*

5.2.1.1 Non-polar amino acids – mAmT

The conformational search of non-polar amino acids and mAmT resulted in 50 structures for aAmAmT, 59 for aVmAmT and 52 for aImAmT. The most stable structure of each aggregate and their relevant parameters may be found in **Figure 5.52** and **Tables 5.27** and **5.28**.

There is a clear difference between the interaction patterns in aAmAmT1 and in the other two aggregates. While aAmAmT is governed by van der Waal interactions, several hydrogen bonds were found in VmAmT1 and aImAmT1. Nevertheless, the contribution of the side chain to the aggregation process seems to be irrelevant. Although some weak interactions were found for the aVmAmT1 and aImAmT side chain, it does not play a central role in the aggregation.

Table 5.27. *Distances (Å) and angles (º) of the intermolecular interactions observed in aAmAmT1, aVmAmT1 and aImAmT1*

	Interaction type			
	H bond		vdW	
	N2H···N7	N5H···O=C2	N2H···N7	R: CH···π
aAmAmT1			2.410/142.4	
aVmAmT1		1.992/164.6		~2.400/ -
aImAmT1	2.099/155.3			~2.200/ -

Table 5.28. *Electron density (a.u.) of the hydrogen bonds observed for the clusters aAmAmT1, aVmAmT1 and aImAmT1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type	
	H bond	
	N2H···N7	N5H···O=C2
aAmAmT1		
aVmAmT1		0.0203
aImAmT1	0.0200	

Regarding the binding energy values, aAmAmT seems to be slightly less stable, probably due to the absence of hydrogen bonds. The trend observed is maintained in the next most stable structures, as data in **Figure 5.53** shows. While the conformers surrounding the global minimum are dominated by vdW interactions, the trend to establish HB with the backbone was also found. A relatively large number of structures were found packed in a very narrow energy window, pointing to a competition between both families for shaping the molecular aggregate. This is especially true in aImAmT, where structures with and without hydrogen bond with the backbone co-exist.

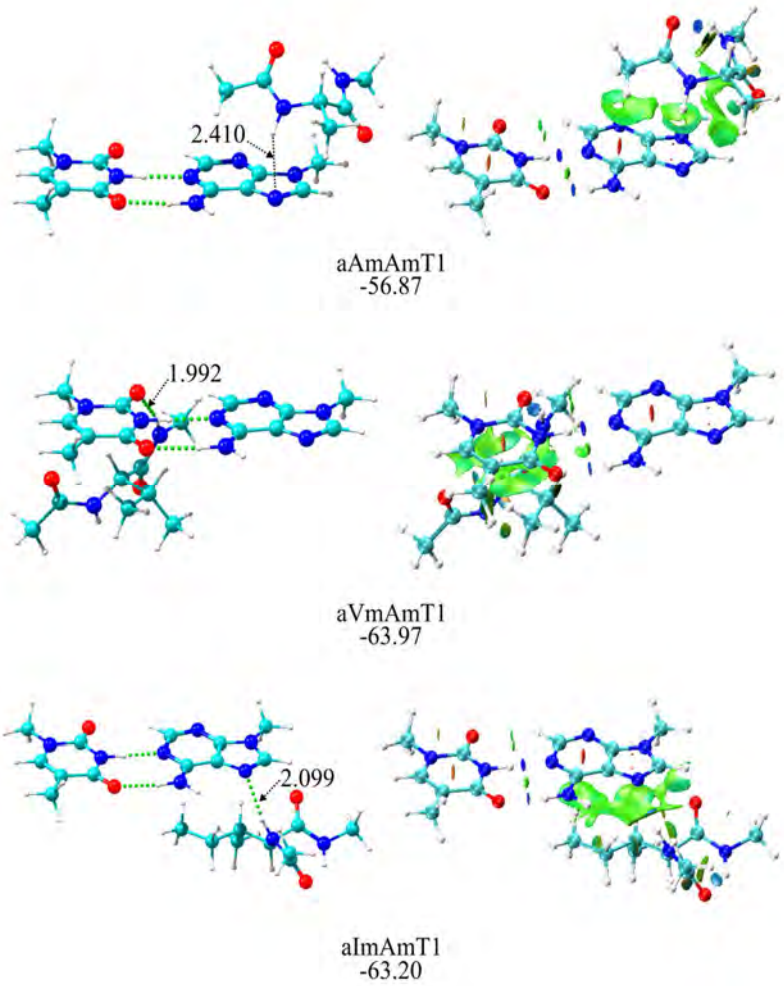


Figure 5.52 *The most stable structure found for aAmAmT, aVmAmT and aImAmT complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

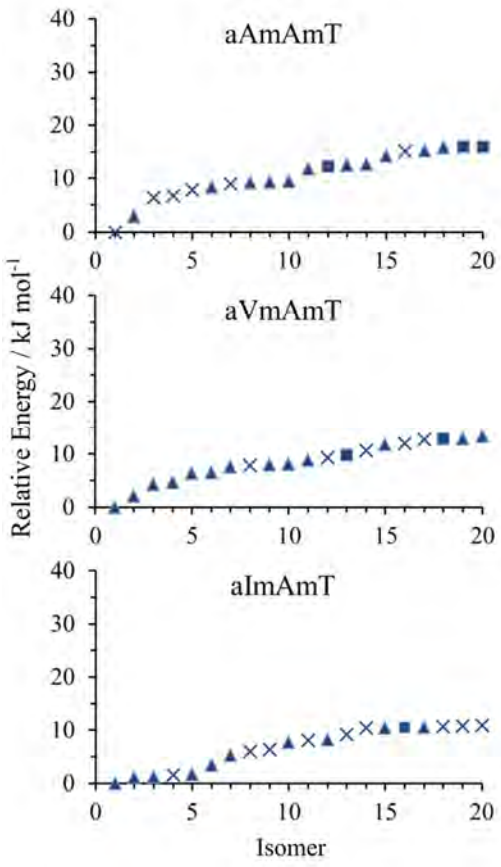


Figure 5.53 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aAmAmT, aVmAmT and aImAmT. Blue color: interaction between the peptidic bond and the base. Types of interactions: HB: hydrogen bond (□: two, and Δ: one HB); vdW: van der Waals interactions (×).*

5.2.2. Polar amino acids – DNA base pairs

5.2.2.1 Polar amino acids – mCmG

The same methodology was applied for the aggregation of the three polar amino acids with mCmG base pair, obtaining 50, 67 and 52 structures for aNmCmG, aQmCmG and aRmCmG respectively that were optimized using DFT. The structure of the most stable polar amino acid-

mCmG aggregate, together with the value of their binding energy, the most relevant interactions and the NCI plots may be found in **Figure 5.54**; the parameters of the most relevant interactions were collected in **Tables 5.29** and **5.30**.

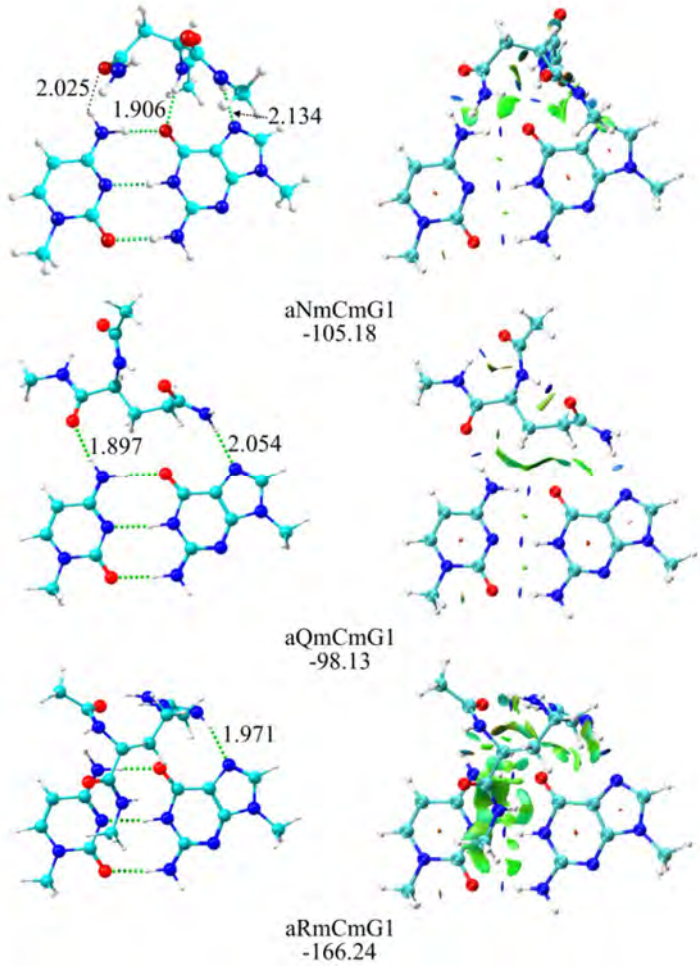


Figure 5.54 *The most stable structure found for aNmCmG, aQmCmG and aRmCmG complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

Following the trend observed in the interaction with a single base, the side chain does take part in the aggregation, establishing extended contacts with the dimer of DNA bases that include formation of hydrogen bonds and at the same time, vdW interactions, as highlighted in the NCI

plots. Those plots also point out that the observed interactions with the side chain contribute with a relevant amount of stabilization energy to the formation of the aggregates.

Table 5.29. *Distances (Å) and angles (º) of the intermolecular interactions observed in aNmCmG1, aQmCmG1 and aRmCmG1.*

	Interaction type						
	H bond				vdW		
	N5H·· ··N7	N2H·· ··O=C6	C4=O·· ··H(NH)	R:H(NH)·· ··N7	R:O(C)·· ··H(NH)	N5H·· ··O(C)	N5H·· ··N1H
aNmCmG1	2.134/141.6	1.906/172.7			2.025/131.6		
aQmCmG1			1.897/168.8	2.054/162.0			
aRmCmG1				1.971/156.1		2.464/156.3	2.543/148.2

Table 5.30. *Electron density (a.u.) of the hydrogen bonds and van der Waals interactions observed for the clusters aNmCmG1, aQmCmG1 and aRmCmG1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type				
	H bond				vdW
	N5H·· ··N7	N2H·· ··O=C6	C4=O·· ··H(NH)	R:H(NH)·· ··N7	R:O(C)·· ··H(NH)
aNmCmG1	0.0188	0.0268			0.0220
aQmCmG1			0.0332	0.022	
aRmCmG1				0.0180	

Regarding the absolute value of the binding energy in these systems, similar values were found compared to the complex of the same amino acids with the components of the dimer: mC and mG. Perhaps, the most remarkable difference is that the small binding energy value found for dimers of aN increased in aNmCmG, which has a higher binding energy than aQmCmG. aRmCmG follows the trend observed previously and it is still by large the most stable aggregate.

Analysis of the 20 most stable aggregates of each system, **Figure 5.55**, depicts a rich conformational landscape, in which several families of isomers exist in a narrow stability window. However, the prevalence of green, red and black colors highlight the propensity of the lateral chain to interact with the dimer of DNA bases, confirming the tendency observed

for the global minimum. Furthermore, despite that mCmG only interacts with the backbone of aN in aNmCmG2, the gap is high-enough (more than 7 kJ mol^{-1}) to consider this group as a possible competitor of the global minimum.

In summary, the previous trend observed for the polar amino acid – single DNA base aggregates is maintained in their aggregation with the canonical DNA base pairs and the side chain plays a main role in the aggregation process, accompanied by hydrogen bonds with the peptidic moieties.

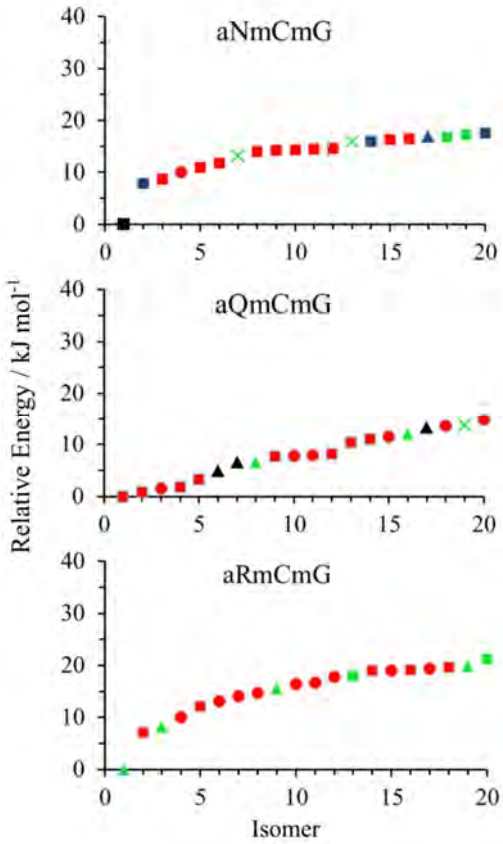


Figure 5.55 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aNmCmG, aQmCmG and aRmCmG. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

5.2.2.2 Polar amino acids – mAmT

A similar conformational landscape was encountered for this three systems, for which 56 structures of aNmAmT, 47 of aQmAmT and 38 of aRmAmT were fully optimized using DFT. The most stable structure of each aggregate and their relevant parameters may be found in **Figure 5.56** and **Tables 5.31** and **5.32**. As it can be seen, aNmAmT1 and aQmAmT1 establish hydrogen bonds with its lateral chain whereas aRmAmT1 only exhibits one hydrogen bond with the amine group of the backbone. This is the first time that the most stable aR aggregate does not present a direct hydrogen bond with the lateral chain, but a strong electrostatic interaction with the CO moiety of the DNA base pair, as observed in the NCI plot.

Table 5.31. *Distances (Å) and angles (º) of the intermolecular interactions observed in aNmAmT1, aQmAmT1 and aRmAmT1.*

	Interaction type							
	H bond				vdW			
	R:H(NH)··· ··N7	R: H(NH)··· ··O=C2	N2H··· ··O=C4	N5H··· ··O=C4	R:O(C)··· ··H(NH)	R: H(NH)··· ··O=C2	R:O···π	R:N···π
aNmAmT1	2.099/145.3				2.299/134.9			
aQmAmT1		2.082/152.7	1.958/147.5				~3.100/-	
aRmAmT1				1.942/165.5		2.215/134.1		~3.100/-

Table 5.32. *Electron density (a.u.) of the hydrogen bonds and van der Waals interactions observed for the clusters aNmAmT1, aQmAmT1 and aRmAmT1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type			
	H bond			
	R:H(NH)··· ··N7	R: H(NH)··· ··O=C2	N2H··· ··O=C4	N5H··· ··O=C4
aNmAmT1	0.0204			
aQmAmT1		0.0182	0.0230	
aRmAmT1				0.0238

Analysis of the binding energy values shows that the trend previously observed still holds, although the dissociation energy is stronger than in mA/mT aggregates. More precisely, the values in **Figure 5.56** are closer to those for mT aggregates and certainly, all three amino

acids have a closer contact with that monomer in mAmT aggregates. On the other hand, the higher binding energy value of aRmAmT compared to the other two systems reinforces the propensity of this amino acid to interact with the DNA bases.

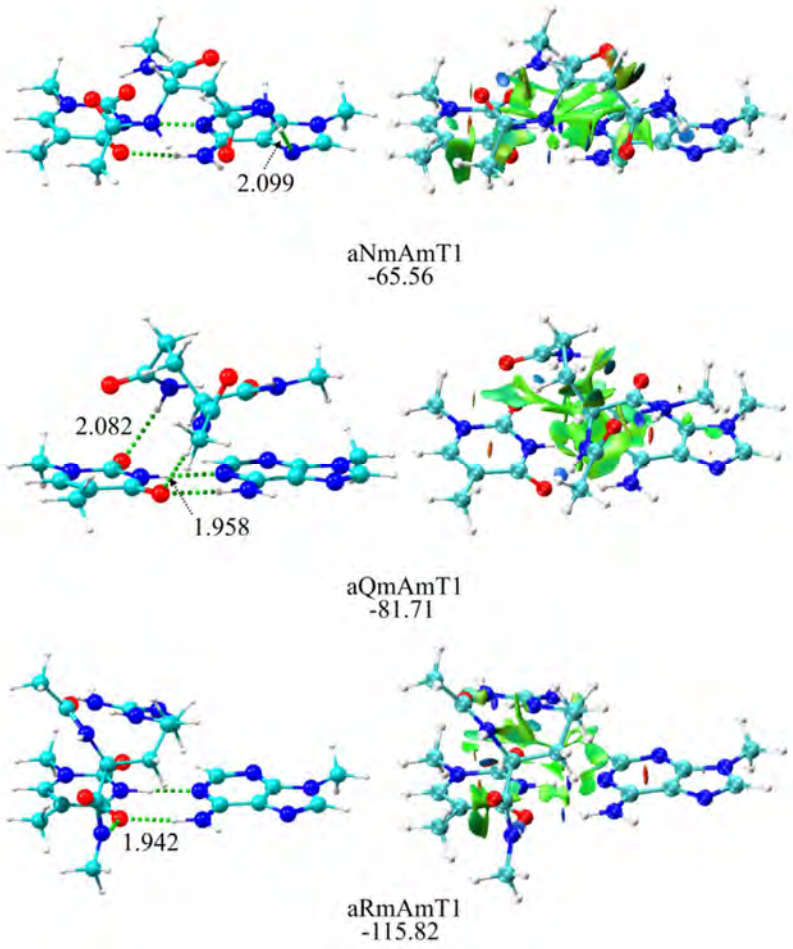


Figure 5.56 *The most stable structure found for aNmAmT, aQmAmT and aRmAmT complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are shown.*

Regarding the general tendency of the systems studied, **Figure 5.57** depicts the energetic order of the first 20 computed structures, with the leading interactions highlighted by the symbols and the color code. The main conclusion is that a rich diversity of structures was found for these systems, although the global minima and their vicinity present relevant

interactions between the side chain and the DNA bases, following the same trend observed in the complexes with individual DNA bases.

In summary, the interaction between polar amino acids and mAmT base pair is primarily guided by the contacts with the lateral chain of the amino acids, rearranging the structure to maximize the number of contacts with the other groups.

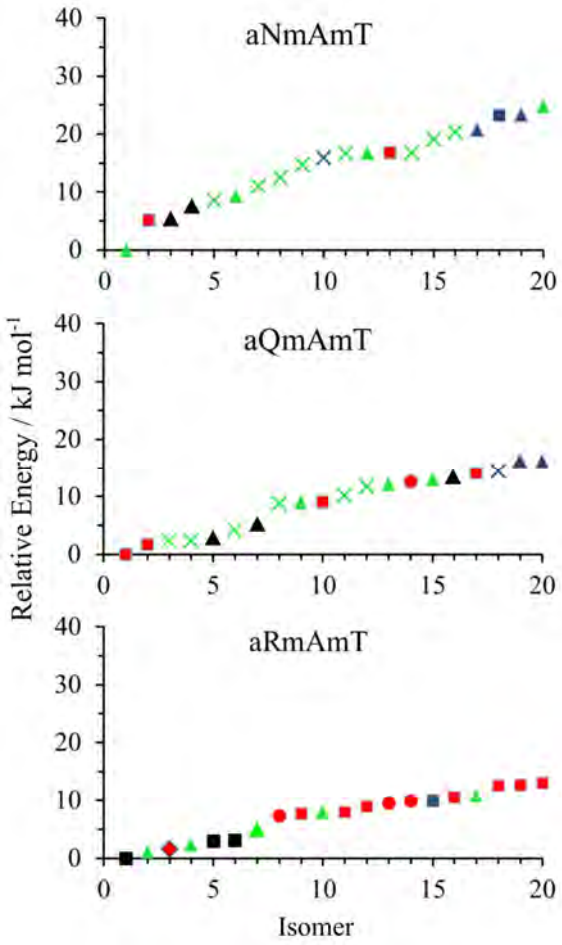


Figure 5.57 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aNmAmT, aQmAmT and aRmAmT. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW with the chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

5.2.3. Aromatic amino acids – DNA base pairs

5.2.3.1 Aromatic amino acids – mCmG

A total of 73, 61 and 43 conformers of aFmCmG, aWmCmG and aYmCmG respectively were explored and the structure of the most stable aromatic amino acid-mCmG complexes, together with their binding energy, the most relevant interactions and the NCI plots were collected in **Figure 5.58**. See also **Tables 5.33** and **5.34** for structural parameters.

Table 5.33. *Distances (Å) and angles (º) of the intermolecular interactions observed in aFmCmG1, aWmCmG1 and aYmCmG1.*

	Interaction type		
	H bond		vdW
	C1=O···H(NH)	N5H···O=C6	C3H···O=C
aFmCmG1	1.903/160.8	2.014/164.6	2.574/129.6
aWmCmG1	1.964/149.9	1.995/160.8	2.476/130.7
aYmCmG1	1.917/161.9	2.001/168.6	2.668/130.9

Table 5.34. *Electron density (a.u.) of the hydrogen bonds and van der Waals interactions observed for the clusters aFmCmG1, aWmCmG1 and aYmCmG1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type	
	H bond	
	C1=O···H(NH)	N5H···O=C6
aFmCmG1	0.0258	0.0208
aWmCmG1	0.0225	0.0213
aYmCmG1	0.0250	0.0211

The main conclusion extracted from the ball&stick structures in **Figure 5.58** is that the stacking interaction observed in previous sections is maintained also in this systems. The preference for the puric base to establish π···π interactions, **Figure 5.45**, also holds for the interaction with

the canonical mCmG dimer. So, the side chain position is in close contact with the mG monomer, resulting in similar dissociation energy values, despite that the peptidic bond is now able to select polar groups from both DNA bases to interact with, and therefore it has a larger number of options to maximize the interaction.

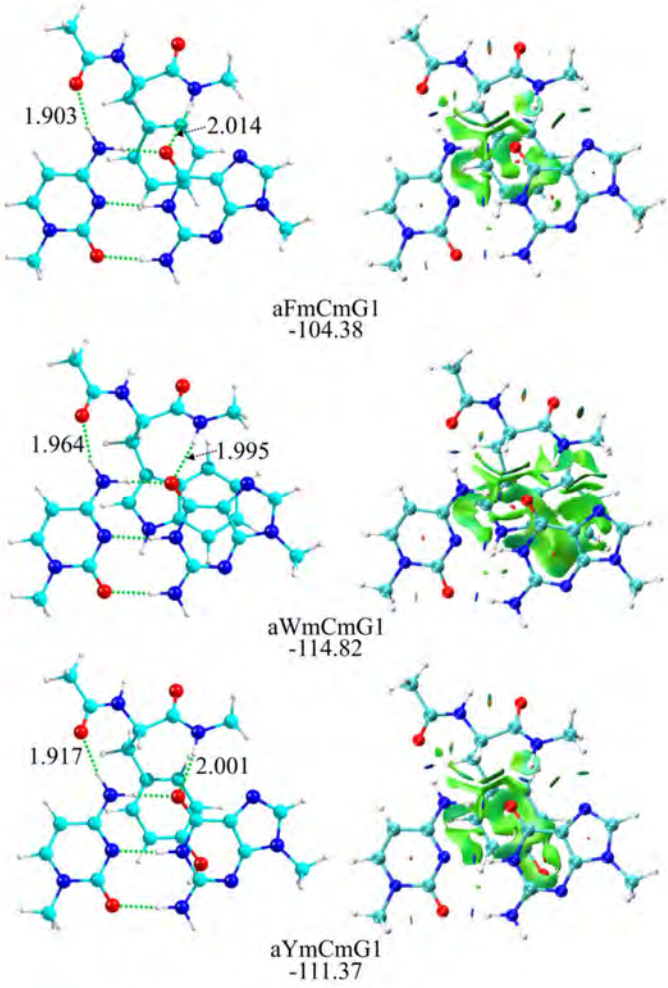


Figure 5.58 *The most stable structure found for aFmCmG, aWmCmG and aYmCmG complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are also shown.*

Thus, the observed differences in binding energy between the three conformers in **Figure 5.58** may be due to the strength of the π···π interaction, as the contacts between the peptidic bond and the DNA bases

is almost identical in all three systems. This trend is also observed in the rest of the optimized structures, **Figure 5.59**, especially in those conformers closer to the global minimum (below ~10 kJ/mol). Above such value, representative structures from other families start showing up.

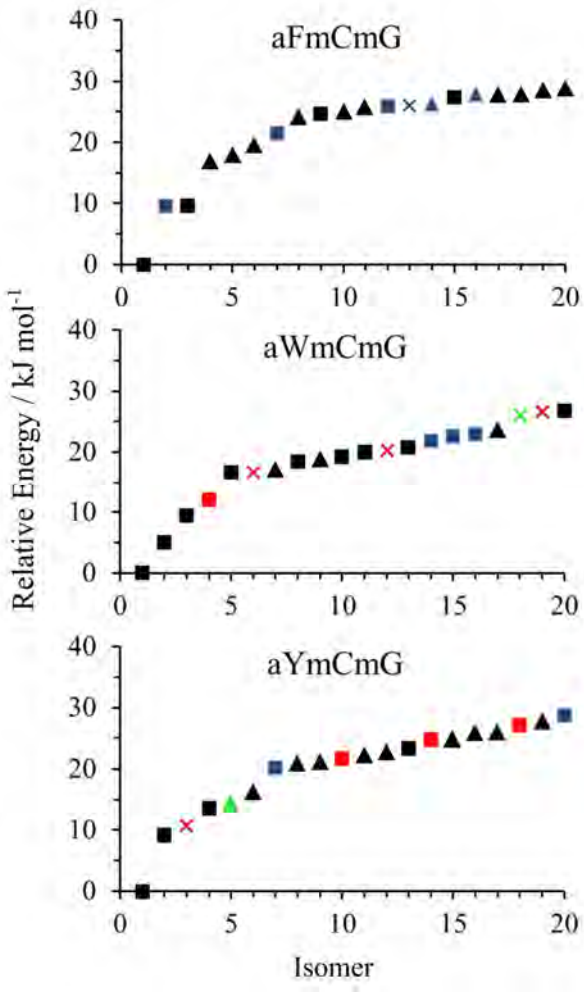


Figure 5.59 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aFmCmG, aWmCmG and aYmCmG. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: backbone and lateral chain exhibit the same interaction (HB or vdW). Black: HB with the backbone and relevant vdW interaction with the side chain. Types of interactions: HB (◇: four, ○: three, □: two, and Δ: one HB), vdW interaction (×).*

5.2.3.2 Aromatic amino acids – mAmT

The last part of the study addressed the aggregation of aromatic amino acids and mAmT base pair. 60, 64 and 103 structures were optimized for aFmAmT, aWmAmT and aYmAmT respectively and **Figure 5.60** and **Tables 5.35** and **5.36** collect the structure and relevant parameter of the most stable aggregates. Surprisingly, the structures are completely different from those in **Figure 5.58**: apparently, the optimal orientation of the peptidic skeleton is significantly modified in order to maximize the interaction with the dimer. This modification leads also to a very different interaction between the aromatic rings. Perhaps the extreme case is aYmAmT, where the whole dipeptide interacts with the aromatic cloud of the canonical dimer.

Table 5.35. *Distances (Å) and angles (°) of the intermolecular interactions observed in aFmAmT1, aWmAmT1 and aYmAmT1.*

	Interaction type				
	H bond				vdW
	N2H···O=C2	N5H···O=C2	N5H···O=C4	R:OH···N7	C3H···N1
aFmAmT1	1.867/160.5				
aWmAmT1		1.957/167.4			2.545/169.0
aYmAmT1			1.921/157.7	1.895/163.9	

Table 5.36. *Electron density (a.u.) of the hydrogen bonds and van der Waals interactions observed for the clusters aFmAmT1, aWmAmT1 and aYmAmT1, obtained from the NCI grid with 0.1x0.1x0.1 increments.*

	Interaction type			
	H bond			
	N2H···O=C2	N5H···O=C2	N5H···O=C4	R:OH···N7
aFmAmT1	0.0263			
aWmAmT1		0.0223		
aYmAmT1			0.0241	0.0322

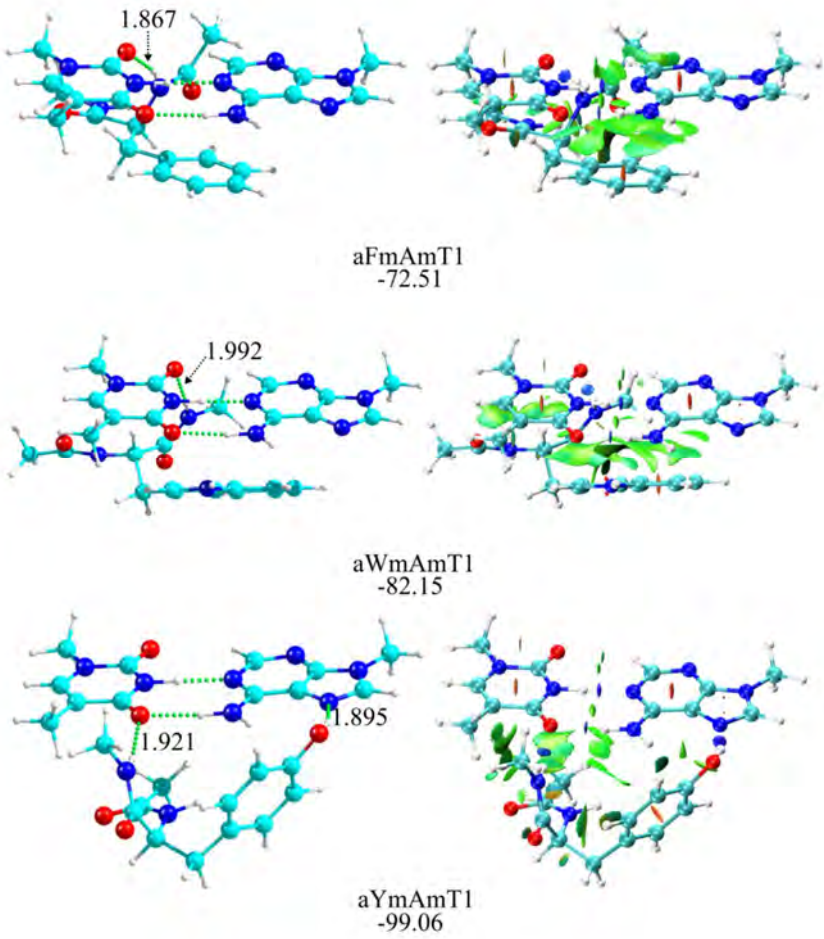


Figure 5.60 *The most stable structure found for aFmAmT, aWmAmT and aYmAmT complexes with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions and NCI plots. The binding energy values (kJ mol^{-1}) are also shown. Altogether, 60, 64 and 103 structures were optimized for aFmAmT, aWmAmT and aYmAmT respectively.*

The highlighted structural differences also result in a collection of binding energy values, which increase from aFmAmT to aYmAmT. Nevertheless, those values are very close to the ones observed for the interaction of the three amino acids with isolated mT, **Figure 5.45**. The interaction with aromatic clouds is characterized by shallow potentials.[32] This is reflected in a large collection of structures packed in a very narrow

energy gap, as highlighted in **Figure 5.61**. Thus the trend observed for the global minimum is not maintained for the rest of the isomers and there is in fact a strong competition between several families of conformers to shape the aggregate. Nevertheless, all of them are characterized by a direct interaction between the side chain and the aromatic cloud of the canonical base pair, which is physically not possible in real histone-DNA interactions, due to the steric hindrance imposed by the rest of the base pairs.

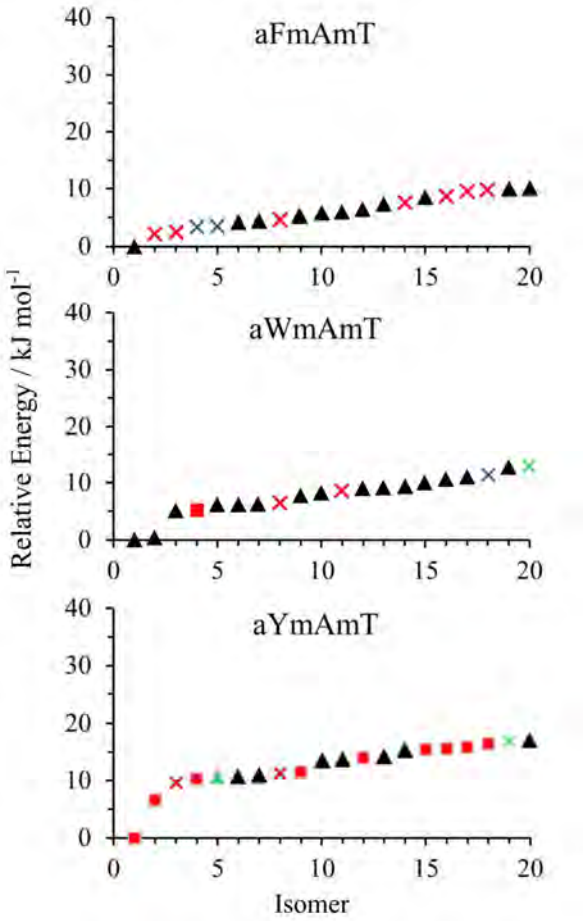


Figure 5.61 *Conformational distribution of the most relevant interactions found for the first 20 isomers of aFmAmT, aWmAmT and aYmAmT. Blue color depicts the interaction between the peptidic bond and the base. Green: Interaction between the lateral chain and the base. Red: Both interactions, peptidic bond and lateral chain, with the base, are relevant. Types of interactions: HB: hydrogen bond (□: two, and Δ: one HB), vdW: van der Waals interactions (×).*

5.2.4. Discussion

The model was increased in order to reproduce double-stranded DNA bases and the occupancy of several points of interactions involved in the complementarity of the DNA bases. The three types of amino acid (polar, non-polar and aromatic) were presented to the two possible canonical DNA base pairs, mGmC and mAmT, and an exhaustive study of the conformational landscape and molecular properties were carried out for the aggregates. Despite that this model is supposed to be closer to the real situation in histone-DNA interaction, the conclusions reached are not very different from those extracted with the simplified model. The interaction patterns observed in the simplified model are still present in the interaction with the canonical pairs. Non-polar amino acids seem reluctant to present their side chain to the base pair. Conversely, the interaction between the side chain of polar amino acids and the base pair takes an active role in the aggregation process. The same may be applied to the interaction with aromatic amino acids, although in this case a more sophisticated model is required to account for the steric factors that condition the interaction in real systems: mainly that a stacking interaction between the amino acid and the base pair is not possible because no physical space is available between stacked base pairs for the side chain to fit in.

Regarding the comparison with the contact propensity between amino acids and base pairs, **Figure 5.62**, no large differences were found with the previous and simplified model. Overall, a good agreement was reached, except for aN aggregates, for which the interaction energy is significantly lower than the propensity of contacts. May be this is due to other structural reasons. For example, N, asparagine, is a very hydrophilic amino acid and therefore it is placed in strategic locations that are required to force the histone to bend in certain ways. Therefore, it would be less exposed and therefore would have a higher probability to get hidden when the histone enters in contact with the DNA double strand.

Also, as explained above, no good agreement was found for the aromatic amino acids. Significantly more elaborated models would be required to improve it, probably involving at least three DNA base pairs, to simulate the steric constrains. Then, sugar and phosphate units would be required, increasing the size of the system beyond what can be tackled by "simple" DFT methods and therefore ONIOM-type methods or similar combinations of molecular mechanics and DFT would be required to

attack the problem. However, exploration of such methods is far beyond the scope of the present study and may be the subject of future works.

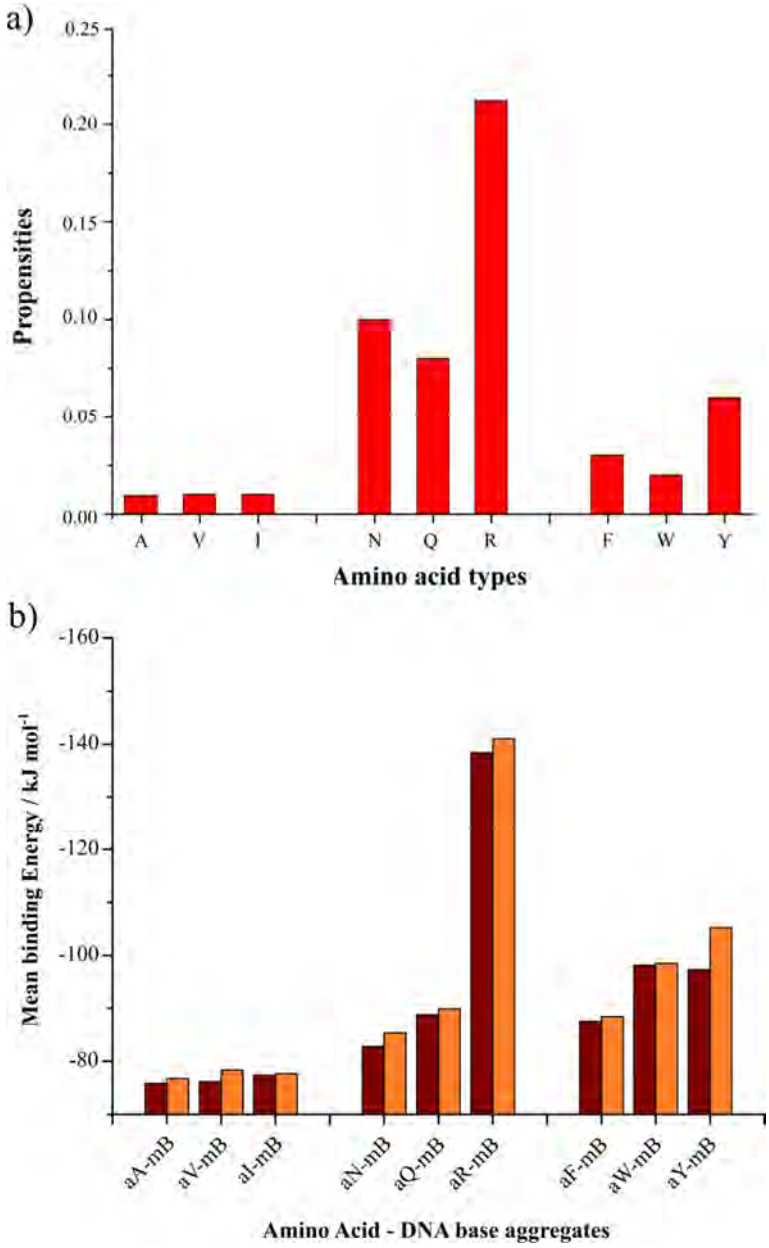


Figure 5.62 (a) *Representation of the propensities of the different amino acids with the DNA bases.*[28] ***(b)*** *Representation of the mean binding energy of amino acid-DNA bases (mB) aggregates (maroon) and amino acid-DNA base pairs aggregates (orange) for the most stable structure.*

5.3 CONCLUSIONS

We investigated the molecular basis of the histone-DNA interactions through a reductionist approach, using a combination of computational methods which included Molecular Mechanics, DFT calculations, and an NCI analysis. The aggregation properties of the 54 dimers and a total of 1080 structures is finally controlled by two kinds of intermolecular interactions, moderate/strong hydrogen bonds and van der Waals dispersive forces, for which we provided a comprehensive conformational study and molecular properties.

The analysis of the interactions found for the most stable clusters shows that the interaction between non-polar amino acids (alanine, valine and isoleucine) and DNA bases always takes place at the peptidic bond and their side chain lies away from the DNA bases. For polar (asparagine and glutamine) or charged (arginine) amino acids, the results show that the side chain plays a main role in the aggregation, adding extra stability to the system which also presents hydrogen bonds with the peptide bond. Aromatic amino acids (phenylalanine, tryptophan and tyrosine) exhibit a similar trend, although in this case the interaction between the side chain and the DNA base is of $\pi\cdots\pi$ character and hydrogen bonds with the backbone are also present.

The correlation between the calculated binding energy and the propensities protein-DNA base found in the literature is noticeable for non-polar and polar amino acids. However, no good agreement was found for the aromatic amino acids and more elaborated models would be required to improve this correlation.

In summary, the balance between hydrogen bonds and van der Waals interactions in the control of the final shape and binding energies of amino acid-DNA base pairs is satisfactorily accounted for in dispersion-corrected DFT molecular models. The computational data offer chemical insight into the dominant interactions representing the histone-DNA interaction, becoming a useful tool for the prediction of aggregation properties of DNA. Moreover, the characterization of the interactions which take place between the amino acids and the DNA bases could help to elucidate the language that rules the process of recognition for the amino acids of a particular histone and specific sequences of DNA: maximizing the interaction between the histone-DNA to force the DNA to wrap around the sections of the histone with higher affinity.

5.4 References:

1 L. Stryer, J. M. Berg and J. L. Tymoczko, *Biochemistry*, Reverté, New York, 5th edn., 2002.

2 D. L. Nelson and M. M. Cox, *Principles of Biochemistry*, Omega, New York, 4th edn., 2000.

3 B. Alberts, A. Johnson, D. Morgan, M. Raff and J. Lewis, *Molecular Biology of the Cell*, Taylor & Francis Inc, New York, 2014.

4 B. D. Strahl and C. D. Allis, *Nature*, 2000, **403**, 41–45.

5 K. Luger, *Chromosom. Res.*, 2006, **14**, 5–16.

6 B. M. Turner, *Cell*, 2002, **111**, 285–291.

7 T. Jenuwein and C. D. Allis, *Science*, 2001, **293**, 1074–1081.

8 R. Z. Jurkowska and A. Jeltsch, *Methods Mol. Biol.*, 2010, **649**, 149–161.

9 M. M. Müller and T. W. Muir, *Chem. Rev.*, 2015, **115**, 2296–2349.

10 G. S. Freeman, J. P. Lequieu, D. M. Hinckley, J. K. Whitmer and J. J. De Pablo, *Phys. Rev. Lett.*, 2014, **113**, 1–5.

11 K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, A. W. M Auml Der, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251–260.

12 T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 730–748.

13 Y. Zhao and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2701–2705.

14 Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.

15 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

16 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 1–19.

17 F. Boys, S.F.; Bernardi, *Mol. Phys.*, 1970, **19**, 553–556.

18 S. Jones, P. van Heyningen, H. M. Berman and J. M. Thornton, *J. Mol. Biol.*, 1999, **287**, 877–896.

19 C. O. Pabo and L. Nekludova, *J. Mol. Biol.*, 2000, **301**, 597–624.

20 J. Wu, H. Liu, X. Duan, Y. Ding, H. Wu, Y. Bai and X. Sun, *Bioinformatics*, 2009, **25**, 30–35.

21 S. A. Coulocheri, D. G. Pigis, K. A. Papavassiliou and A. G. Papavassiliou, *Biochimie*, 2007, **89**, 1291–1303.

22 B. Schneider, J. Černý, D. Svozil, P. Čech, J. C. Gelly and A. G. De Brevern, *Nucleic Acids Res.*, 2014, **42**, 3381–3394.

23 R. Sathyapriya, K. V. Brinda and S. Vishveshwara, *J. Chem. Inf. Model.*, 2006, **46**, 123–129.

24 N. M. Luscombe, R. a Laskowski and J. M. Thornton, *Nucleic Acids Res.*, 2001, **29**, 2860–2874.

25 N. M. Luscombe, S. E. Austin, H. M. Berman and J. M. Thornton, *Genome Biol.*, 2000, **1**, 1–10.

26 N. M. Luscombe and J. M. Thornton, *J. Mol. Biol.*, 2002, **320**, 991–1009.

27 D. Lejeune, N. Delsaux, B. Charloteaux, A. Thomas and R. Brasseur, *Proteins*, 2005, **61**, 258–271.

28 R. Sathyapriya, M. S. Vijayabaskar and S. Vishveshwara, *PLoS Comput. Biol.*, 2008, **4**, e1000170.

29 XCluster Macromodel Manual, 2012.

30 F. H. C. Watson, J. D.; Crick, *Nature*, 1953, 171, 737–738.

31 R. E. Franklin and R. G. Gosling, *Nature*, 1953, **171**, 740–741.

32 D. J. Walles, *Intermolecular Forces and Clusters I*, Springer Berlin Heidelberg, Heildelberg, 2005.

CHAPTER 6

AGGREGATION OF DNA BASES

6. - AGGREGATION OF DNA BASES

DNA is formed by a combination of sugar phosphates attached to four bases, cytosine, guanine, adenine and thymine (CGAT) or nucleotides. This four DNA bases play a variety of roles in cellular metabolism, but their main purpose is the storage of genetic information.[1–4] The reasons why these four molecules were chosen for this task is still a subject of discussion.[5–14] The canonical structure is the dimeric C-G/A-T base pairing. However, these structures are not the only ones found in DNA. Formation of trimers and quadruplexes of DNA bases is not unusual and formation of G-quadruplexes, i-motifs (made by the aggregation of cytosines) or thymine and adenine triplex and tetrads has been reported.[15–29] These observations rise the question of which are the real aggregation preferences of the DNA bases and which is the role that such preferences played in the appearance of the first DNA strands.

In this chapter, the conformational preferences in the aggregation of DNA bases were studied by a combination of spectroscopic techniques and theoretical calculations. Laser ablation and supersonic expansions were combined to create the required conditions for the DNA bases to aggregate. The collisions in the first instants of the expansion cool the molecules to ~100 K of vibrational temperature and stabilize the aggregates. Mass-resolved spectroscopic techniques: REMPI (resonance enhanced multiphoton ionization) and IR/UV double resonance were used to obtain the IR spectrum of the DNA aggregates, which were interpreted on the light of accurate calculations. This approach has been previously used in a large number of systems, including DNA bases.[30–38]

As stated above, the main goal of this study is to determine the natural tendency of the DNA bases during aggregation, without any solvent effect and the comparison of the two different interaction trends present in DNA: stacking vs. H-bonding. Furthermore, the conformational behavior of the aggregates was explored, evaluating the effect related with the cluster size. Moreover, the tautomeric preferences of the DNA bases and their aggregates were studied along with their dependence on the temperature and compared with that of their methylated derivatives in the position where the sugar attaches to the DNA base, although the latter will only be carried out at a theoretical level.

Two different DNA bases were considered: cytosine as a purine DNA base and guanine as a pyrimidinic DNA base. Two experimental set ups were used to obtain the experimental results: the one available at the

UPV/EHU and that available at FELIX facility. Detailed description of the experiments may be found in Section 2.1.3.

Regarding the calculations, they are described in the corresponding section. Just as a remainder, MMFFs was used as force field, while DFT calculations were conducted using M06-2X and Popple's double- and triple-ζ basis sets. The 15 most stable structures of each system were selected to calculate Gibb's free energy in a broad interval of temperatures: from 0 to 700 K, following the procedure described in section 3.2.5. Due to the large number of conformational isomers studied, we can only include a limited number of them along the text, but the structure of the 15 most stable conformers and the Gibbs free energy diagrams of each system may be found in Appendix 8.3. Finally, comparison between experimental and theoretical results will lead us to propose an assignment for the experimental spectra.

The different tautomers of DNA bases make difficult to find a simple systematic to name the aggregates. Following the convention in previous works,[30–36,39–41] we will label the aggregates as aXaYn, where a is only present in guanine complexes. It is related to the H in N7 or N9 position, which gives different tautomers, and it can take the numbers 7 or 9. Furthermore, X and Y refer to the keto-enol-imino tautomery and they can take k (keto), e (enol) or ik (iminoketo) values. Finally, n is related to the structure's energetic stability in the conformational search, according to the ZPE-corrected ΔH value. Furthermore, for methylated species, 1m- or 9m- is added in front of the labelling to indicate the methylation of position 1 (cytosine) or 9 (guanine), where the sugar is attached to form the nucleoside.

So a structure named 9k7e3 is the third structure in stability of the aggregate made by the combination of a keto tautomer of guanine with the hydrogen in position 9, and an enol tautomer of guanine with an hydrogen placed in position 7.

6.1 Cytosine

Several studies on cytosine isolated base and its dimers in supersonic expansions may be found in the literature, most of them carried out by the group of de Vries.[31,38,39,42] Some pure computational studies were also published, focusing in the formation of the dimers.[43–51] Furthermore, an exhaustive exploration of the aggregation preferences of keto-cytosine up to the hexamer appeared, using a force field fitted to reproduce the *ab-initio* computed structure of the dimer.[52]

In this section, we present a systematic study of the aggregation preferences of cytosine in supersonic expansions up to the tetramer, using a combination of laser spectroscopic techniques and theoretical calculations. The structure and the atom labelling of cytosine is presented in **Figure 6.1.**

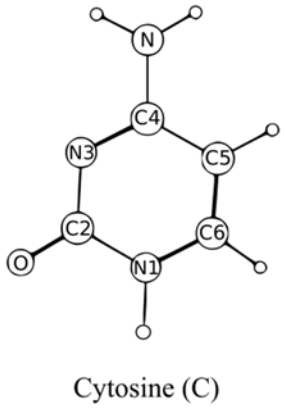


Figure 6.1 *Structure and atom labeling of cytosine.*

6.1.1 Spectroscopy of cytosine aggregates

Cytosine is one of the pyrimidinic DNA bases. It has different electronegative groups, such an NH_2 in C4, a C=O in C2 and a NH in position 1, as **Figure 6.1** shows. Thus, it presents several tautomers. Alonso et al.[53] found the five tautomers of the cytosine presented in **Figure 6.2** in supersonic expansions, using Fourier Transform Microwave (FT-MW) spectroscopy.

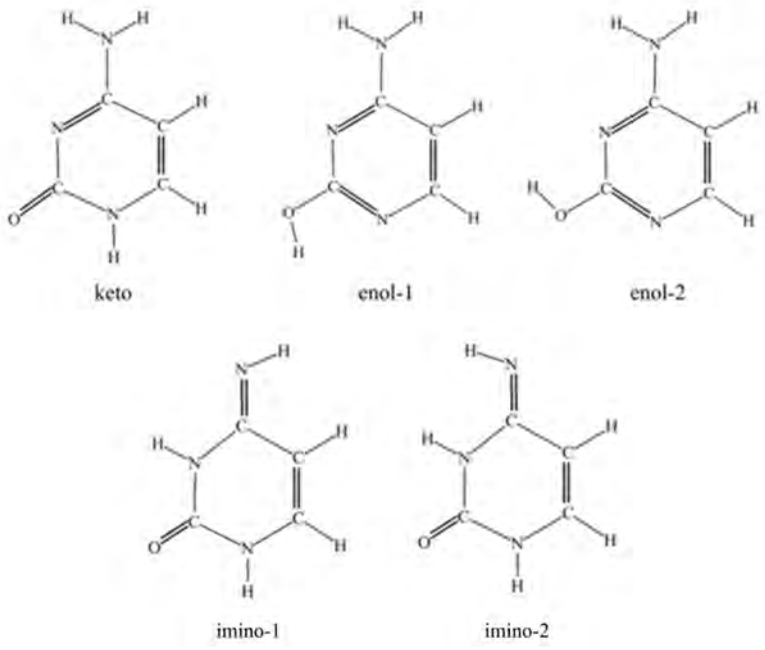


Figure 6.2 *Structure of the five most stable cytosine tautomers.*

The conditions of the expansion favored generation and detection of the two most abundant tautomers, which correspond to keto and enol forms. The enol tautomer also presents two different isomers, due to the rotation of the hydroxilic hydrogen.

These two tautomeric forms (keto and enol) present UV absorptions in different parts of the spectrum. While the keto form displays a 0_0^0 transition at 31826 cm^{-1} and a well-resolved spectrum with narrow bands, the enol tautomer presents a very low S/N ratio, what complicates the identification of the origin band, which lies around 36225 cm^{-1}.[31,39] The species also present very different $D_0^+ \leftarrow S_1$ transitions.[7,8]

REMPI spectra of both species have been reported in refs. 31 and 40 using 1+1' REMPI spectroscopy. An excimer laser emitting at 193 nm was used to obtain energetic-enough radiation to produce the ionization of the keto cytosine. Unfortunately, this type of laser is not available in our laboratory, and therefore we were not able to record the REMPI spectrum of keto-cytosine. On the other hand, the lower $D_0^+ \leftarrow S_1$ transfer of the enol tautormer permitted us to obtain the noisy trace of the enol monomer presented in **Figure 6.3**, in the 36200-36650 cm^{-1} region.

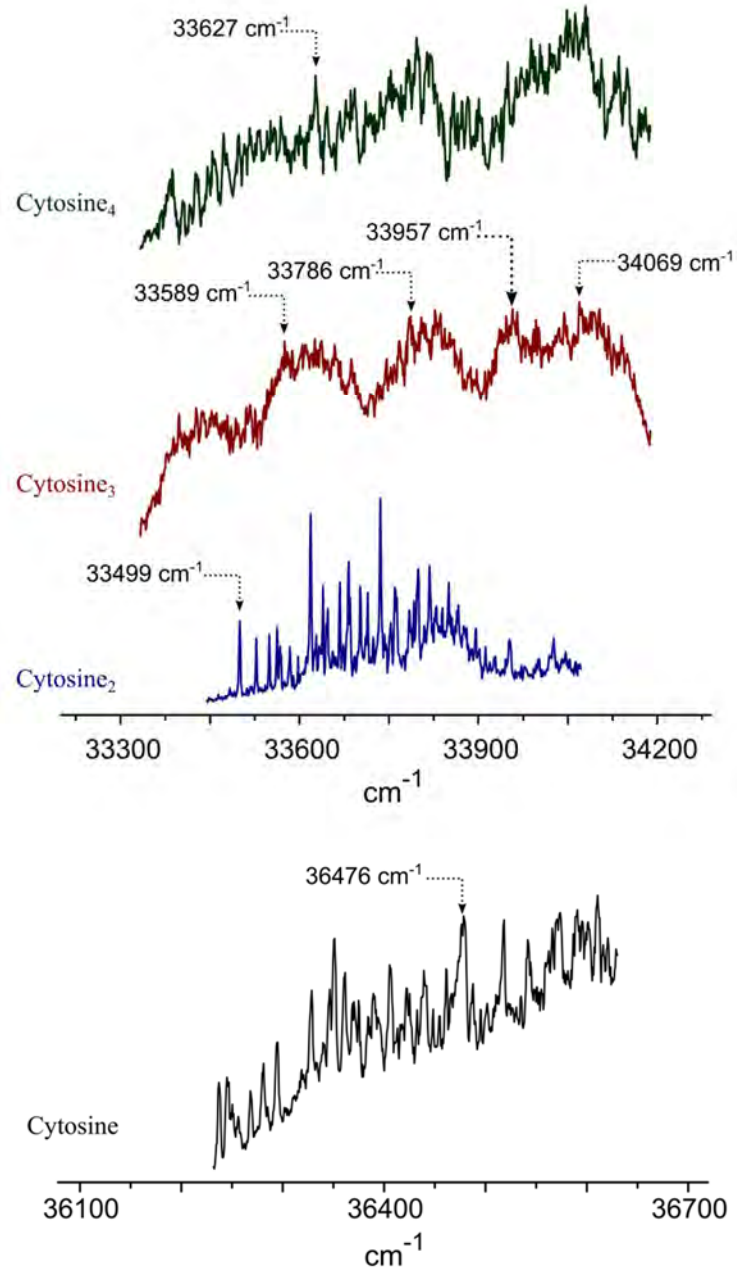


Figure 6.3 *REMPI spectrum of enol cytosine monomer (black) recorded between 36200-36650 cm^{-1}. REMPI spectra in the keto region for dimer (blue), trimer (red) and tetramer (green) of cytosine, scanned at 33300-34200 cm^{-1}. The arrows indicate the wavenumbers tested to record IR/UV spectra.*

Formation of the dimer usually results in a red shift both on the $S_1 \leftarrow S_0$ and $D_0^+ \leftarrow S_1$ transitions, allowing us to explore the spectroscopy of the complex built on the keto tautomer. Thus, the UV spectrum of cytosine dimer, which is shown in **Figure 6.3**, was collected in the surroundings of the 0_0^0 transition of the cytosine's keto tautomer, between 33300-34200 cm^{-1}. It presents a well-resolved spectrum with narrow bands and the first band appears at 33499 cm^{-1}, which is in good agreement with previous publications (33483 cm^{-1} in Ref. 31). Therefore, the appearance of the cytosine dimer in this region leads to another important conclusion: at least one of the two molecules of cytosine has to be in keto conformation.

The spectrum of the cytosine dimer in the vicinity of the enol-cytosine could not be recorded, perhaps due to a loss of the intensity in the signal or to the absence of enol tautomer as part of the dimer.

The electronic spectra of trimer and tetramer were also recorded in the region of the keto cytosine, confirming the presence of at least one monomer in keto form **(Figure 6.3)**. This is the first time that the spectrum of the trimer and tetramer of cytosine are reported in gas phase. The traces collected between 33300-34200 cm^{-1} present broad absorptions and a small red shift from the spectrum of the dimer. The broad absorptions may be related to dynamics of the excited state or to a large number of low-frequency vibrations. Although the REMPI spectra do not offer structural information on the clusters, they allowed us to record the IR/UV spectra collected in **Figures 6.3** and **6.4**.

The black spectrum corresponds to enol-cytosine monomer, as it was recorded tuning the UV laser at 36476 cm^{-1}. This is the first time that the IR spectrum of the enol monomer of cytosine is reported. Previous studies of the monomer were focused in the UV/UV hole burning and the IR/UV spectrum of the keto cytosine, and only the REMPI spectra of the enol tautomer was previously reported.[31,38,39] Three different peaks are observed at 3613, 3577 and 3462 cm^{-1}. This spectrum will be important to estimate the correction factors to account for the anharmonicity and that will be used for the rest of aggregates of cytosine.

Formation of the dimer resulted in drastic changes in the IR/UV spectrum **(Figure 6.4)**. The spectrum, recorded tuning the UV laser at 33499 cm^{-1} shows four well-resolved bands in the 3400-3700 interval and a broad absorption centered around 3000 cm^{-1}, probably related to the presence of one or several strong hydrogen bonds.

The addition of a third cytosine resulted in the appearance of a fifth peak and some shoulders between 3400-3600 cm^{-1}, while the broad band shifted slightly to the blue. Several UV wavelengths were probed in an

attempt to unveil additional isomers, hidden in the broad absorption of the REMPI spectrum (**Figure 8.3.** in **Appendix 8.3.15)**, but the spectra recorded at those alternative wavelengths were always similar.

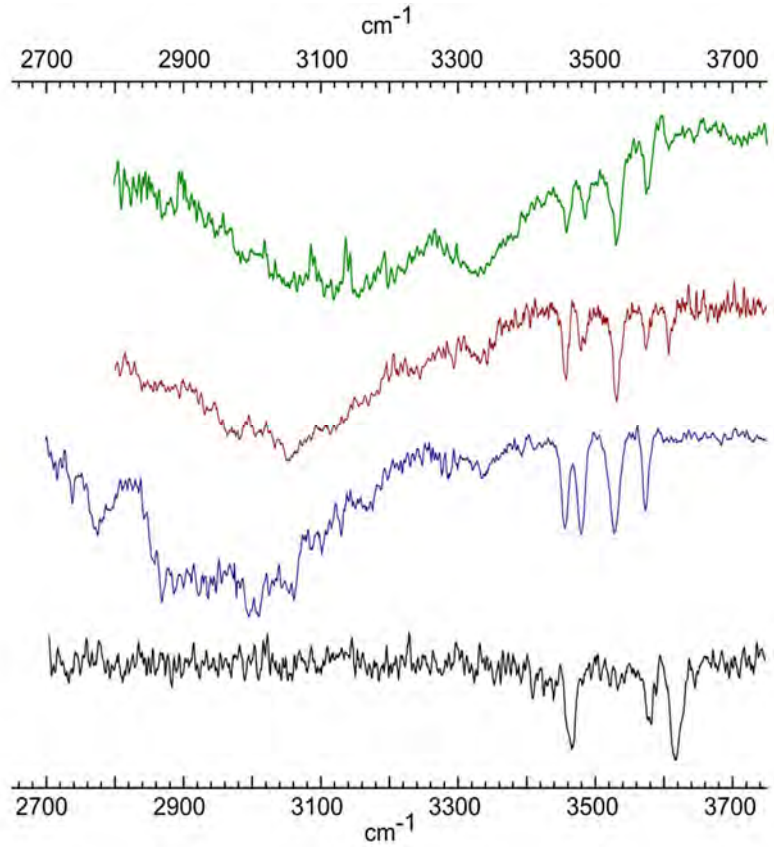


Figure 6.4 *IR spectra of enol monomer (black) and keto dimer (blue), trimer (red) and tetramer (green) of cytosine, scanned at 2700-3750 cm^{-1}.*

Finally, the green trace in **Figure 6.4** corresponds to the IR/UV spectrum of the tetramer. The addition of a fourth cytosine reduced the number of resolved bands and a second broad absorption, centered around 3333 cm^{-1}, and appeared between the first broad band and the narrow transitions. Furthermore, all the discrete bands present shoulders, which seem to suggest the presence of several isomers.

In the next sections, the theoretical study of each aggregate will be presented and the comparison with the experimental results will be discussed.

6.1.2 Cytosine monomer

The study of the Gibbs free energy of the monomers in the 0-700 K interval is presented in **Figure 6.5**. As the picture shows, both the keto and the enol species must be considered in the comparison with the experiments because they are very close in stability in the 100-200 K region, which is the temperature window of our expansion. Furthermore, the keto tautomer is favored by entropy, as the increase in relative stability with temperature indicates. In fact, it becomes isoenergetic with enol2 at 300 K and becomes even relatively more stable at higher temperatures. On the other hand, the imino forms are high enough in energy to rule their presence in the expansion out. Therefore, they will not be considered for the calculations of the aggregates.

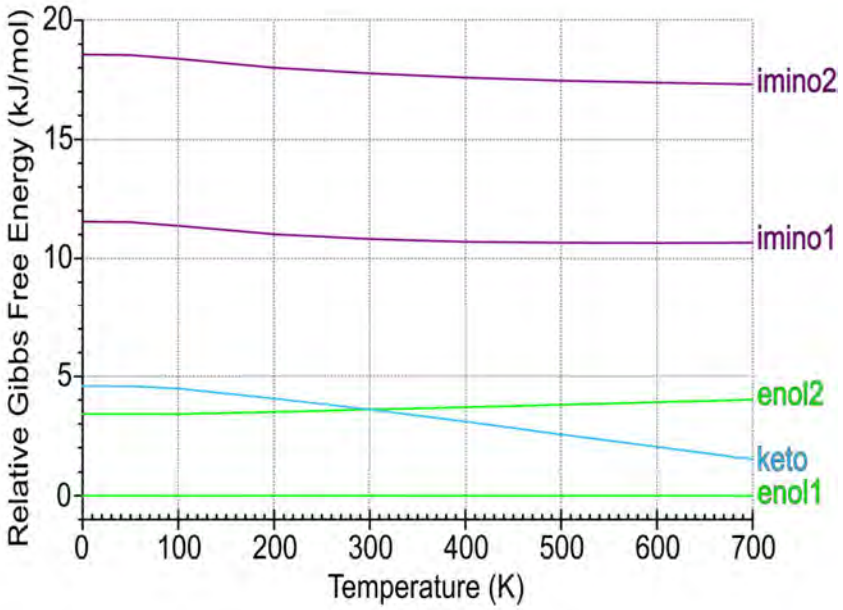


Figure 6.5 *Relative Gibbs free energy of the five possible tautomers of cytosine in the 0-700 K interval.*

Comparison of the simulated spectra of enol- and keto-cytosine with the experimental IR spectrum may be found in **Figure 6.6**, together with a ball&stick representation of the structure of the enol and keto tautomers and their relative energy (ΔH) in kJ mol^{-1}. The two enol tautomers only differ in the orientation of the hydroxyl group. Consequently, they are very close in stability, and their IR spectra are very similar and reproduce

reasonably well the experimental trace. On the other hand, the presence of shoulders on the peaks suggests the co-existence of both isomers in the expansion.

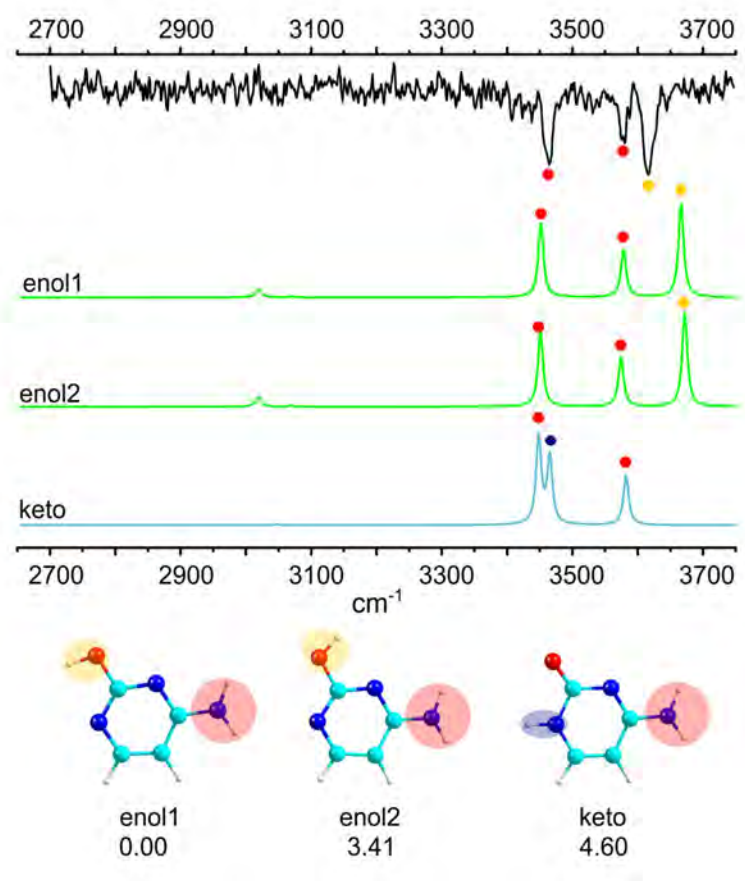


Figure 6.6 *Comparison between the experimental IR spectrum (black trace) and the theoretical prediction of the three most stable structures of the cytosine (Green traces: enol isomers; Blue trace: keto isomer). The structure of the three species has been presented with their relative energy in kJ mol^{-1}.*

The theoretical calculations reveal that the peaks at 3577 and 3462 cm^{-1}, highlighted with red dots in the figure, are due to the symmetric and antisymmetric stretching of the NH_2 group respectively and suit perfectly the theoretical predictions. On the other hand, the third band corresponds to the stretching of the hydroxyl group and the position of the band is shifted due to the anharmonic nature of the vibration. No single scaling

factor permits a perfect match between experimental and theoretical spectra. As a compromise, a scaling factor of 0.9483 was selected, which will be used for scaling the rest of the simulated spectra of cytosine aggregates.

6.1.3 Cytosine dimer

As mentioned in the previous section, cytosine dimer was measured in the keto zone, which means that at least one of the molecules has to be in keto form. However, enol-enol dimer was also considered in order to study the relative stability of all possible aggregates. Hence, 10, 21 and 31 structures of keto-keto, keto-enol and enol-enol dimers were respectively subjected to full optimization. The study of the Gibbs free energy of the most relevant aggregates in the 0-700 K range may be found in **Figure 6.7**. Clearly, keto-keto dimers are more stable than the rest of the aggregates. In the region of interest, the two most stable keto-keto dimers are isoenergetic and therefore their presence in the expansion, and consequently in the spectrum, cannot be ruled out *a priori*.

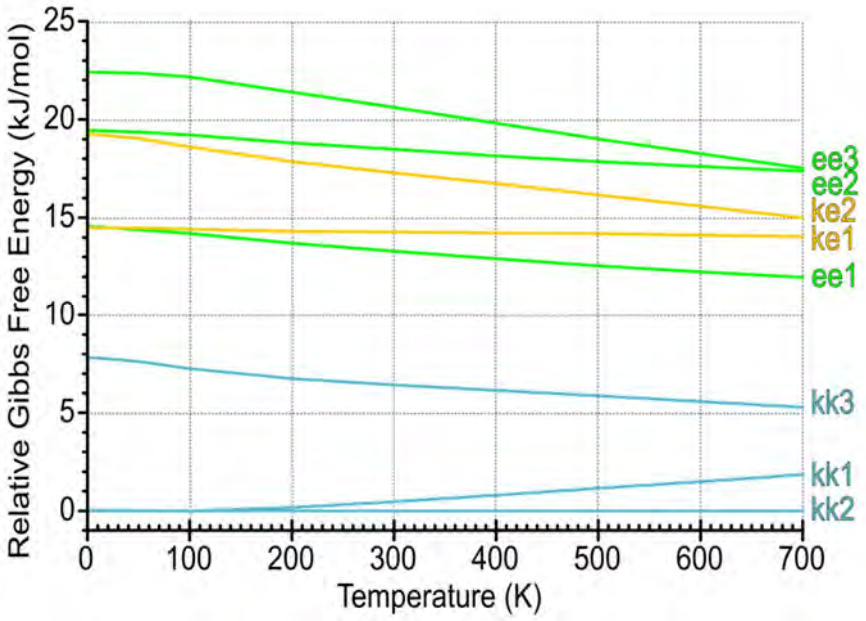


Figure 6.7 *Relative Gibbs free energy of some selected calculated species of cytosine dimer in the 0-700 K interval. Blue: keto-keto; green: enol-enol; yellow keto-enol isomers.*

Focusing the study in the aggregation tendency, **Figure 6.8** highlights two different trends: planar hydrogen-bonded dimers, depicted with continuous line, and stacking structures (dotted lines). The latter are more than 40 kJ mol^{-1} less stable in the case of homodimers (keto-keto and enol-enol) and around 30 kJ mol^{-1} for mixed keto-enol dimers. This tendency increases with the temperature.

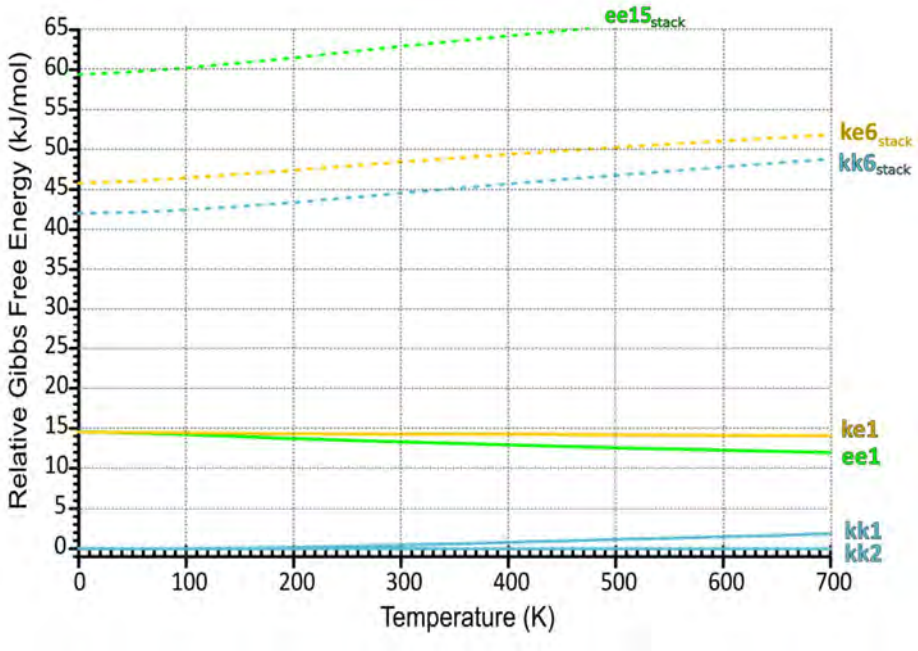


Figure 6.8 *Relative Gibbs free energy of the most stable planar H-bonded and stacked conformations of cytosine dimer. Continous line: planar structure; Dot line: stacking structure. Blue color: keto-keto; green: enol-enol; yellow keto-enol structures.*

Finally, the comparison with the experimental results was collected in **Figure 6.9**. The conformers selected were those below 15 kJ mol^{-1} in the 100-200 K interval. The selected structures included the most stable species of keto-enol and enol-enol aggregates, although only the keto-keto aggregates are stable-enough to present a significant population in the beam. As can be seen in the figure, the experimental spectrum can be fully explained by the most stable keto-keto structure: the four bands can be assigned to the symmetric and antisymmetric stretches of the free NH_2 (3575 and 3457 cm^{-1}), the stretching of the NH (3481 cm^{-1}) and the stretching of the free hydrogen in the bounded NH_2 (3530 cm^{-1}). The

second conformer, which is very close in energy (0.49 kJ mol^{-1}) and becomes more stable above 200 K, presents only two discrete bands, which can be overlapping with the bands of the most stable conformer. Consequently, its presence in the gas expansion cannot be excluded.

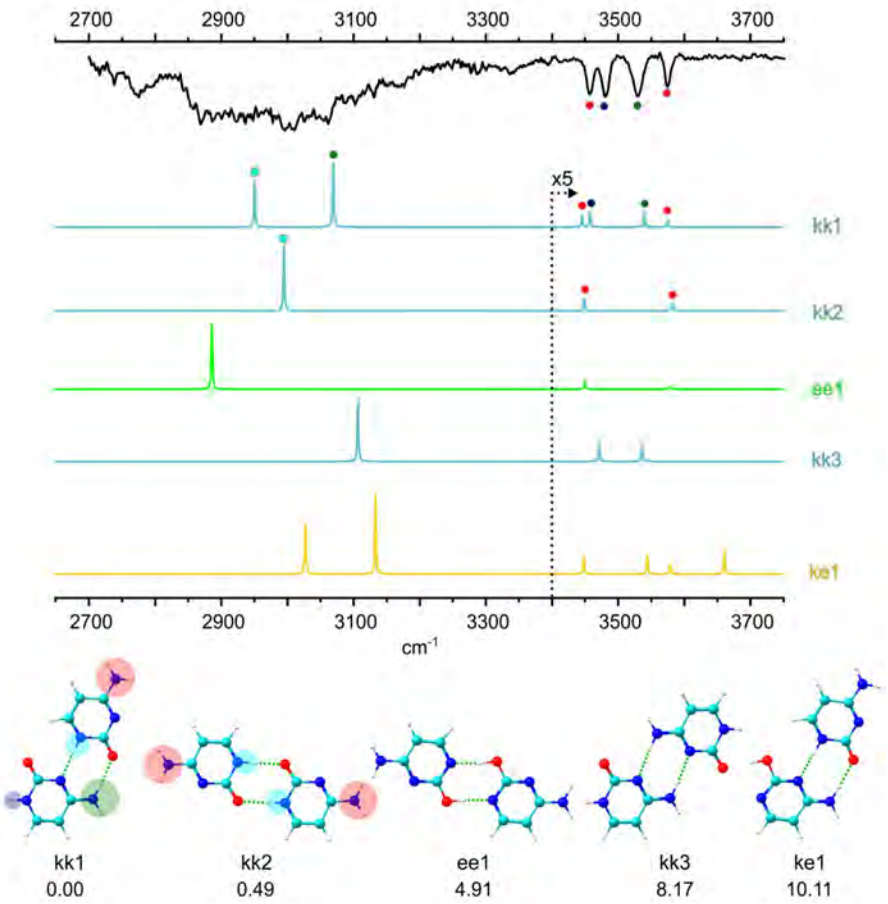


Figure 6.9 *Comparison between the mass-resolved IR/UV spectrum of cytosine dimer and the simulated spectra of the five more probable species that may be contributing to the spectrum. The structure of these species and their relative energies in kJ* mol^{-1} *are also included.*

The broad absorption between 2900-3250 cm^{-1} corresponds to the stretch of the NH groups involved in intermolecular interactions: two bands in the kk1 dimer and only one band in kk2. Thus, the shape of the band and the large shift experienced by this vibration indicates a surprisingly strong interaction and a high anharmonicity. In fact, theoretical calculations predict an intensity for this band ~17-20 times

stronger than the most intense band on the 3400-3650 cm^{-1} region. Both effects may be due to a partial delocalization of the proton between the molecules, as previous studies on protonated cytosines demonstrated.[54]

Both keto-keto structures exhibit two hydrogen bonds between the members. However, the hydrogen bond pattern is different: kk1 isomer presents $HN\text{-}H^{I}\cdots O{=}C_2{}^{II}$ and $N_3{}^{I}\cdots H\text{-}N_1{}^{II}$ hydrogen bonds, while kk2 displays two symmetrical hydrogen bonds between $N_1\text{-}H^{I}\cdots O{=}C_2{}^{II}$ and $C_2{=}O^{I}\cdots H\text{-}N_1{}^{II}$. **Figure 6.10** collects the structure of both isomers, together with the binding energy values, the distance of the hydrogen bonds and the representation of the non-covalent interactions. It is worthy to note the extraordinary strength of the hydrogen bonds, displayed as dark blue zones in the NCI plots, and as the binding energy values highlight, also confirmed by the distances and the electron density values collected in **Table 6.1**.

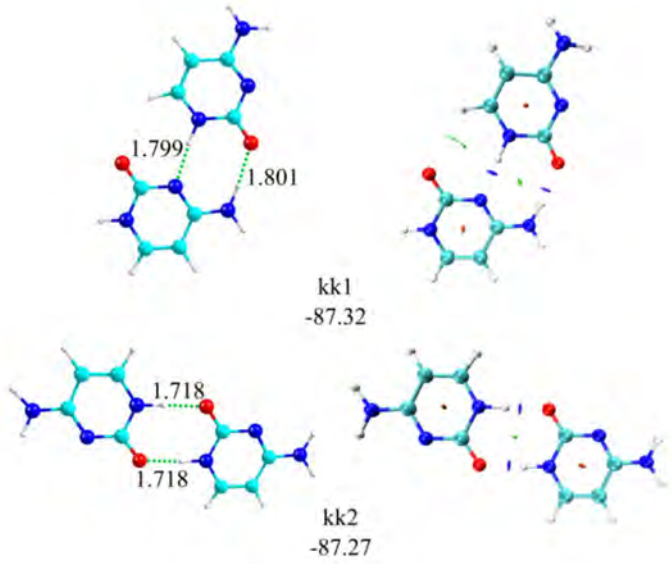


Figure 6.10 *kk1 and kk2 dimers of cytosine, together with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions depicted in NCI plots. Binding energy values (kJ* mol^{-1}*) are also shown.*

Finally, the structure of the second isomer also explains the absence of the enol tautomers in the dimers of the cytosine. Focusing in the structure of the most stable enol-enol isomer, it can be easily interconverted into the kk2 isomer throughout a double proton transfer. Furthermore, the most stable mixed keto-enol dimer presents more bands than the experimental IR spectrum, what permits us to discard it. The rest

of keto-enol and enol-enol species lay very high in energy compared with pure keto-keto structures and their presence in the expansion can be safely ruled out.

Table 6.1. *Electron density (a.u.) of the hydrogen bonds observed for the kk1 and kk2 aggregates.*

	Electron Density (a.u.)	
	kk1	**kk2**
HN-H^{I}···O=C2^{II}	0.0334	
N3^{I}···H-N1^{II}	0.0417	
C2=O^{I}···H-N1^{II}		0.0422
N1-H^{I}···O=C2^{II}		0.0422

6.1.4 Cytosine trimer

The addition of a third cytosine makes the conformational landscape more complicated. Taking into account that the first enol-enol dimer can easily interconvert into the keto-keto dimer, that the remaining enol-enol dimers appeared at high energy and that the aggregate needs to have at least one cytosine in keto form, the pure enol trimer has not been considered. Therefore, 29, 41 and 57 structures for keto-keto-keto, keto-keto-enol and keto-enol-enol respectively were analyzed in the conformational study. **Figure 6.11** displays the variation of the Gibbs free energy of the aggregates along the selected temperature interval. The pure keto trimers are significantly more stable than the mixed keto-enol trimers. Thus, the first keto-keto-enol conformer appears at 14 kJ mol^{-1} above the global minimum and the first keto-enol-enol trimer appears even higher in energy, at 30 kJ mol^{-1}. Focusing in the region of 100-200 K, it is observed that kkk1 and kkk2 are clearly the most probable structures.

Figure 6.12 shows a comparison between the relative stability of planar and stacked structures for the trimers. The tendency observed in the dimers is maintained also in the trimers: stacking structures are less favored than planar ones, formed by hydrogen bonds. However, the difference between them is smaller than in the dimers: trimers in planar

conformation are separated by 16-18 kJ mol^{-1} from the stacking ones. Nevertheless, this difference increases with the temperature, as in the dimers. Apparently, the former are favored by entropy.

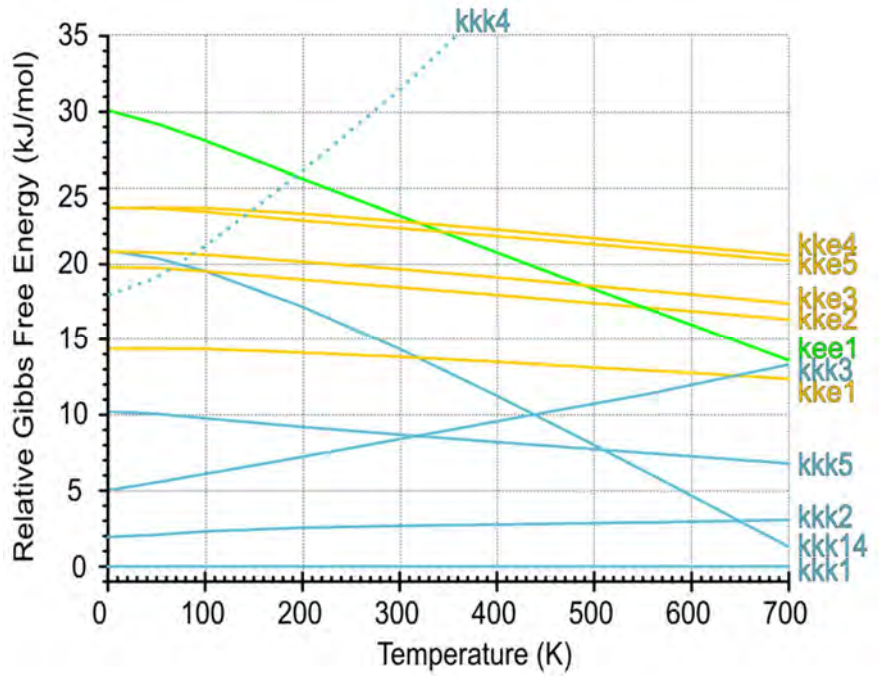


Figure 6.11 *Relative Gibbs free energy of the trimers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure. Blue color: keto-keto-keto; green: keto-enol-enol; yellow keto-keto-enol structures.*

Comparison with the experimental results, **Figure 6.13**, further confirms that pure keto trimers are formed in the beam. The spectrum predicted for the global minimum kkk1 alone is not able to reproduce all the peaks observed in the experimental trace. This conformer presents four narrow peaks: 3566 and 3438 cm^{-1}, related to the symmetric and antisymmetric stretch of the free NH_2; the band at 3538 cm^{-1} is due to the stretching of the free N-H of the bonded NH_2 and the peak at 3472 cm^{-1} corresponds to the free NH. However, the contribution of the second most stable structure is necessary to explain the remaining feature at 3608 cm^{-1}. In fact, the experimental spectrum can be totally explained by this second isomer. Nevertheless, the presence of kkk1 cannot be ruled out, due to the overlapping of the bands from both isomers and the small difference in

stability: 1.9 kJ mol^{-1}, which probably is within the calculation error. Furthermore, the combination of the bands of both conformers (at 3472 and 3453 cm^{-1}) can also explain the doublet at 3481-3486 cm^{-1}, reinforcing the idea of the presence of both isomers in the expansion.

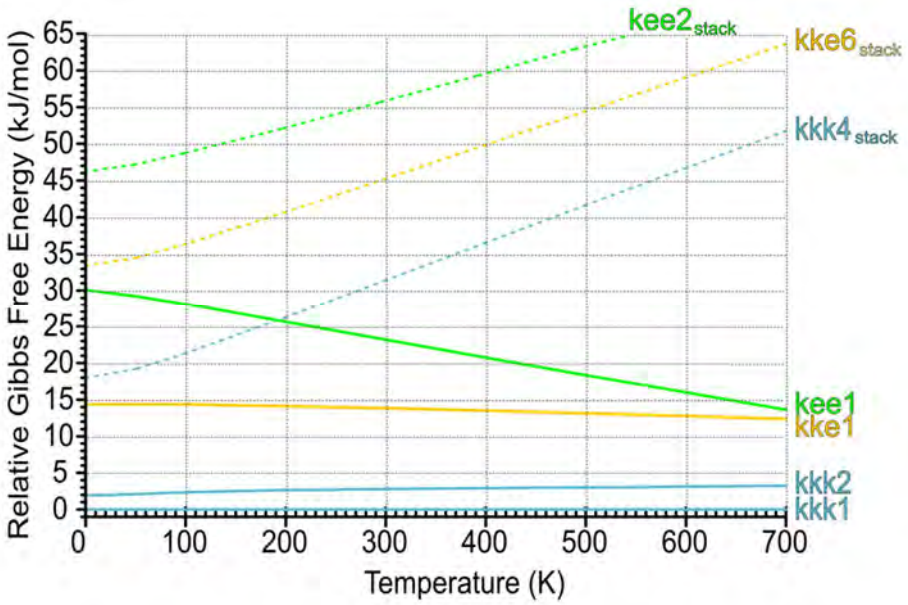


Figure 6.12 *Relative Gibbs free energy of the most stable planar H-bonded and stacked conformations of cytosine trimer. Continous line: planar structure; Dot line: stacking structure. Blue color: keto-keto-keto; green: keto-enol-enol; yellow keto-keto-enol structures.*

The band at 3608 cm^{-1} raises questions about a possible enol contribution to the spectrum, as a similar band was recorded by Nir et al. for cytosine-5-methylcytosine dimer, which was assigned to a free OH.[31] However, there are two evidences against this hypothesis. On the one hand, as stated above, the trimers containing enol tautomers appear at relatively high energy (15 kJ mol^{-1}), and therefore one would not expect to find a large population of those species in the beam. On the other hand, the absence of enol species in the dimer suggests that the tautomerization process efficiently convert any enol tautomer into a different tautomer when the complex is formed. A final test would be to record the IR/IR/UV triple resonance spectrum;[55] however, the low signal intensity of this system precluded the use of such complicated technique.

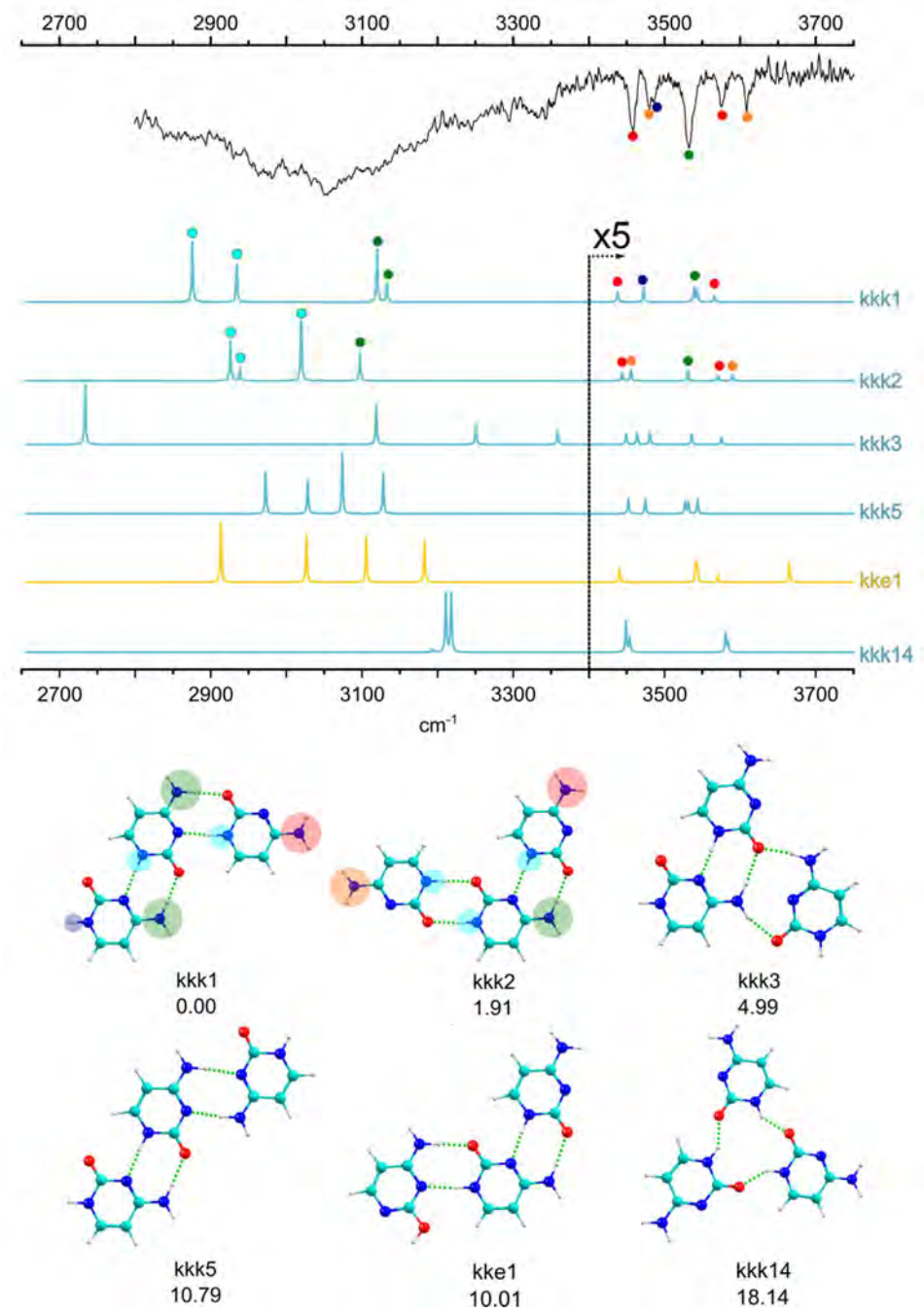


Figure 6.13 *Comparison between the mass-resolved IR/UV spectrum of cytosine trimer and the simulated spectra of the six more probable species that may be contributing to the spectrum. The structure of these species and their relative energies in kJ mol^{-1} are also included.*

The broad absorption between 2800-3350 cm^{-1} is due to the stretch of the NH groups involved in intermolecular interactions, as in the case of the dimer. In fact, this absorption is broader than in the dimer. A careful analysis of the kkk1 and kkk2 structure reveals that both conformers are formed by a third keto-cytosine interacting with a kk2 kernel: In the kkk1 trimer, the third cytosine interacts through $HN\text{-}H^{II}\cdots O{=}C_2{}^{III}$ and $N_3{}^{II}\cdots H\text{-}N_1{}^{III}$ hydrogen bonds, forming a second kk1 dimer, while, kkk2 trimer is a combination of kk1 plus kk2 dimers and the third cytosine exhibits $C_2{=}O^{I}\cdots H\text{-}N_1{}^{III}$ and $N_1\text{-}H^{I}\cdots O{=}C_2{}^{III}$ hydrogen bonds, as in kk2 dimer. The structure of both isomers, together with the binding energy values, the distance of the hydrogen bonds and the NCI plots were collected in **Figure 6.14**. The extraordinary strength of the hydrogen bonds observed in the dimers is maintained in the trimers, as the values of the electron density collected in **Table 6.2** show. These evidences point to a sequential mechanism for the aggregation of cytosine.

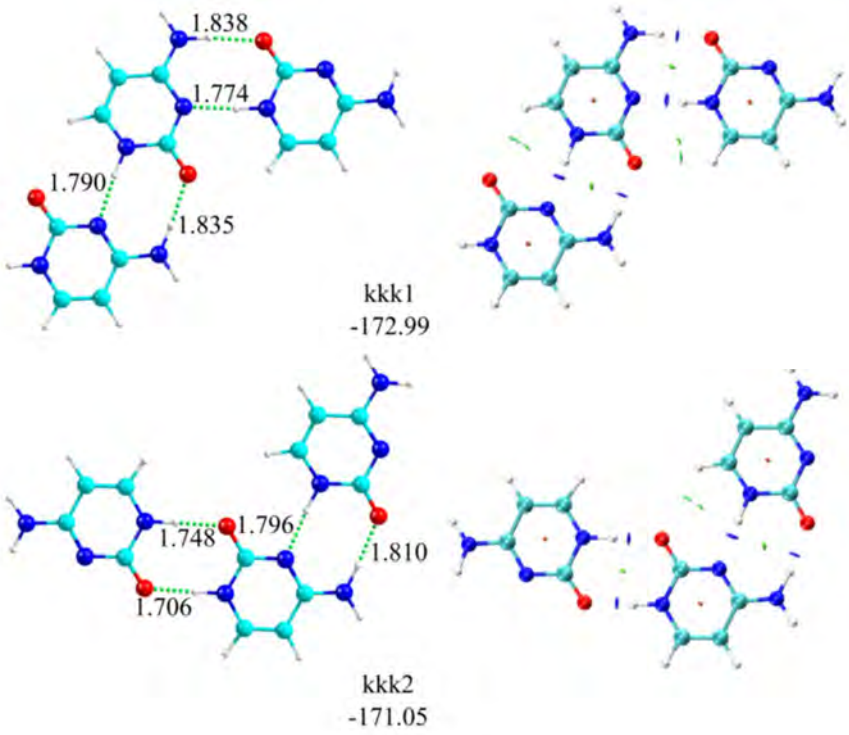


Figure 6.14 *kkk1 and kkk2 trimers of cytosine, together with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions depicted in NCI plots. Binding energy values (kJ* mol^{-1}*) are also shown.*

Table 6.2. *Electron density (a.u.) of the hydrogen bonds observed for the kkk1 and kkk2 aggregates. The dash line signs the two similar hydrogen bonds present in both aggregates, related to the structure of the kk1 dimer.*

	Electron Density (a.u.)	
	kkk1	**kkk2**
HN-H^{I}···O=C2^{II}	0.0309	0.0327
N3^{I}···H-N1^{II}	0.0427	0.0423
HN-H^{II}···O=C2^{III}	0.0305	
N3^{II}···H-N1^{III}	0.0447	
C2=O^{I}···H-N1^{III}		0.0393
N1-H^{I}···O=C2^{III}		0.0435

6.1.5 Cytosine tetramer

As mentioned above, there was a reduction in the number of the resolved bands in the spectrum from the trimer to the tetramer. In addition, a second broad band appeared in the experimental IR spectrum. However, from a theoretical point of view, the conformational landscape is more complicated than in the previous cases. Four different combinations, pure keto tetramer and mixed keto-enol tetramers were considered for full optimization, amounting 80, 70, 67 and 72 structures for kkkk, kkke, kkee and keee respectively. The study of the Gibbs free energy was collected in **Figure 6.15**. Pure keto dimers are favored over the rest of the combinations in the whole interval of temperatures, as observed in previous aggregates. In fact, the first keto-keto-keto-enol structure appears at 15 kJ mol^{-1} and the first keto-keto-enol-enol appears above 33 kJ mol^{-1}.

Furthermore, the study of the aggregation pattern collected in **Figure 6.16**, reveals a change in the previous tendency. At very low temperatures, the most stable stacking and planar structures exhibit similar values of ΔG. However, a minimum increase in temperature is enough to favour the planar structures, following the tendency observed in previous aggregates. On the other hand, the difference in stability between stacking and planar structures is in general smaller and planar structures of keto-enol-enol-enol were not found.

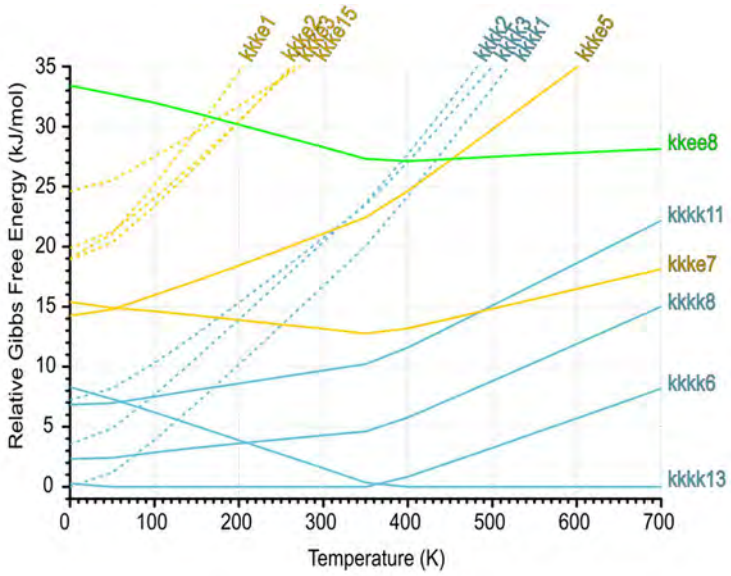


Figure 6.15 *Relative Gibbs free energy of the tetramers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure. Blue: keto-keto-keto-keto; green: keto-keto-enol-enol; yellow: keto-keto-keto-enol structures.*

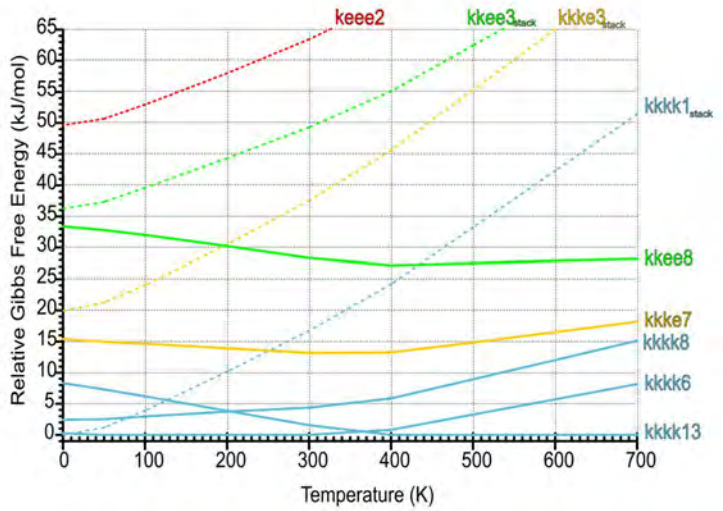


Figure 6.16 *Relative Gibbs free energy of the most stable planar H-bonded and stacked conformations of cytosine tetramer. Continous line: planar structure; Dot line: stacking structure. Blue: keto-keto-keto-keto; green: keto-keto-enol-enol; yellow: keto-keto-keto-enol; red: keto-enol-enol-enol structures.*

Finally, the most stable structure at 0 K, kkkk1, adopts a stacking disposition, as the dot blue line depicts. In fact, the first planar structure is kkkk6. Above 350 K, another planar conformer, kkkk13, becomes the most stable structure (see **Figure 6.17**). Therefore, a competition between these two different trends of aggregations is clearly observed: planar structures are more favored by entropy and stacking structures are more stable.

Figure 6.17 also offers a comparison between the experimental IR/UV spectrum of cytosine tetramer and the simulated spectra for some selected structures. Thus, planar structures are expected to be the most abundant species around 100-200 K, although the stacking structures were also taken into account. As it can be seen, the whole spectrum can be explained by isomer kkkk6, the most favored conformer by entropy at 100 K, except for the broad absorption centered at 3333 cm^{-1}, shaded in yellow in **Figure 6.17**. The second most stable isomer kkkk8, can also explain the shoulders observed in the spectrum, but still it does not present any band that could explain the broad absorption. In fact, none of the planar structures in a reasonable range of stability can explain that feature, and a stacking structure is required. Certainly, conformer kkkk1, which appears around 4 kJ mol^{-1} at the temperature of the expansion, presents an intense band at 3413 cm^{-1} in its simulated spectrum that fits reasonably well the broad absorption in the experimental spectrum. Thus, the IR spectrum may be pointing to the co-existence on the beam of planar and stacked structures of cytosine tetramer.

The structures of the pure keto tetramers (kkkk conformers) calculated in this work may be found in **Figure 6.18**, together with the hydrogen bond distances and binding energies and the NCI plots. Clearly, the stacked structure is formed by two kk1 dimers, held together mainly by π···π interactions, although some weak hydrogen bonds between the dimers are also formed. This pattern of hydrogen bonds is also observed in kkkk2 and kkkk3 tetramers. The values of the electron density of the hydrogen bonds may be found in **Table 6.3**, while the NCI plots highlight the nature of the interactions: hydrogen bonds, appear as dark blue lentils and π···π interaction as extended green surfaces. According to the electron density, $N3^{I}\cdots H\text{-}N1^{II}$ / $N3^{III}\cdots H\text{-}N1^{IV}$ are very strong, with a value of 0.0476 a.u in comparison to the other hydrogen bonds of the NH_2 groups (0.0200 and 0.0186 a.u).

The mechanism that leads the formation of such structures in the gas phase is not clear. They can be formed by the collision of dimers or the addition of one cytosine to a trimer structure. Although the most abundant

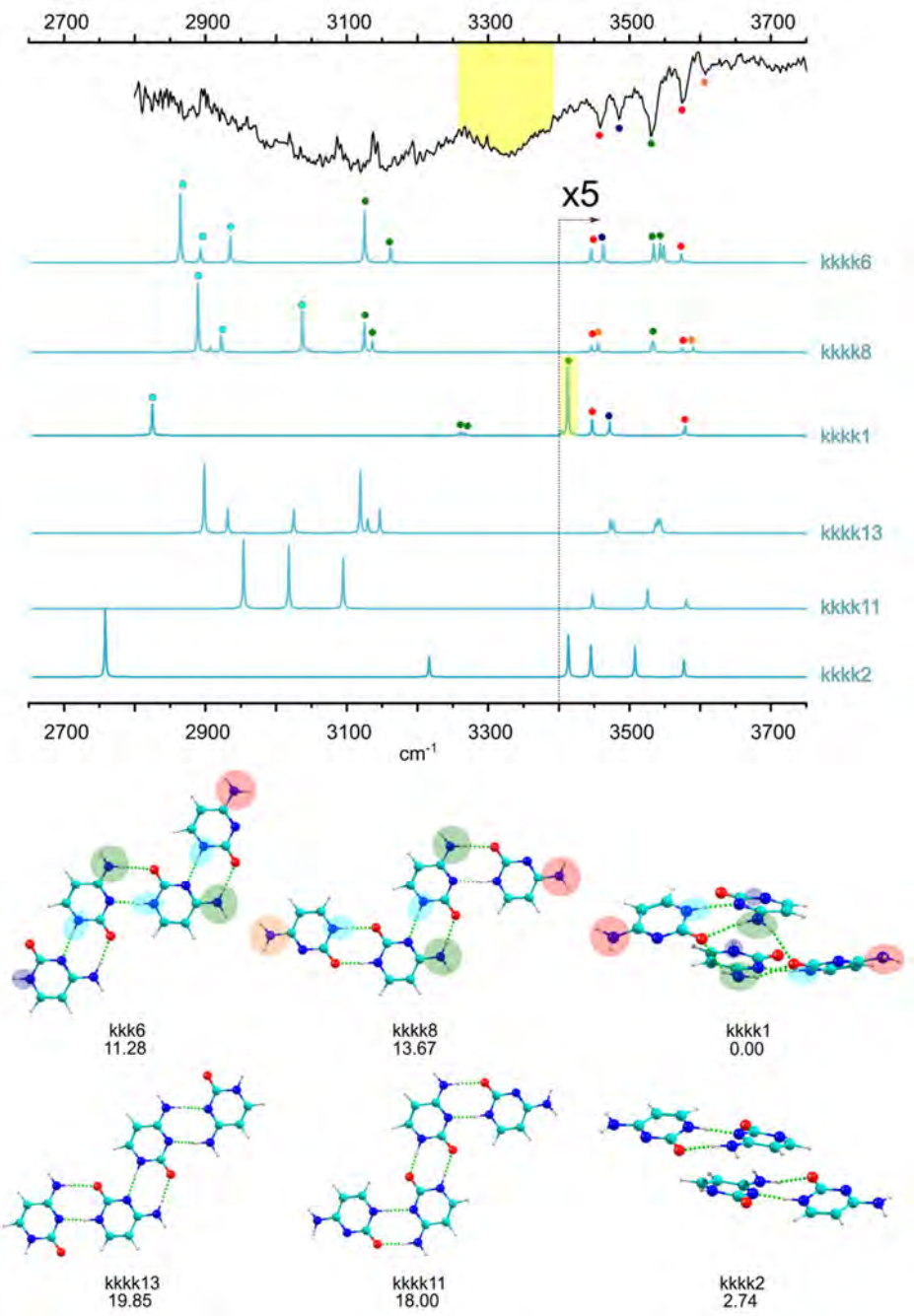


Figure 6.17 *Comparison between the mass-resolved IR/UV spectrum of cytosine tetramer and the simulated spectra of the six more probable species. The structure of these species and their relative energies in kJ mol*$^{-1}$ *are also included. The broad absorption between 3250-3400 cm*$^{-1}$ *and its explanation were depicted in yellow.*

species is the monomer, dimers are significantly more abundant than trimers. However, the comparison of the collision cross section for the aggregation of a monomer plus a trimer or for collision of two dimers is difficult. Thus, both mechanisms need to be considered as responsible for the detection of the staking structures. In any case, the geometry of the kk1 dimer seems to govern the aggregation of cytosine.

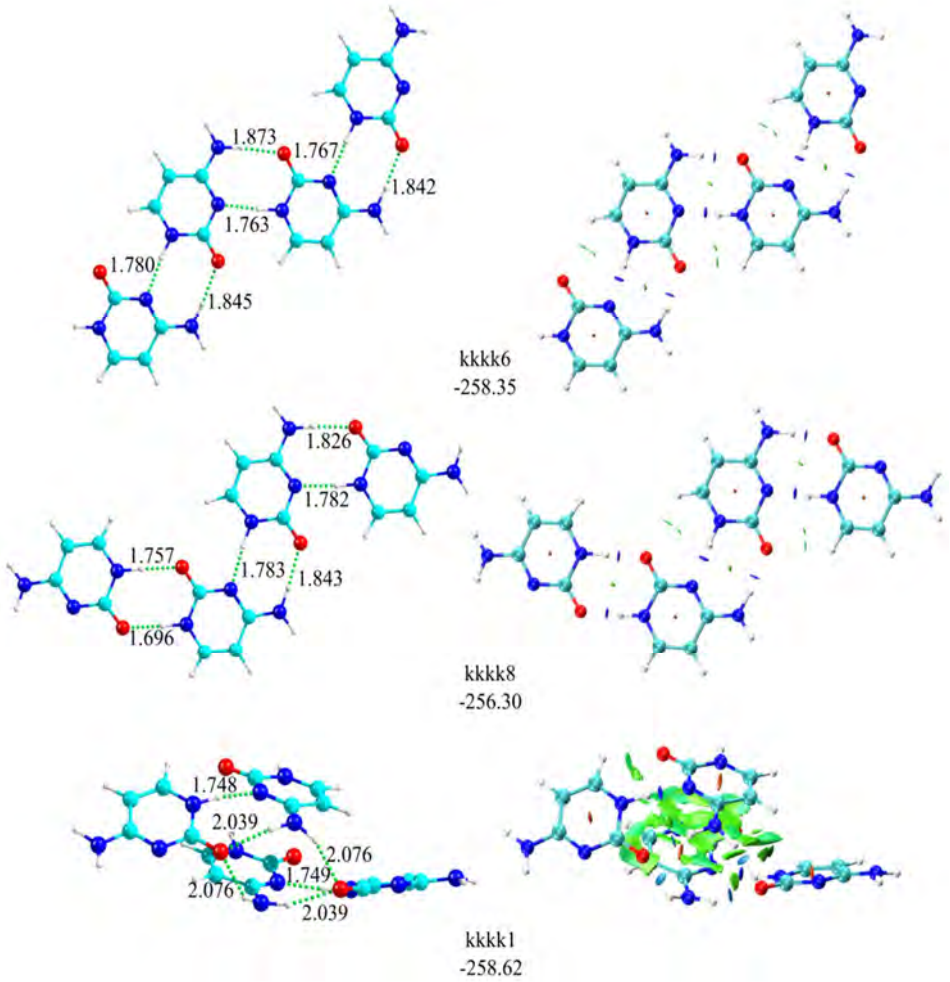


Figure 6.18 *kkkk1, kkkk6 and kkkk8 tetramers of cytosine, together with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions depicted in NCI plots. Binding energy values (kJ mol^{-1}) are also shown.*

Table 6.3. *Electron density (a.u.) of the hydrogen bonds observed for the kkkk1, kkkk6 and kkk8 aggregates. The first dash line signs the first kk1 dimer structure and the second dash line signs the structure kkk1 trimer, both observed in the structure of the tetramer.*

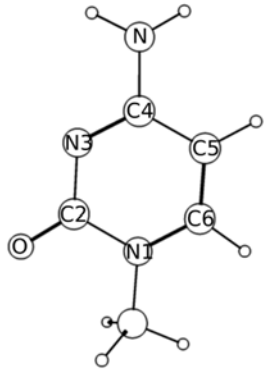
	Electron Density (a.u)		
	Planar structures		Stacking structures
	kkkk6	kkkk8	kkkk1
$HN\text{-}H^{I}\cdots O{=}C2^{II}$	0.0302	0.0304	0.0200
$N3^{I}\cdots H\text{-}N1^{II}$	0.0438	0.0437	0.0476
$HN\text{-}H^{II}\cdots O{=}C2^{III}$	0.0282	0.0315	
$N3^{II}\cdots H\text{-}N1^{III}$	0.0460	0.0438	
$HN\text{-}H^{III}\cdots O{=}C2^{IV}$	0.0303		0.0200
$N3^{III}\cdots H\text{-}N1^{IV}$	0.0454		0.0476
$C2{=}O^{I}\cdots H\text{-}N1^{IV}$		0.0384	
$N1\text{-}H^{I}\cdots O{=}C2^{IV}$		0.0447	
$HN\text{-}H^{I}\cdots O{=}C2^{IV}$			0.0186
$HN\text{-}H^{III}\cdots O{=}C2^{II}$			0.0186

6.2 1-methylcytosine

6.2.1 1-methylcytosine monomer

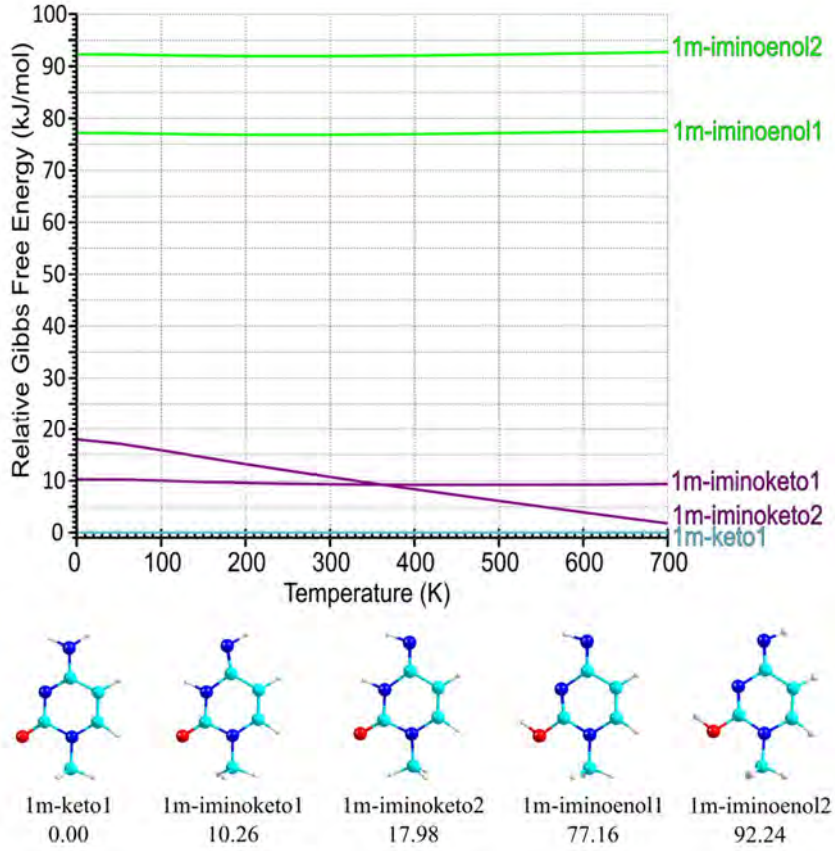
This part of the study was purely computational, although some experimental results may be found in the literature, all of them related to the monomer or to its aggregation with pure cytosine.[31,39,56,57] The structure and the atom labelling of the 1-methylcytosine (1-mC) is presented in **Figure 6.19.**

The methylation in N1 position reduces the number of possible tautomers to only five, as **Figure 6.20** shows. According to $\Delta(\Delta G)$, the keto tautomer is the most stable structure in the whole interval of temperatures, followed by imino tautomers, which appear at 10.3 and 18.0 kJ mol^{-1} respectively. Despite that imino2 becomes relatively more stable with temperature, it is high-enough in energy to rule out its presence in the expansion.



1-Methylcytosine (C)

Figure 6.19 *Structure and atom labeling of 1-methylcytosine (1-mC).*



. ***Figure 6.20*** *Relative Gibbs free energy of the five possible tautomers of 1-methylcytosine in the 0-700 K interval and the structure of these four tautomers with their relative energy in kJ mol^{-1}.*

The low stability of the enol tautomer is remarkable, especially taking into account that enol tautomers were the most stable species in cytosine. Here, enol1 and enol2 tautomers appear at 77.16 and 92.24 kJ mol^{-1} respectively, which clearly discard their presence in the beam.

6.2.2 1-methylcytosine dimer

As the keto-tautomer is significantly more stable, we focused only in keto-keto dimers. The conformational search located a total of 23 possible conformers of 1-mC dimer. **Figure 6.21** collects the variation of the relative Gibbs free energy with the temperature, together with the structure of the more relevant dimers and their relative stability in kJ mol^{-1}. Surprisingly, 1m-kk1 dominates the conformational space, followed by 1m-kk2, a stacking structure, at a relatively high energy of 19 kJ mol^{-1}. Apparently, the methyl substitution greatly simplifies the conformational landscape.

In addition, the study of the aggregation tendency underlines that planar structures are the preferred ones in all the temperatures. In fact, the first stacking structure appears at 19 kJ mol^{-1}. Nevertheless, the energy difference between stacking and planar structures is lower than in the case of cytosine dimer, for which a difference of 42 kJ mol^{-1} was found. Furthermore, the number of planar structures found was significantly smaller compared with cytosine. The methylation of the N_1 position eliminates an important point of contact observed in cytosine dimer. Consequently, the two most stable structures of cytosine dimer disappeared and the first structure observed in 1-mC dimers corresponds to cytosine kk3 conformer, as **Figure 6.22** shows. It is also worthy to note that another planar conformer, 1m-kk12, was found at 39 kJ mol^{-1}; the rest of the conformers present a variety of aggregation pattern, as displayed in **Figure 6.21**.

The 1m-kk1 dimer exhibits two symmetric hydrogen bonds: HN-$H^{I}\cdots N_3{}^{II}$ and $N_3{}^{I}\cdots$H-NH^{II}, highlighted in **Figure 6.22**. Comparison of the electron density values between cytosines and 1-methylcytosine dimers may be found in **Table 6.4**. Both sets of values are very similar, underlying the analogous character of both structures. Therefore, when N1 position is blocked, cytosine tends to form planar structures by formation of $H^{I}\cdots N_3{}^{II}$ and $N_3{}^{I}\cdots$H-NH^{II} hydrogen bonds, adopting a kk3 disposition. If N1 position is available for docking, the unusual strength of the interactions with N1-H guides the molecules towards formation of kk1- or kk2-like cytosine dimers.

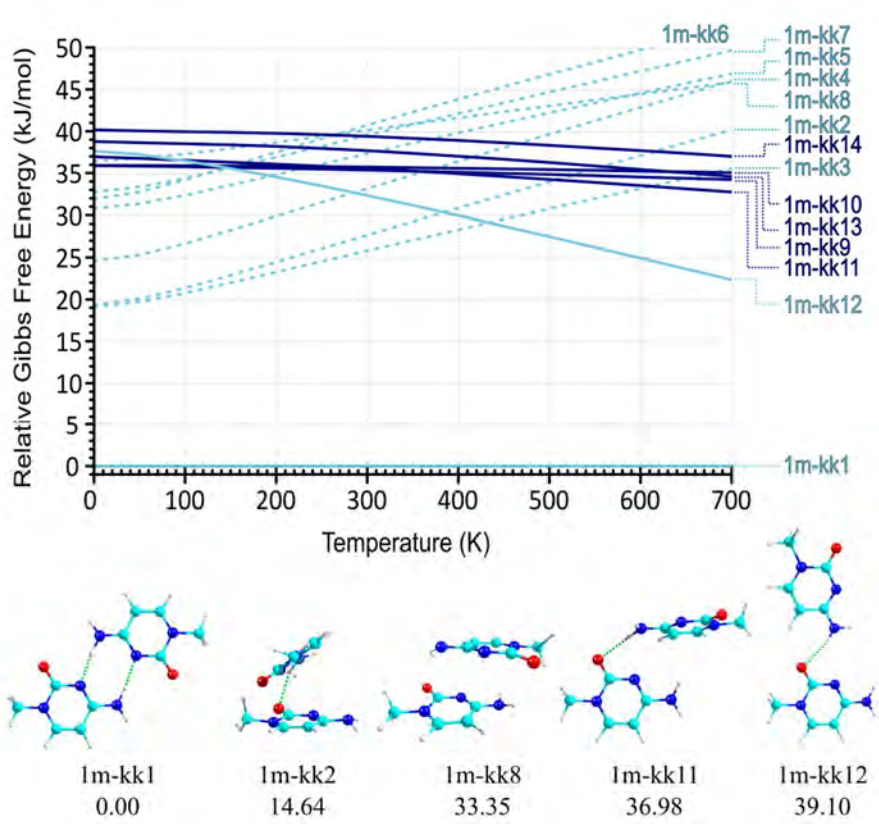


Figure 6.21 *Relative Gibbs free energy of the dimers of 1-methylcytosine in the 0-700 K interval and the structure of the most relevant tautomers with their relative energy in kJ mol^{-1}. Continous line: planar structure; Dot line: stacking structure. Dark blue: non-planar structures.*

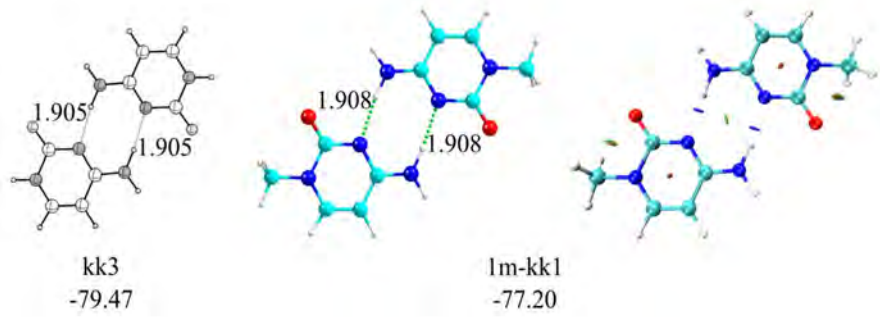


Figure 6.22 *The structure of the kk3 cytosine (left) and 1m-kk1, together with distances (Å) of the hydrogen bonds and the binding energy values in kJ mol^{-1}. The most relevant van der Waals interactions for the 1m-kk1 dimer are depicted in the NCI plot (right).*

Table 6.4. *Electron density (a.u.) of the hydrogen bonds observed for the kk1, kk2 and kk3 dimer of cytosines and 1m-kk1 dimer of 1-methylcytosine.*

	Electron Density (a.u)			
	Cytosine			1-methylcytosine
	kk1	kk2	kk3	1m-kk1
HN-H^{I}···O=C2^{II}	0.0334			
N3^{I}···H-N1^{II}	0.0417			
C2=O^{I}···H-N1^{II}		0.0422		
N1-H^{I}···O=C2^{II}		0.0422		
HN-H^{I}···N3^{II}			0.0319	0.0317
N3^{I}···H-NH^{II}			0.0319	0.0317

6.2.3 1-methylcytosine trimer

Inclusion of a third molecule resulted in a modest increase in the number of conformers, from 23 to 30, but still, the global minimum is separated from the rest of the structures from an energetic point of view, as can be seen in **Figure 6.23**. Furthermore, it is also favored by entropy and therefore it becomes relatively more stable with temperature.

The stability difference between stacking and planar structures was also reduced in the trimer. In fact, 1m-kkk1 is a stacking structure and the first planar structure, 1m-kkk4, is separated by 6.21 kJ mol^{-1} from the global minimum. On the other hand, when the temperature and the BSSE correction are taken into account, the planar structure becomes the most favored ones, and the first stacking structure appears at 3 kJ mol^{-1}. Thus, the same competition observed in the case of the tetramer of cytosine is also observed for the trimer of 1-methylcytosine. Apparently, methylation of N1 position prompts the formation of stacking structures.

Figure 6.24 collects the ball&sticks model of four selected trimers of 1-methylcytosine, together with the relative energy, binding energy values (bold), hydrogen bond distances and the representation of the non-covalent interactions. The values of the electron density are collected in **Table 6.5**

A close look at the structures in **Figure 6.25** shows that all of them contain an 1m-kk1 kernel, to which a third 1m-cytosine attached in

different positions. Therefore, the energy difference between the four isomers is due to the interaction of the third molecule.

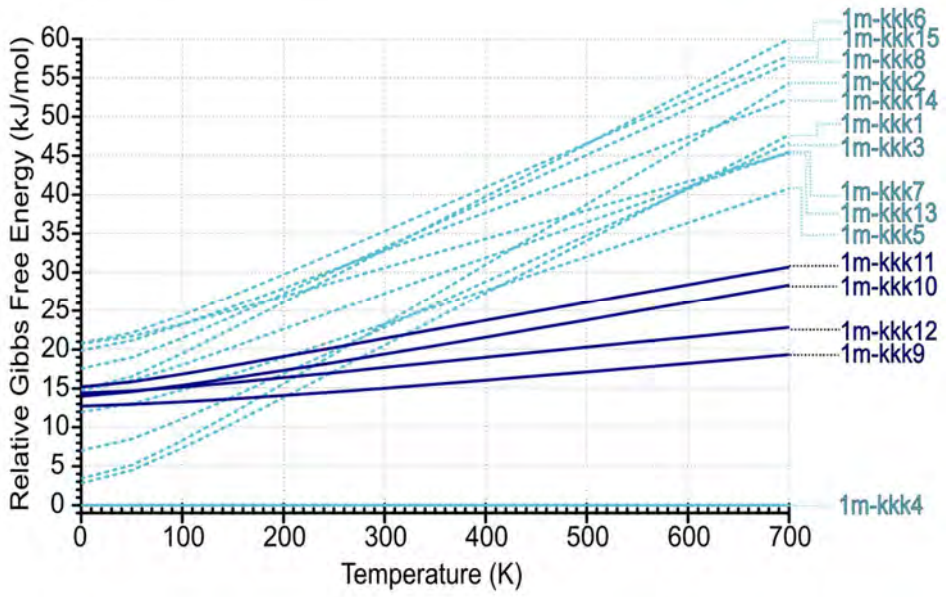


Figure 6.23 *Relative Gibbs free energy of the trimers of 1-methylcytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure. Dark blue: non-planar structures.*

Table 6.5. *Electron density (a.u.) of the hydrogen bonds observed for the four possible trimers of 1-methylcytosine. Dash line signs the first kk1 dimer structure.*

	Electron Density (a.u)			
	1m-kkk1	**1m-kkk2**	**1m-kkk3**	**1m-kkk4**
HN-H^{I}···N3^{II}	0.0343	0.0372	0.0343	0.0404
N3^{I}···H-N1^{II}	0.0212	0.0247	0.0284	0.0245
N3^{I}···H-NH^{III}	0.0149			
C2=O^{I}···H-NH^{III}		0.0206		0.0252
HN-H^{II}···N3^{III}				0.0234

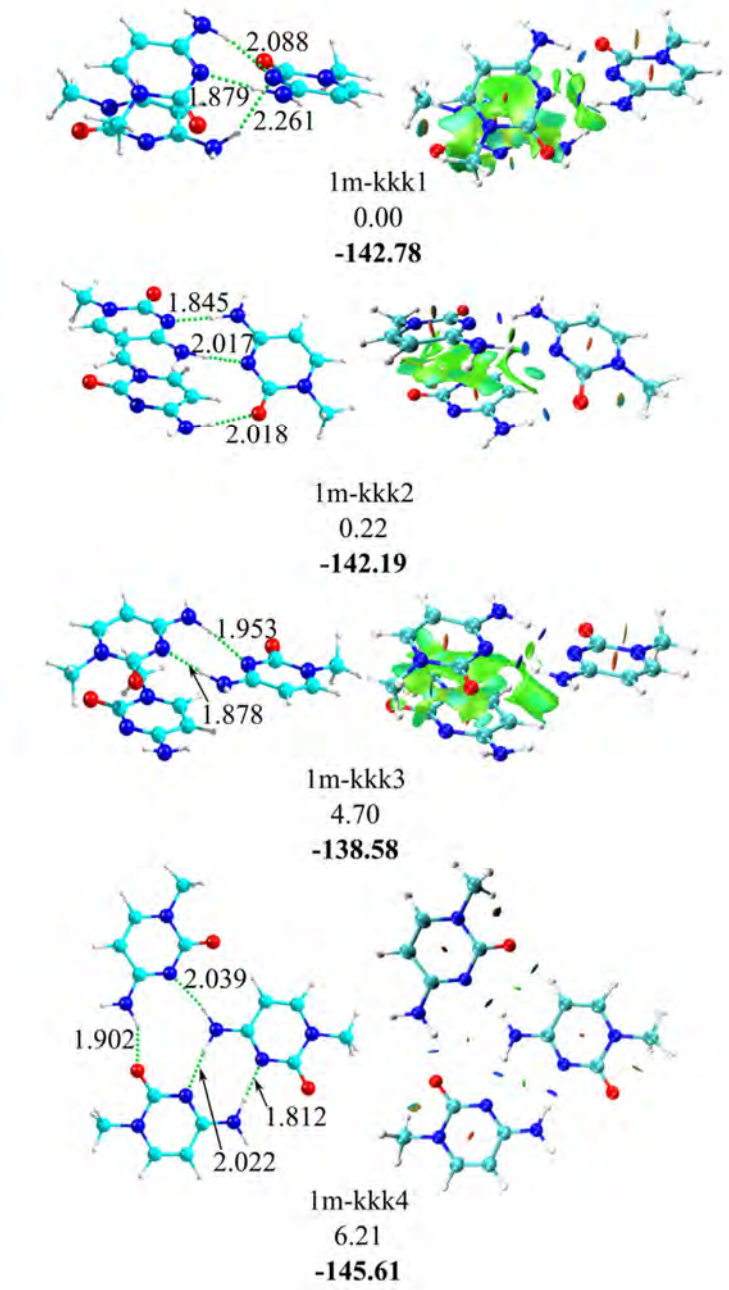


Figure 6.24 *The four possible trimers of 1-methylcytosine with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions depicted in NCI plots. Relative energy and binding energy values (bold) in kJ mol^{-1} are also shown.*

6.2.4 1-methylcytosine tetramer

Addition of the fourth 1-methylcytosine molecule changed the conformational landscape drastically. First, an important increase in the number of structures was observed and therefore, it was necessary to take into account 75 conformers during the full optimization stage. Second, as highlighted in **Figure 6.25**, at least four structures are almost isoenergetic and their relative stability changes with the temperature.

With addition of the fourth molecule, the most stable conformer, 1m-kkkk1, presents a stacking structure. The most stable planar structure, 1m-kkkk38, appears at 14.6 kJ/mol, compared to 26.97 kJ mol^{-1} at 0 K. The prevalence of this structure is maintained even after inclusion of the BSSE correction, up to 200 K. Further increase in temperature results in additional stabilization of the planar structures, which become the most stable ones above 300 K.

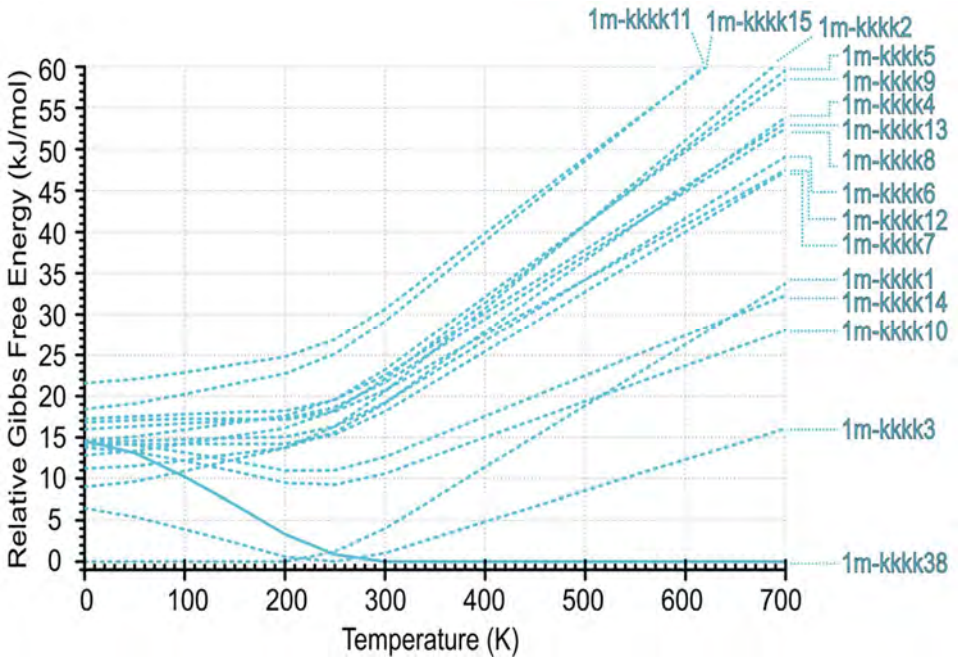


Figure 6.25 *Relative Gibbs free energy of the tetramers of 1-methylcytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

The structure of the three most stable tetramers of 1-methylcytosine at 200 K, together with their relative energy and the binding energy values, the distance of the hydrogen bonds and the representation of the NCI plots were collected in **Figure 6.26** and the values of the electron density may be found in **Table 6.6**. The first two structures present a stacking disposition (1m-kkkk1 and 1m-kkkk3), whereas 1m-kkkk38 exhibits a planar configuration, as explained above.

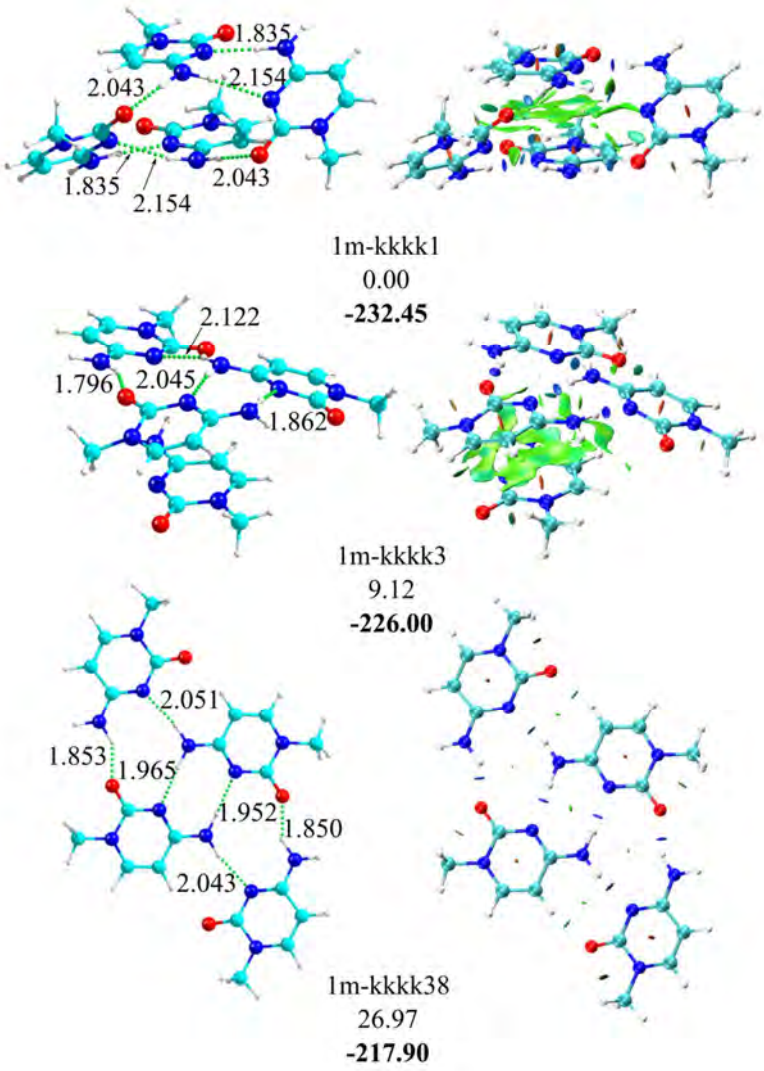


Figure 6.26 *The three more probable tetramers of 1-methylcytosine, together with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions depicted in NCI plots. Relatives energies and binding energy values (bold) in kJ mol^{-1} are also shown.*

A close inspection of the ball&stick representations in **Figure 6.26** shows that all three structures contain an 1m-kk1 kernel. More precisely, 1m-kkkk1 is composed by a pair of 1m-kk1 dimers; two of the rings are stablished by π···π interactions, while at the same time, several polar groups form intermolecular hydrogen bonds. Meanwhile, 1m-kkkk3 and 1m-kkkk38 contain an 1m-kkk4 trimer with a fourth 1-methylcytosine staked to one of the 1m-cytosines in the former and forming a complex hydrogen bond network in the latter.

Table 6.6. *Electron density (a.u.) of the hydrogen bonds observed for the three possible tetramers of 1-methylcytosine. Green dash line signs the first kk1 dimer structure and red dash line signs the structure kkk1 trimer, both observed in the structure of the tetramer.*

	Electron Density (a.u)		
	1m-kkkk1	**1m-kkkk3**	**1m-kkkk38**
$HN\text{-}H^{I}\cdots N3^{II}$	0.0381	0.0317	0.0228
$N3^{I}\cdots H\text{-}N1^{II}$	0.0195	0.0234	0.0226
$HN\text{-}H^{III}\cdots N3^{IV}$	0.0381		
$N3^{III}\cdots H\text{-}N1^{IV}$	0.0195		
$C2{=}O^{I}\cdots H\text{-}NH^{III}$		0.0357	0.0284
$HN\text{-}H^{II}\cdots N3^{III}$		0.0197	0.0145
$HN\text{-}H^{II}\cdots C2{=}O^{IV}$	0.0184		
$C2{=}O^{I}\cdots H\text{-}NH^{III}$	0.0184		
$HN\text{-}H^{I}\cdots N3^{IV}$			0.0147
$C2{=}O^{II}\cdots H\text{-}NH^{IV}$			0.0287

6.3 Guanine

As in the case of the cytosine, guanine has been studied both theoretically and experimentally in depth. Most of the experimental studies were centered in isolated guanine and its dimers.[30,32,33,41,58–63] Some pure computational studies were also carried out, focusing in its tautomeric equilibrium and in the methylated forms.[64–70]

In this section, the aggregation preferences of guanine in supersonic expansions will be studied up to the trimer, using the same methodology as in cytosine. The structure and the atom labelling of guanine is presented in **Figure 6.27.**

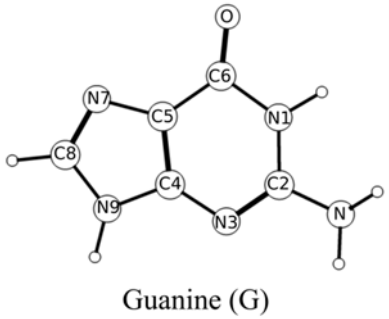


Figure 6.27 *Structure and atom labeling of guanine.*

6.3.1 Spectroscopy of guanine aggregates

Guanine is a bicyclic molecule formed by fused pyrimidine-imidazole rings. It has a number of electronegative moieties, such as an NH_2 at position 2, a C=O at position 6 and two NH groups at position 1 and 9. This combination results in an increased number of tautomers compared with cytosine. FT-MW studies by Alonso et al.[62] determined that at least four different tautomers of guanine, related to the keto/enol tautomerism and to the migration of the hydrogen from position 9 to 7, are present in the expansion. Seefeld et al.[41] carried out IR/UV experiments, finding also imino tautomers. Thus, the conformational landscape becomes very complicate. All the different tautomers may be found in **Figure 6.28**. Enol and imino forms present in addition several rotamers, originated in the rotation of the hydroxyl and imino groups (not shown in the figure).

In contrast with cytosine, the REMPI spectra of guanine tautomers do not present a large gap between keto an enol forms. Several studies using REMPI spectroscopy have been reported, most of them carried out by de Vries group [30,32,33] and Mons group.[58–61] All of them reported the absence of the three most stable tautomers (9-keto, 7-keto and one of the rotamers of the 9-enol) in the REMPI spectrum due to sub-picosecond excited state lifetimes. Four different tautomers were found in the REMPI spectra, as demonstrated by UV/UV hole burning. Two of them were assigned by Mons et al.[58] to 9-enol and 7-enol species, by comparison with

methylated derivatives. The other two monomers are related to imino tautomers, as established by Seefeld et al.[41] by the measurement of the C=N stretching frequency at 1700 cm^{-1}. The REMPI spectrum collected in this work in the 32800 and 34100 cm^{-1} region may be found in **Figure 6.29**, black trace. The 0_0^0 transition appears at 32874 cm^{-1}, in good agreement with the value of 32878 cm^{-1} reported by Nir et al.[33]

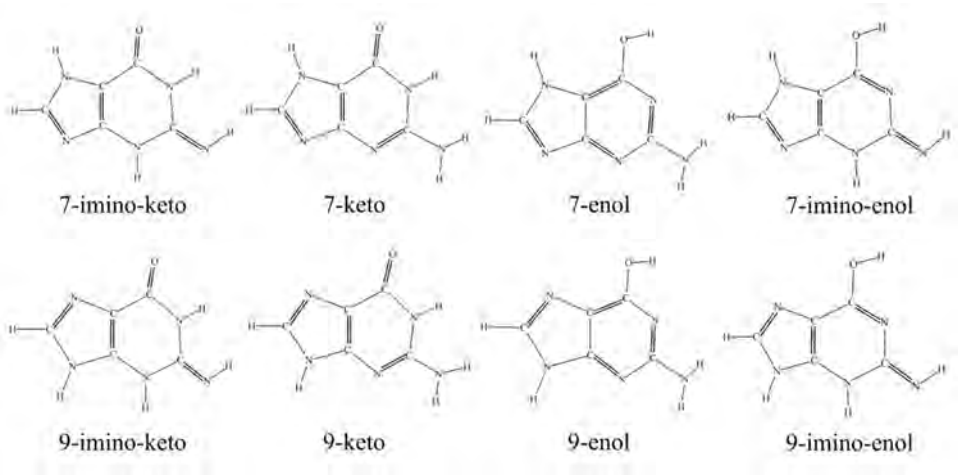


Figure 6.28 *Structure of the eight possible tautomers of guanine.*

REMPI spectrum of guanine dimer was also reported by Nir et al.,[63] who determined the position of the 0_0^0 transition at 33103 cm^{-1}, a value very similar to the value of 33108 cm^{-1} obtained in this work (blue trace in **Figure 6.29**).

The REMPI spectra of guanine monomer and dimer were also recorded during the experiments carried out at FELIX and the spectra are also shown at **Figure 6.29**. In general, there is a good agreement between both spectra. The bands appear at the same positions, although with different relative intensity, probably due to a slight laser saturation of the spectrum recorded at the UPV/EHU. An additional source of differences may be the temperature of the beam, which is significantly colder at FELIX, due to the use of a different valve (see Section 2.1.3).

Despite the reduced signal intensity, it was also possible to record the electronic spectrum of trimer in the 32300-33800 cm^{-1} region, **Figure 6.29**. This is the first report of the REMPI spectrum of guanine trimer. It is a broad absorption that starts slightly to the red of guanine 0_0^0 transition. This sequence of blue shift in the dimer and back to the red in the trimer is

typical of formation of cyclic structures, as observed before in water complexes.[71–75] As in the cytosine trimer and tetramer, a likely explanation for the broad absorption may be related to the dynamics of the excited state or to the large number of low-frequency vibrations.

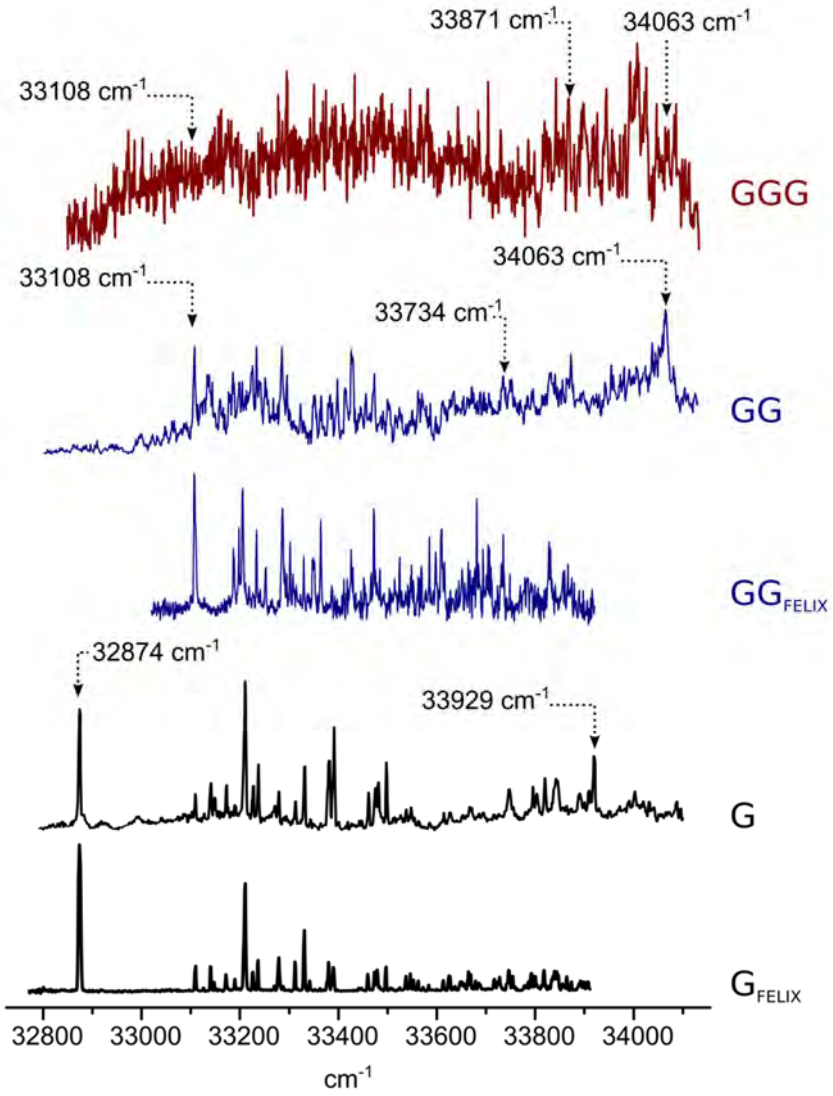


Figure 6.29 *REMPI spectra of guanine monomer (G), depicted in black, guanine dimer (GG) depicted in blue and guanine trimer (GGG) depicted in red, scanned between 32800-34100 cm^{-1}. The arrows indicate the wavenumbers tested to record IR/UV spectra and the FELIX tags indicates the spectra collected during the experiments in FELIX laboratory.*

Despite the low S/N ratio, it was possible to obtain the IR spectra of the species up to the trimer. The wavelength selected for the application of the IR/UV spectroscopy is indicated with arrows in **Figure 6.29**, while the four different IR spectra obtained tuning the UV at those wavelengths may be found in **Figure 6.30**. *A priori* we cannot rule out the presence of several isomers contributing to each of the IR spectra, especially because UV/UV hole burning spectroscopic technique was not applied.

The black traces in **Figure 6.30** correspond to two different tautomers of guanine monomer, and were obtained tuning the UV laser at 32874 cm^{-1} (a) and 33929 cm^{-1} (b). These two spectra will be important for the calculation of the correction factor to account for the anharmonicity and that will be maintained during the study of the aggregates.

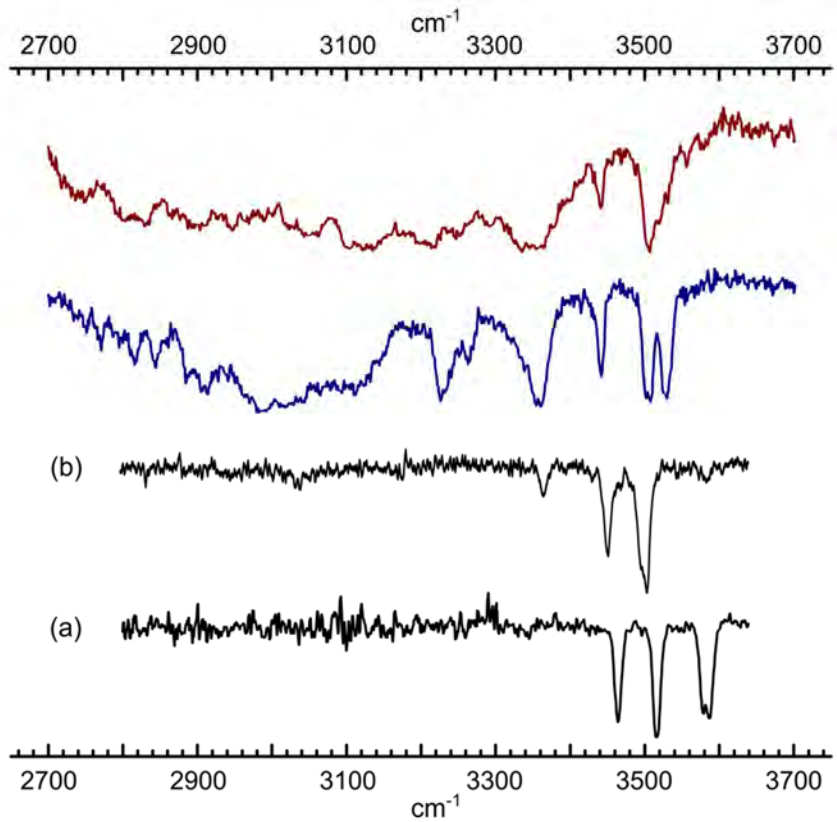


Figure 6.30 *IR spectra of guanine monomer (black), fixing the UV wavelength at 32874 cm^{-1} (a) and 33929 cm^{-1} (b); guanine dimer (blue) and guanine trimer (red), scanned at 2700-3750 cm^{-1}.*

The blue spectrum on **Figure 6.30** corresponds to guanine dimer, and it is similar to the one obtained by Nir et al. in the 3350-3700 cm^{-1} region.[63] The IR spectrum presented in this study extends to 2700 cm^{-1}. Three different wavelengths were probed with the UV laser, looking for evidences of additional conformers, but similar IR spectra to those in **Figure 6.30** were always obtained (see **Figure 8.3.63** in **Appendix 8.3**).

To the best of our knowledge, this is the first time the mass-resolved IR spectrum of the trimer is reported (red trace in the figure). Its shape is relatively similar to that of the dimer, although some of the well-resolved bands disappeared and the broad absorption extended to the blue. As in the case of the dimer, several UV wavelengths were also tested, **Figure 8.3.81** in **Appendix 8.3**, but a similar spectrum was always obtained. In the next sections, the theoretical study of each aggregate will be presented and a possible assignment of the experimental spectra will be discussed.

6.3.2 Guanine monomer

Figure 6.31 collects the variation of the relative Gibbs free energy of guanine tautomers with temperature in the 0-700 K interval. No large variations in $\Delta(\Delta G)$ were observed, which means that none of the tautomers is heavily favored by entropy. Based on the energetic order depicted in **Figure 6.31**, one would expect to find four tautomers in the expansion: 9-keto, 7-keto, 9-enol and 7-enol. The rest of the tautomers are in principle to high-enough in energy to expect them not to present a significant population in the expansion. However, Seefeld et al.[41] probed that 7-iminoketo species are present in a similar experiment, finding 7-enol tautomer and two isomers of 7-iminoketo. Probably high energy barriers isolate such tautomers from the most stable species, allowing part of the population to be trapped. It is also worthy to note that the most stable species, 9-keto, 7-keto and 9-enol1 have not been previously detected in REMPI experiments.[61,64,76] The authors of such studies argued that short, sub-picosecond, excited state lifetimes could be the cause for the lack of signal. This is a very likely explanation, specially taking into account that those tautomers were detected using FT-MW instead of REMPI. Thus, despite they are high in energy, in the following all the tautomers up to 7-iminoketo will be considered as possible participants in the aggregation process in the rest of the study.

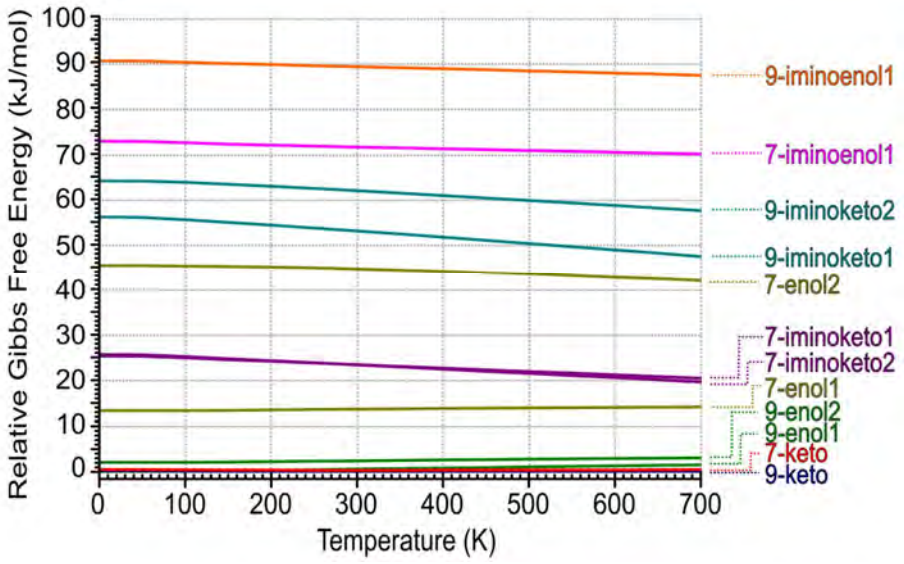


Figure 6.31 *Relative Gibbs free energy of the possible tautomers of guanine in the 0-700 K interval.*

Assignment of the experimental IR may be found in **Figure 6.32**. The experimental spectra were compared to the computed tautomers assigned in previous works. The first experimental trace (a) probably contains contribution from two different enol conformers, differing in the position of the NH group: 9-enol and 7-enol. The theoretical calculations reveal that the peak at 3587 cm^{-1} is due to the stretching of the hydroxyl group. As in the case of enol-cytosine, the calculation does not predict very accurately the position of this band, probably due to effects that were not included in the Hamiltonian. However, a deep study of such effects is away from the scope of the present work.

The peaks at 3577 and 3463 cm^{-1} are due to the symmetric and antisymmetric stretches of the NH_2 group, labelled with red dots in the figure, and the band at 3578 cm^{-1} is due to the stretching of the NH group in position 9 or 7, depending on which conformer is selected. The position of those bands suits very well the position of the experimental peaks. Mons et al.[58] assigned this trace to the 7-enol1 conformer by comparison with the peaks observed in 7-methylguanine and assigned as the 9-enol2 conformer an additional band observed at 34755 cm^{-1}.

Seefeld et al.[41] assigned trace (b) to 7-iminoketo1 and 2 tautomers. As can be seen, our predicted spectra for those species also fit well the experimental trace, confirming such assignment. Both species are almost

isoenergetic, separated only by 0.34 kJ mol^{-1}. Furthermore, the trace presents some shoulders, suggesting the presence of several species. The calculations predict the presence of bands that match very well the position of such features. Thus, we cannot exclude the existence of both species in the expansion. The two peaks at 3502 and 3495 cm^{-1} may be then assigned to the stretching of the N3H and N7H groups and the third peak at 3451 cm^{-1} corresponds to the stretch of N5H. Finally, the stretching of the imino moiety appears at 3365 cm^{-1}.

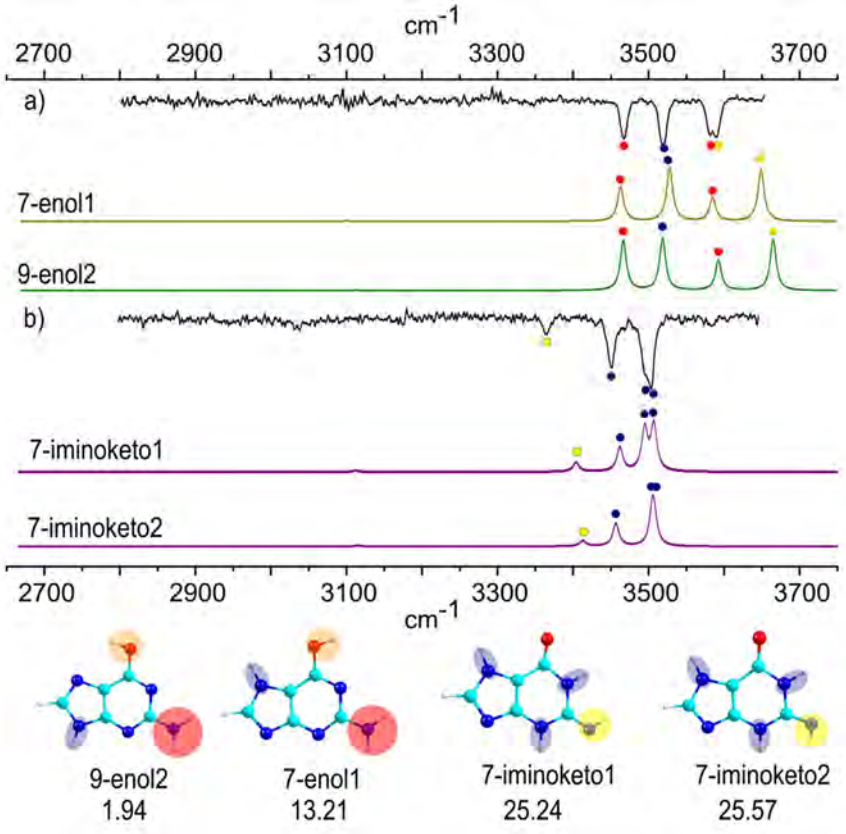


Figure 6.32 *Comparison between the experimental IR spectrum at 32874 cm^{-1} (a) and 33929 cm^{-1} (b) and the theoretical prediction of the four most probable structures of the guanine, with their structure and relative energy in kJ mol^{-1}.*

It is worthy to mention that, as in the case of cytosine, the spectrum of the monomer will be used to calculate the scaling factor to account for the anharmonicity, which was estimated to be 0.9520, and it will be used to correct also the calculated spectra of guanine aggregates.

6.3.3 Guanine dimer

The possible existence of several guanine tautomers in the expansion: 9-/7-keto, 9-/7-enol and 9-/7-iminoketo, largely complicates the calculations as all possible combinations have to be taken into account. Altogether, 15 possible combinations exist. **Figure 6.33** collects the study of the Gibbs free energy in the 0-700 K interval for the most relevant aggregates of each combination of tautomers, adopting either planar or stacking conformations. Nevertheless, the preference for planar structures in the whole range of temperatures is clear. The first stacking structure (9k9k2) appears at ~18 kJ mol^{-1} and the rest of stacking geometries are above 30 kJ mol^{-1}. Therefore, observation of such structures would be very unlikely. Nevertheless, the stacked structures are relatively more stable than in the case of cytosine.

It is interesting to note how the difference in stability between the global minimum, 9k9k homodimer, and the next most stable structure, 9k7k1, is ~ 6 kJ mol^{-1} which is a significant difference. Furthermore, the four most stable structures contain a 9k tautomer. Apparently, formation of guanine dimers seems to favor those aggregates containing a 9k tautomer. Interestingly, despite the stability of the enol tautomer, formation of enol dimers is significantly less favored and the first enol homodimer appears around 23 kJ mol^{-1}, in contrast with the most stable enol heterodimer, 9k7e1, which is at 14 kJ mol^{-1} above the global minimum.

In order not to lose any relevant conformers, we selected a wide energy window of 20 kJ mol^{-1}: all the families of structure within this window at 100-200 K were considered: 9k9k, 9k7k, 9k7ik, 9k7e, 7k7k, 7k7ik, 7e7ik and 7k7e. The remaining families are higher-enough in energy to be excluded from the rest of the study.

The analysis of the Gibbs free energy of these families in the selected range of temperatures may be found in **Figure 6.34**. Still, 9k9k1 is the preferred conformer in the whole interval, but several additional conformers are stable enough to be present in the expansion. For example, if an energy window of 15 kJ mol^{-1} is selected, 9k9k1, 9k7k1, 9k7k2, 9k7ik1, 9k7ik2, 7k7k1 and 9k7e1 lie in that window. The predicted spectra for such structures were compared with the experimental IR spectrum recorded in the beam (**Figure 6.35**).

This dimer was studied by Nir et al.[63] by UV/UV hole burning and IR/UV double resonance combined with *ab-initio* calculations, although at

a significantly lower computational level. They found two isomers in the expansion. Comparison of the results obtained for guanine dimer with those from 1- and 9-methylguanine dimers, led them to conclude that the two species detected corresponded to 9k7k1 and 9k7k2. Absence of 9k9k1, the most stable dimer, was attributed to a possible strong exciton splitting and hydrogen bond that induced a shift in the allowed $S_0 \leftarrow S_2$ transition, moving it away from their spectral window.[32,63]

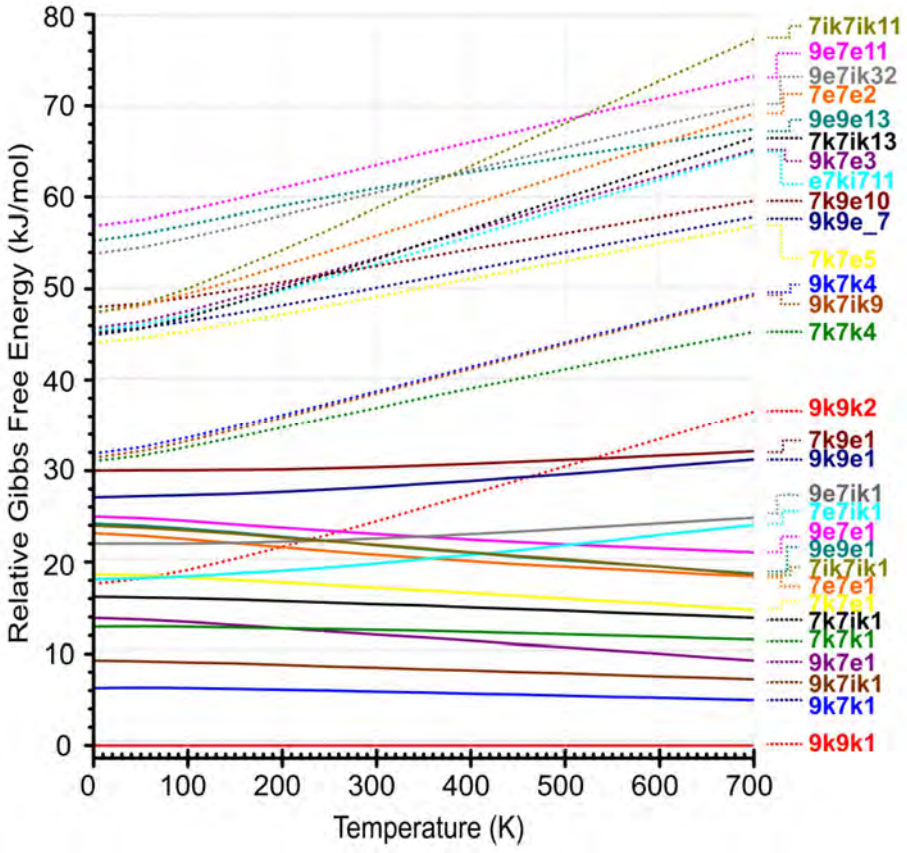


Figure 6.33 *Relative Gibbs free energy of the most stable representant of the guanine dimers in the 0-700 K interval. Continous line: the first planar dimer; Dot lines: the first stacking structure.*

The high level calculations conducted here seem to confirm the assignment proposed by Nir et al.. However, some bands in the spectrum below 3300 cm^{-1}, which were outside their observation window and therefore were not reported in their work, point to the existence of additional isomers contributing to the spectrum (**Figure 6.35).** Certainly,

the bands at 3263 and 3227 cm^{-1} are not present in the simulated spectra of the assigned conformer, 9k7k1 and the most stable isomers whose spectra are able to reproduce those bands contain iminoketo tautomers: 9k7ik1 and 9k7ik2. The former is less than 10 kJ/mol above the global minimum and the latter is at ~12 kJ/mol and therefore their presence may not be *a priori* discarded, especially taking into account that the corresponding monomers were detected in the expansion. However, their spectra do not fit well with the experimental bands around 3550 cm^{-1}.

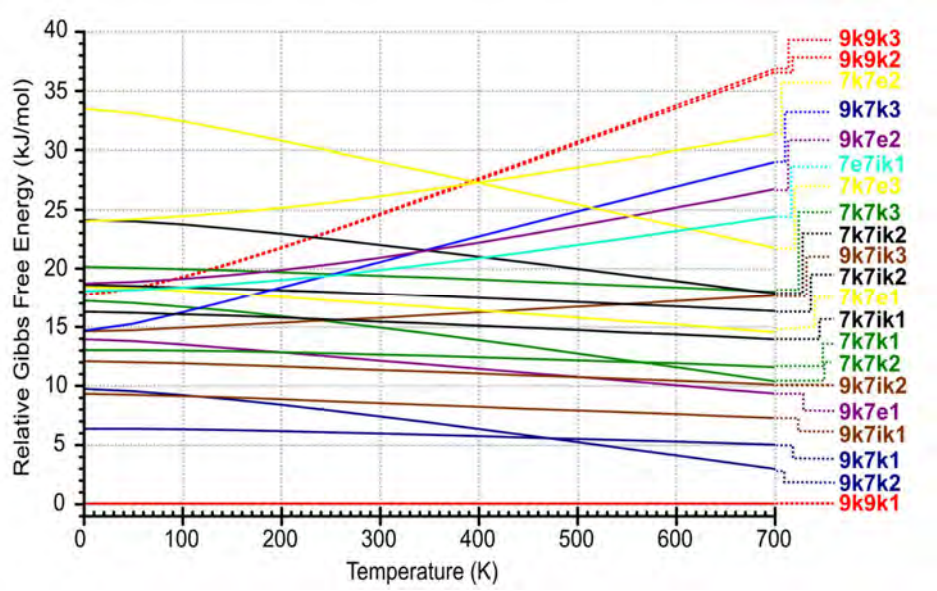


Figure 6.34 *Relative Gibbs free energy of the most favored dimers of guanine in the 0-700 K interval.*

Looking for additional confirmation of the proposed assignments, the IR spectrum of the detected isomer was also recorded in the 300-1800 cm^{-1} region. The spectrum obtained may be found in **Figure 6.36**. As it can be seen, the presence of 9k7k1 dimer is clearly confirmed due to the band at 1740 cm^{-1} that can only be explained by this conformer. Regarding the rest of conformers, only 7k7k1 could be safely discarded. To solve the question about the number of isomers present, additional experiments would be required, such as IR/IR/UV triple resonance. However, such technique requires of signal intensity difficult to achieve with this system, due to the fluctuations in signal intensity introduced by the laser desorption system.

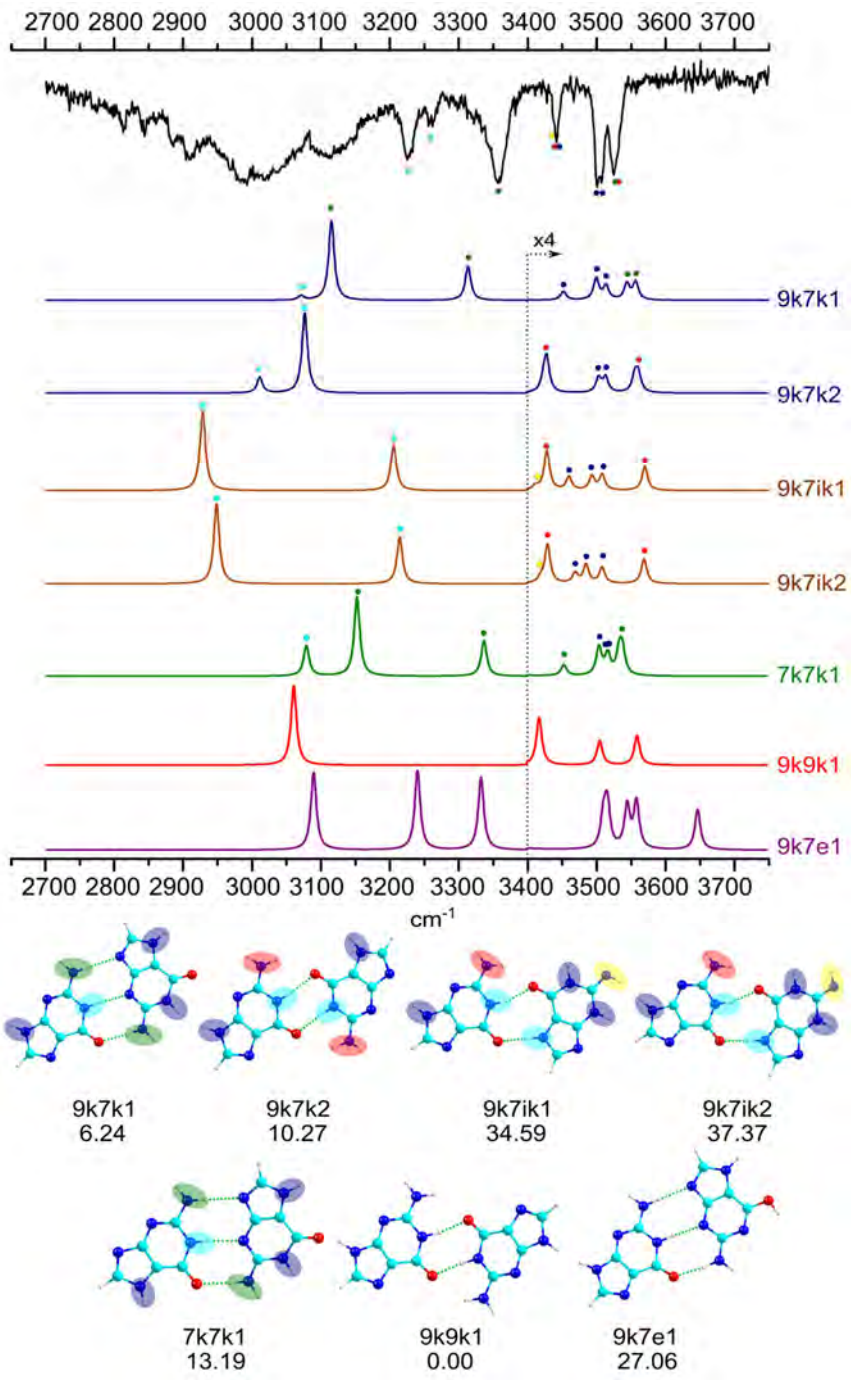


Figure 6.35 *Comparison between the mass-resolved IR/UV spectrum of guanine dimer and the simulated spectra of the seven more probable species that may be contributing to the spectrum, together with the structure of these species and their relative energies in kJ mol^{-1}.*

Finally, ball&stick models of the six conformers with the binding energy values and the most relevant interactions depicted in the NCI plots may be found in **Figure 6.37** and the values of the electron density of the hydrogen bond in **Table 6.7**.

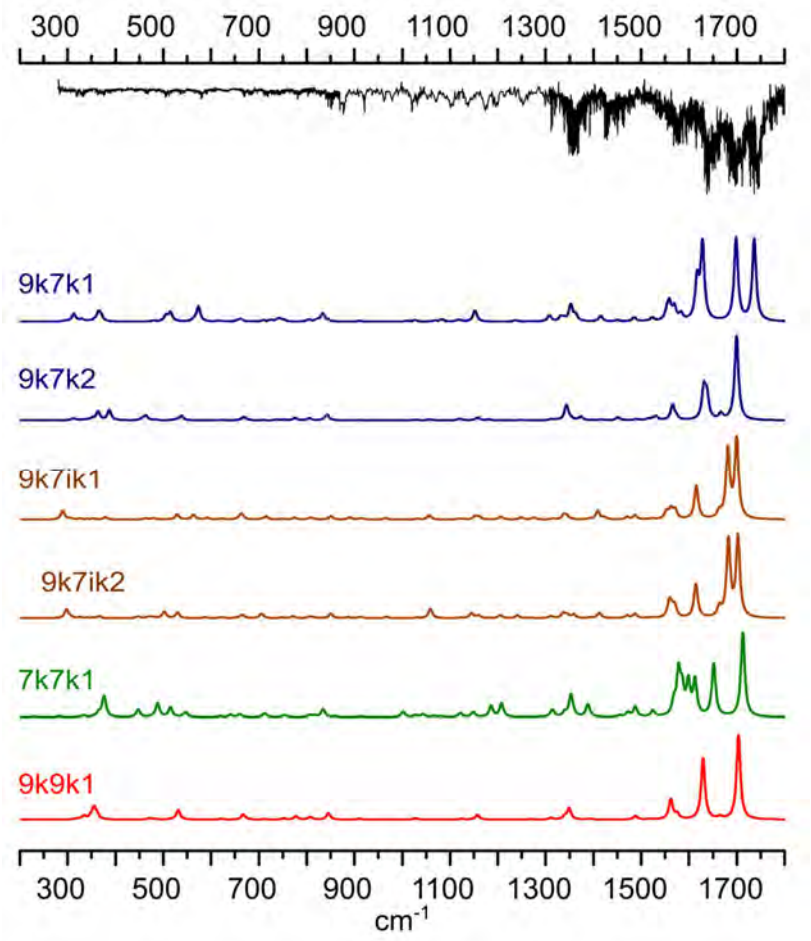


Figure 6.36 *Comparison between the mass-resolved IR/UV spectrum of guanine dimer in 300-1800 cm^{-1} IR region and the simulated spectra of the six species that may be contributing to the spectrum.*

Three types of interactions may be identified in the dimer: formation of two symmetric hydrogen bonds between N_1-$H^{I}\cdots O$=$C_6{}^{II}$ and C_6=$O^{I}\cdots$H-$N_1{}^{II}$ in 9k9k1 and 9k7k2 (type I); formation of three hydrogen bonds between N_1-$H^{I}\cdots N_3{}^{II}$, C_6=$O^{I}\cdots$H-$N_7{}^{II}$ and H-$NH^{I}\cdots N_9$ in 9k7k1 and 7k7k1 conformers (type II), and formation of N_1-$H^{I}\cdots O$=$C_6{}^{II}$ and

C_6=O^I···H-N_7^{II} hydrogen bonds in 9k7ik1 and 9k7ik2 (type III). As usual, several other weaker van der Waals interactions are also present in all the conformers, as highlighted by the NCI plots.

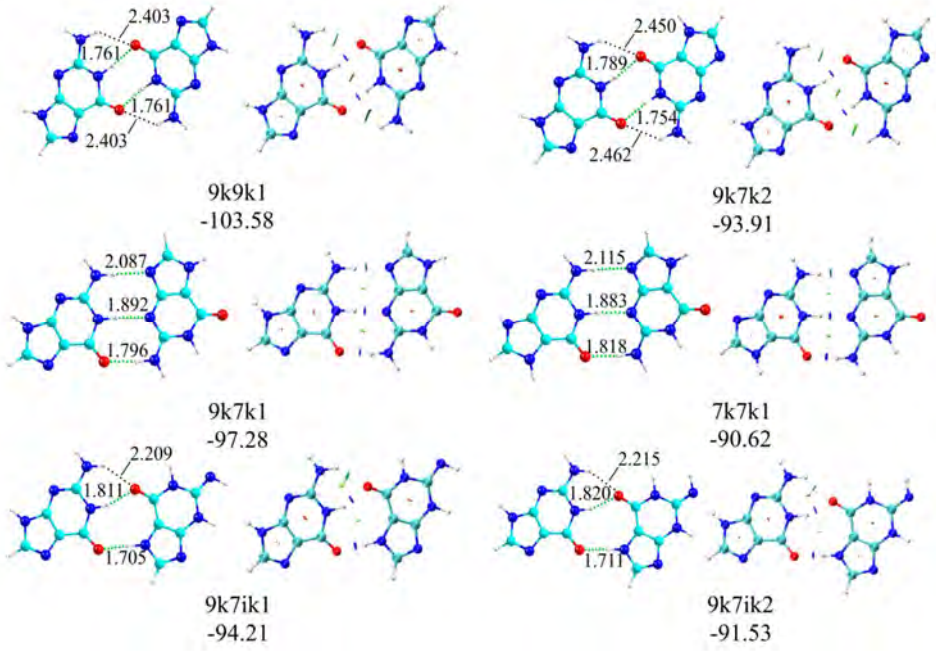


Figure 6.37 *The structure of the six more probable guanine dimers with distances (Å) of the hydrogen bonds and most relevant interactions depicted in NCI plots. Binding energy values in kJ mol^{-1} are also shown.*

Table 6.7. *Electron density (a.u.) of the hydrogen bonds observed for the six possible dimers of guanine.*

	Electron Density (a.u.)					
	I		II		III	
	9k9k1	9k7k2	9k7k1	7k7k1	9k7ik1	9k7ik2
N1-H^I···O=C6^{II}	0.0398	0.0371			0.0337	0.0330
C6=O^I···H-N1^{II}	0.0398	0.0403				
HN-H^I···N9^{II}			0.0205	0.0193		
N1-H^I···N3^{II}			0.0331	0.0338		
C6=O^I···H-NH^{II}			0.0336	0.0320		
C6=O^I···H-N7^{II}					0.0430	0.0423

6.3.4 Guanine trimer

Addition of a third molecule of guanine produces a very rich conformational landscape. From a combinatorial point of view, there are 34 possible combinations for guanine trimer. In order to make a rational computational study and taking into account the observations for guanine dimer, we selected the most favorable dimers (9k9k, 9k7k, 9k7ik and 7k7k) and considered all their possible combinations with 9k, 7k and 7ik, the monomers found in the beam as candidates for the most stable trimers. Therefore, the number of possible combinations was reduced to 8, each of which presents its own collection of conformers. The variation of the relative Gibbs free energy with the temperature of the most relevant aggregates of each combination of tautomers, adopting planar or stacking structures, may be found in **Figure 6.39**.

Surprisingly, the most stable structure, the planar 9k9k9k1conformer, is separated by ~24 kJ mol^{-1} from the next family of isomers. Furthermore, planar structures are favored by entropy, and therefore stacking structures are not expected to be found in the expansion. The data in **Figure 6.39** seem to indicate that only the most stable isomer would present enough population in the beam to be detected. However, as complexes formed by other tautomers were also detected and taking into account the large barriers for some tautomerizations, for example, imino-keto tautomerization, additional families were considered for the assignment. Thus, 9k9k9k, 9k9k7k, 9k7k7k and 9k9k7ik were subjected to full analysis and the results may be found in **Figure 6.40**. However, no new conformers closer in energy to the global minimum were found in this in-depth analysis. Nevertheless, one must keep in mind that 9k9k dimer was not observed because its $S_1 \leftarrow S_0$ transition has almost no transition dipole moments and the $S_2 \leftarrow S_0$ lies away from the observation window. Taking into account the previously considered guanine dimers, 9k9k9k1, 9k9k7k1, 9k9k7k5, 9k9k7k6, 9k7k7k6, 9k9k9k2 and 9k9k7ik5 were chosen for the assignment of the experimental IR spectra in **Figures 6.41** and **6.42**.

Assignment of the experimental spectra is not easy. The spectrum in the 2700-3700 cm^{-1} region cannot be reproduced by the computed spectrum of the global minimum, 9k9k9k1, due to the band at 3442 cm^{-1}, and therefore at least one additional species must be present in the expansion. The shape and intensity of the bands between 3450-3550 cm^{-1} of 9k9k9k2 and 9k9k7k5 does not fit well with the experimental observation and these structures will not be further considered. On the

other hand, focusing in the 1800-300 cm^{-1} region, 9k9k7k1, 9k9k7k5 and 9k7k7k6 exhibit an intense band at 1750 cm^{-1} that is not present in the experimental trace, at least with such intensity. Therefore, 9k9k7k6 and 9k9k7ik5 are the only remaining candidates, as their theoretical spectra present a band at 3400 cm^{-1}. Furthermore, 9k9k7k6 does not present a band at 1750 cm^{-1} and 9k9k7ik5 exhibits only a weak band at 1740, what fits better with the experimental trace.

From an energetic point of view, if both conformers were present, the expected high population of the global minimum should result in stronger bands, making difficult the detection of 9k9k7k6 or 9k9k7ik5. However, the transition dipole moment of the global minimum is very small compared with that of 9k9k7k6 or 9k9k7ik5 (0.30, 5.32 and 3.99 D respectively), somehow compensating the difference in population. In this assignment, we have considered that 9k7k and 9k7ik species were detected in the expansion and therefore one cannot discard that they take part in formation of upper aggregates, despite their huge difference in

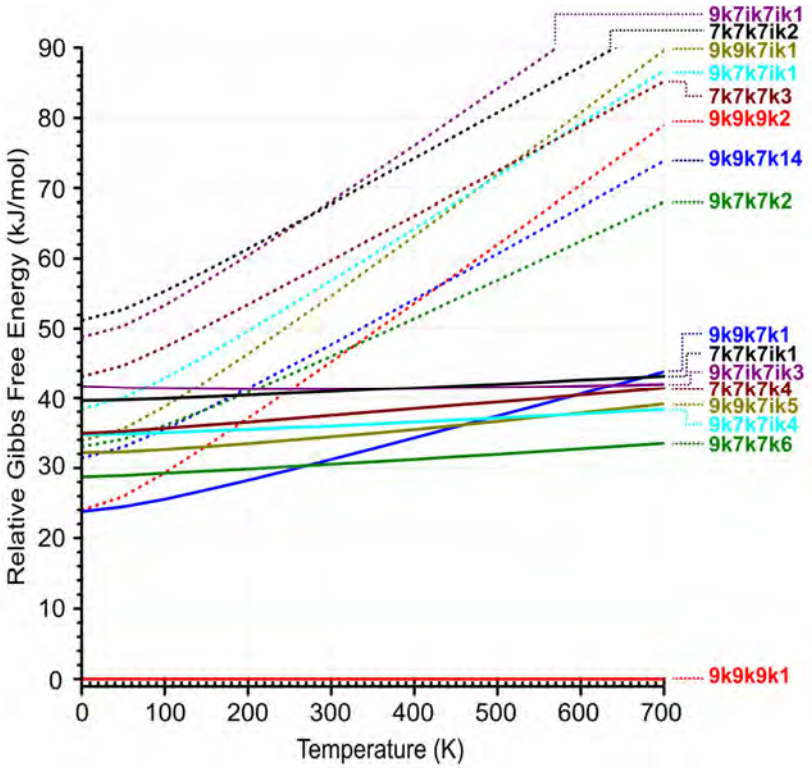


Figure 6.39 *Relative Gibbs free energy of the most favored representant of the guanine trimers in the 0-700 K interval. Continous line: planar strucutures; Dot lines: stacked structures.*

stability. Indeed, the barriers for tautomerization from imino-keto into keto or enol are expected to be significantly high and therefore, the energy required for the process may be lost in the first instants of the cooling process. However, other isomers can suit with the experimental trace and consequently, further experiments are required to determine the validity of the proposed assignment.

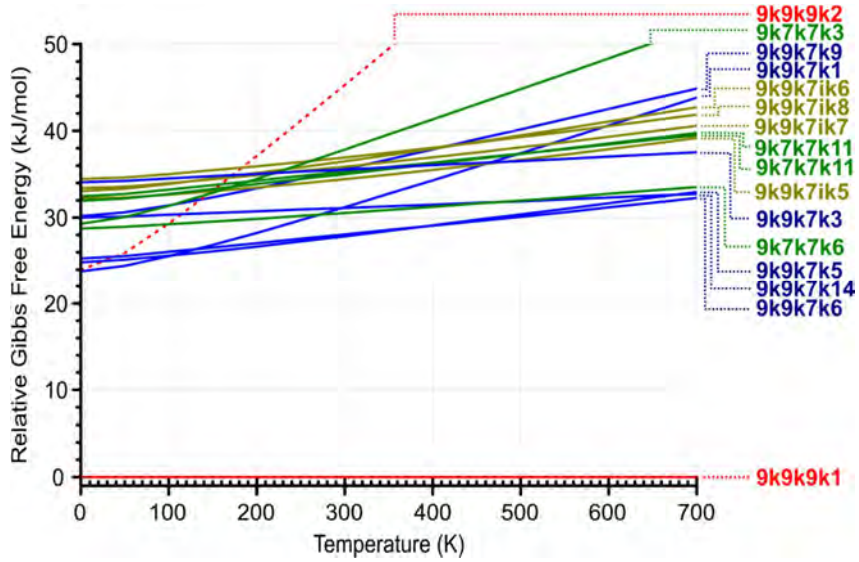


Figure 6.40 *Relative Gibbs free energy of the most favored trimers of guanine in the 0-700 K interval. Continous line: planar strucutures; Dot lines: stacked structures.*

Finally, the structure of the five conformers with the binding energy values and the most relevant interactions depicted in the NCI plots are collected in **Figure 6.43**. The values of the electron density of the hydrogen bond are also presented in **Table 6.8**.

Three different families were observed. The family of the 9k9k9k1 structure (type I) is not only the most stable one, but also the one that presents the highest binding energy, reinforcing the prevalence of this family. The elegant and symmetric 9k9k9k1 structure is a derivative of the 9k9k1, formed by the same N_1-H···O=C_6 and HN-H···O=C_6 interactions between the three molecules, forming a cyclic structure. On the other hand, 9k9k7ik5 maintains the pure structure of 9k9k1, adding the third 7ik guanine by three hydrogen bonds.

The symmetry is broken in the second family (type II), which includes 9k9k7k6 and 9k7k7k6 conformers, built around a 9k7k1 core, with the third guanine interacting in different ways.

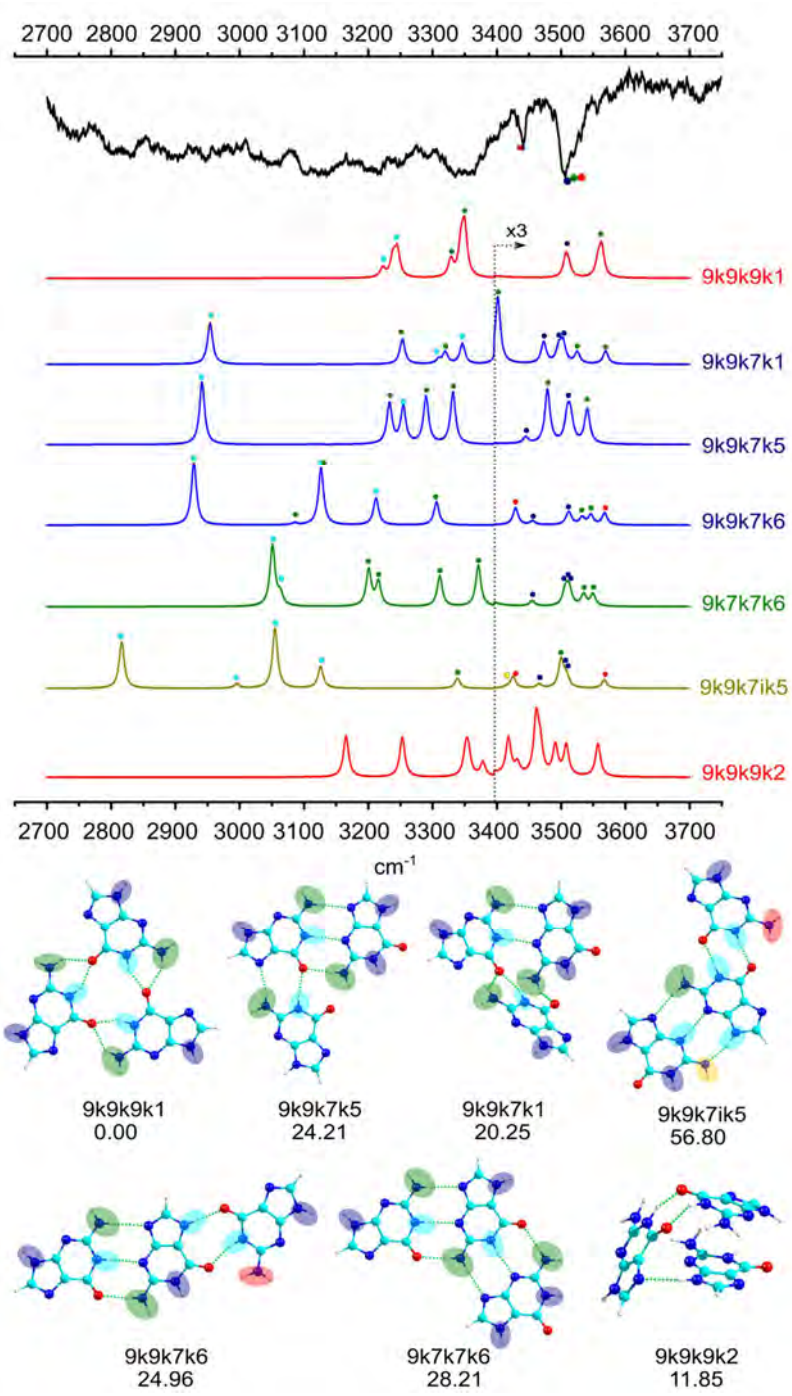


Figure 6.41 *Comparison between the mass-resolved IR/UV spectrum of guanine trimer and the simulated spectra of the seven more probable species that may be contributing to the spectrum, together with the structure of these species and their relative energies in kJ mol^{-1}.*

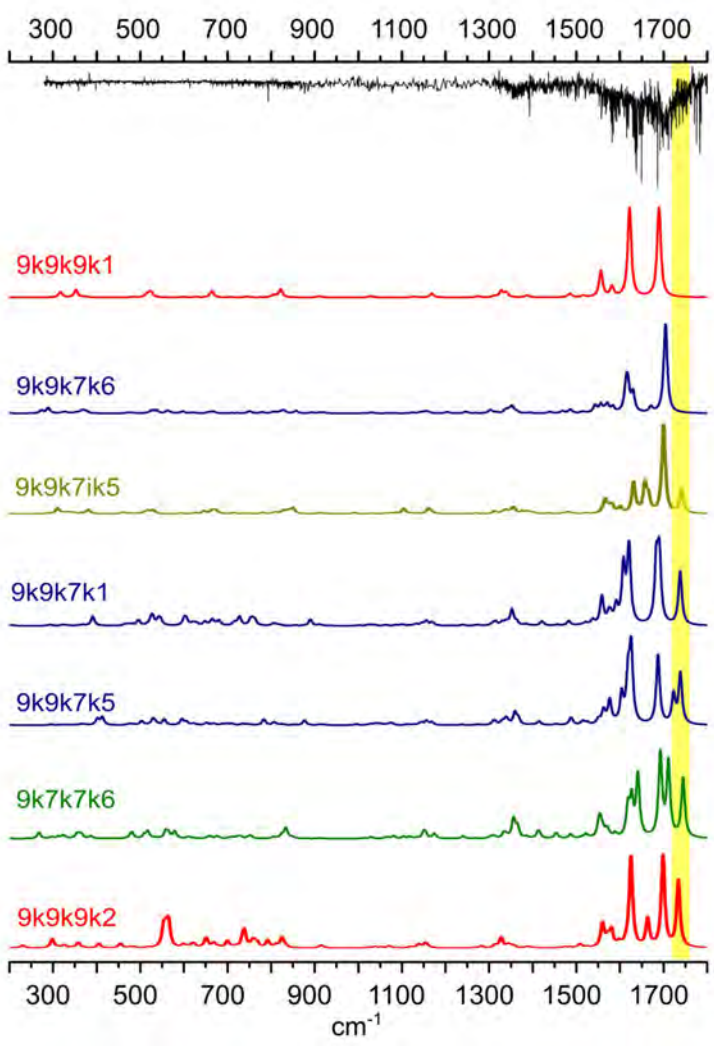


Figure 6.42 *Comparison between the mass-resolved IR/UV spectrum of guanine trimer in 300-1800 cm^{-1} IR region and the simulated spectra of the seven species that may be contributing to the spectrum. The region between 1730-1760 was highlighted in yellow.*

Finally, the third family (type III), whose representative conformer is 9k9k7k1, contains non-planar conformers, allowing the third structure to stablish the same double N_1-H···O=C_6 / HN-H···O=C_6 hydrogen bond with the 9k-guanine and a HN-H···O=C_6 hydrogen bond with the 7k guanine, together with a strong N_1-H···O=C_6 van der Waals interaction, depicted in mid blue in the NCI plot, due to the small angle of the interaction.

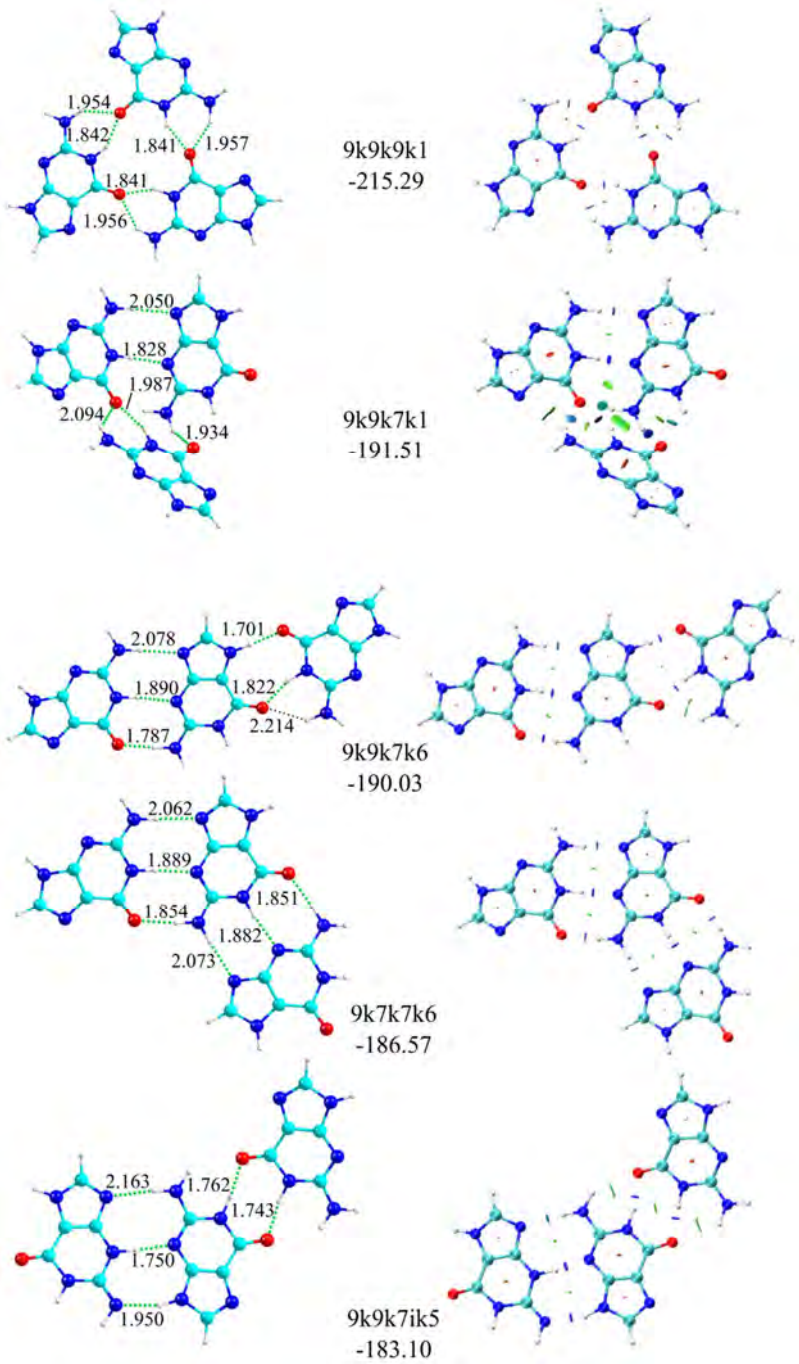


Figure 6.43 *The structure of the five more probable trimers of guanine with distances (Å) of the hydrogen bonds and most relevant van der Waals interactions depicted in NCI plots. Binding energy values in kJ mol^{-1} are also shown.*

In summary, the structure of the dimer is present in the trimers, as observed in cytosine aggregates. It seems that trimers are primarily formed by a favored dimer structure plus a monomer of guanine.

Table 6.8. *Electron density (a.u.) of the hydrogen bonds observed for the six possible trimers of guanine. Dot lines highlight the structure of the 9k7k1 dimer, observed in type II of the trimers.*

	Electron Density (a.u.)				
	I		II		III
	9k9k9k1	9k9k7ik5	9k9k7k6	9k7k7k6	9k9k7k1
N1-H^{I}···O=C6^{II}	0.0290	0.0417			
HN-H^{I}···O=C6^{II}	0.0238				
C6=O^{I}···H-N1^{III}	0.0290				0.0203
C6=O^{I}···H-NH^{III}	0.0238				0.0188
N1-H^{II}···O=C6^{III}	0.0290				
HN-H^{II}···O=C6^{III}	0.0238				0.0270
C6=O^{I}···H-N1^{II}		0.0396			
HN-H^{I}···N9^{II}			0.0244	0.0216	0.0226
N1-H^{I}···N3^{II}			0.0365	0.0335	0.0384
C6=O^{I}···H-NH^{II}			0.0270	0.0292	
HN-H^{II}···N9^{III}		0.0174		0.0212	
N3 ^{II}···H-N1^{III}		0.0466			
N9-H ^{II}···H-N^{III}		0.0289			
C6=O^{II}···H-N1^{III}			0.0329		
N7-H^{II}···O=C6^{III}			0.0434		
N7-H^{I}···H-NH^{III}					
N1-H^{II}···N3^{III}				0.0338	
C6=O^{II}···H-NH^{III}				0.0294	

6.4 9-methylguanine

As in section 6.2, the study on the aggregation preferences of guanine was complemented with exploration of the aggregates of 9-methylguanine. The purpose of the methylation in position N9 is to reproduce the effect of the deoxyribose attached to the nucleoside in the aggregation process. The study will be pure computational again, although comparison with the experimental results published by other groups[30,58,63,77–80] will be offered to determine the accuracy of the calculations.

The structure and the atom labelling of the 9-methylguanine (9-mG) is presented in the **Figure 6.44.**

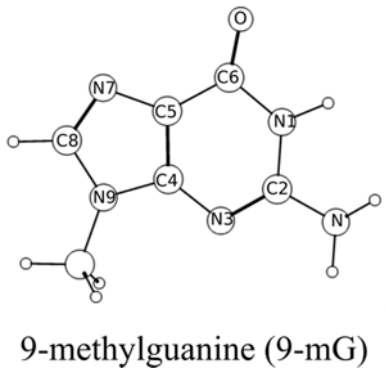


Figure 6.44 *Structure and atom labeling of 9-methylguanine.*

6.4.1 9-methylguanine monomer

Methylation of the N9-position substantially reduces the number of possible tautomers, because 7-k, 7-e and 7-ik tautomers are no longer possible, However, there are still several isoenergetic species, as the data in **Figure 6.45** demonstrate. Also, three tautomers are favored in the whole range of temperatures: 9m-k1, 9m-e1 and 9m-e2, that in addition present a small energy difference of ~ 1.5 kJ mol^{-1}. The imino forms appear above 55 kJ mol^{-1} and therefore, it is very unlikely that they will be present in the expansion. Therefore, only the pure keto and the enol conformers will be selected as "building blocks" to explore the aggregation.

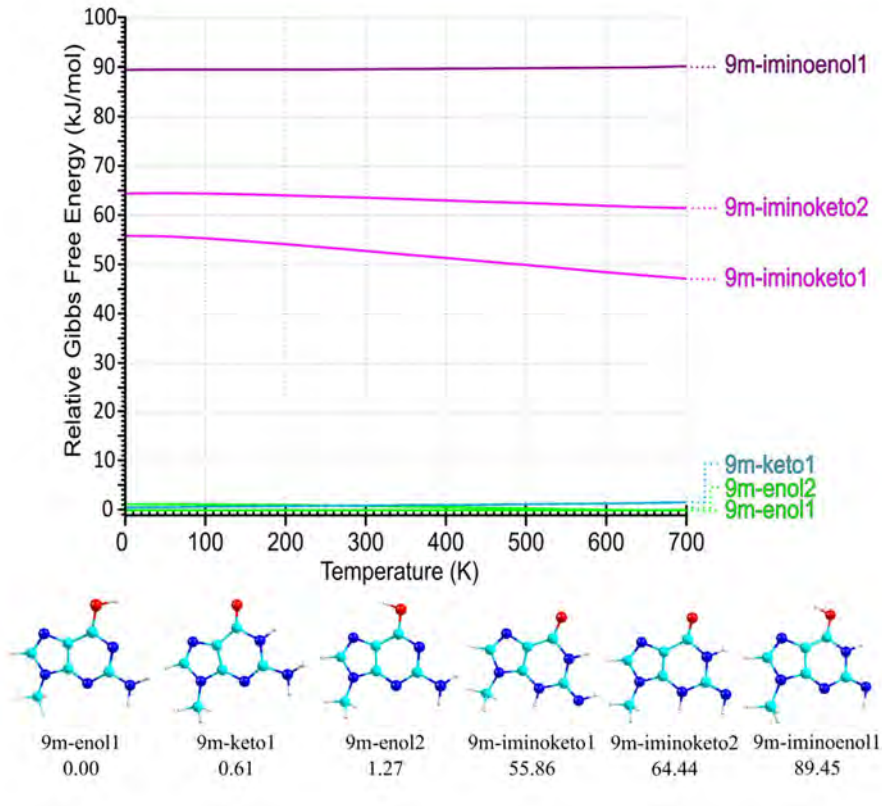


Figure 6.45 *Relative Gibbs free energy of the possible tautomers of 9-methylguanine in the 0-700 K interval and the structure of the tautomers with their relative energy in kJ mol*$^{-1}$.

6.4.2 9-methylguanine dimer

Despite guanine dimer was one of the most studied dimers of DNA bases, only Urashima and co-workers[80] studied the 9-methylguanine dimer, and their hydrated derivatives, by IR/UV double resonance spectroscopy. Thus, at the end of the section, a comparison with our computational results will be offered.

The analysis of 9-mG was focused in the combination of 9m-keto and 9m-enol conformers. The methodology used for the exploration of the conformational landscape resulted in 12, 72 and 79 structures for 9m-kk, 9m-ee and 9m-ke respectively, which were subjected to full optimization using DFT. The resulting situation is significantly simplified respect to guanine dimers, as the data in **Figure 6.46** highlight.

Clearly, 9m-kk1 is more stable than any other conformer in the whole range of temperatures, followed by 9m-ee1, which appears above 24 kJ mol^{-1}. This second conformer is formed by two enol tautomers, establishing N···H-O symmetric hydrogen bonds and therefore, very likely it can relax into 9m-kk1 spontaneously. Furthermore, the first planar keto-enol dimer, 9m-ke4, appears above 30 kJ mol^{-1}. Consequently, only pure keto-keto dimers will be considered as a core for the formation of the trimer and the rest of the families can be safely ruled out.

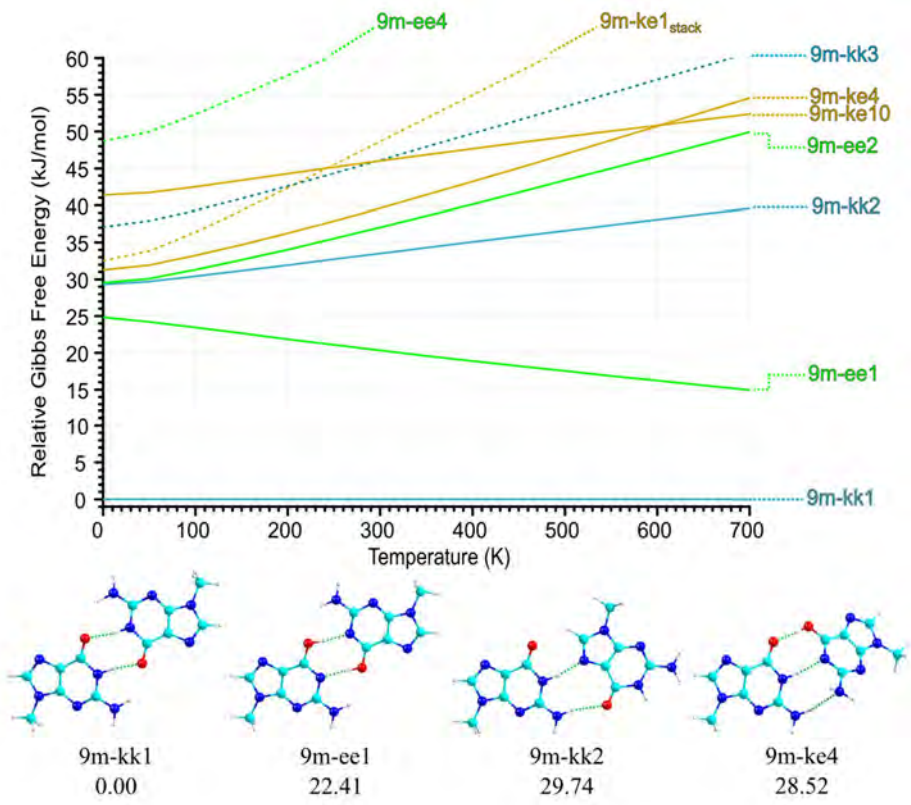


Figure 6.46 *Relative Gibbs free energy of the dimers of 9-methylguanine in the 0-700 K interval and the structure of the most relevant tautomers with their relative energy in kJ* mol^{-1}. *Continous line: planar structure; Dot line: stacking structure. Blue: keto-keto; green: enol-enol; yellow: keto-enol structures.*

In addition, the study of the aggregation tendency underlines that planar structures are preferred over stacking ones, in all the temperatures both for keto-keto and enol-enol homodimers. For example, planar 9m-ee

dimers are separated from staking dimers by 25 kJ mol^{-1} and this difference is even larger for 9m-kk (37 kJ mol^{-1}), remarking the high stability of the planar structures in comparison with stacking ones. In the case of keto-enol heterodimer, the most stable structure, 9m-ke1, adopts a stacking geometry. However, when entropy and temperature are taken into account, 9m-ke4 becomes significantly more stable. This conformer is not completely planar because the interaction between the two NH2 groups. The first truly planar structure is 9m-ke10, which is significantly higher in energy.

As in the case of 1-methylcytosine, the number of planar structures is reduced in comparison with guanine, by the elimination of the N7 tautomerism.

The 9m-kk1 dimer exhibits two symmetrical hydrogen bonds between N_1-$H^{I}\cdots O{=}C_6^{II}$ and $C_6{=}O^{I}\cdots H$-N_1^{II} groups and two weak van der Waals interactions between the NH_2 and the C_6=O. This structure adopts the same disposition than 9k9k dimer of guanine. Comparison between both dimers may be found in **Figure 6.47**. The corresponding electron densities may be found in **Table 6.9**. As it can be seen the strength of the interactions are similar in both conformers.

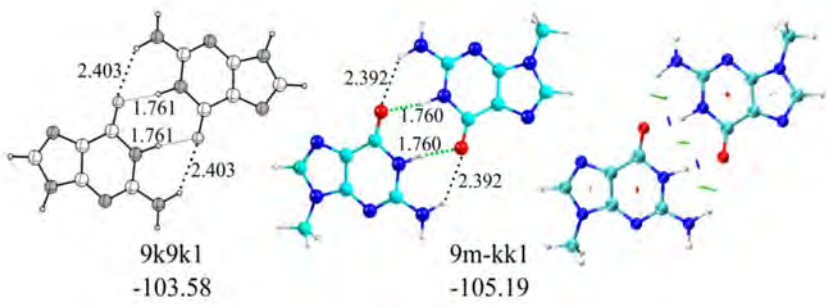


Figure 6.47 *The structure of the 9k9k1 guanine (left) and 9m-kk1(middle), together with the distances (Å) of the hydrogen bonds and the binding energy values in kJ* mol^{-1}*. The most relevant van der Waals interactions for the 9m-kk1 dimer are depicted in the NCI plot (right).*

The experimental results obtained by Urashima and co-workers[80] and a comparison with our theoretical results may be found in **Figure 6.48**. Clearly, the experimental trace is reproduced by the spectrum of 9m-kk1 dimer, the most stable one. The peaks at 3404 and 3543 cm^{-1} are related to the stretching of the NH_2 and fits satisfactorily with the

prediction of the theoretical calculation, using the scaling factor of guanine (0.9520). On the other hand, the predicted NH stretch is slightly shifted compared to the experimental peak.

Table 6.9. *Electron density (a.u.) of the hydrogen bonds observed for the 9k9k1 dimer of guanine and 9m-kk1 dimer of 9-methylguanine.*

	Electron Density (a.u)	
	Guanine	9-methylguanine
	kk1	9m-kk1
$C6{=}O^{I}\cdots H{-}N1^{II}$	0.0398	0.0399
$N1{-}H^{I}\cdots O{=}C6^{II}$	0.0398	0.0399

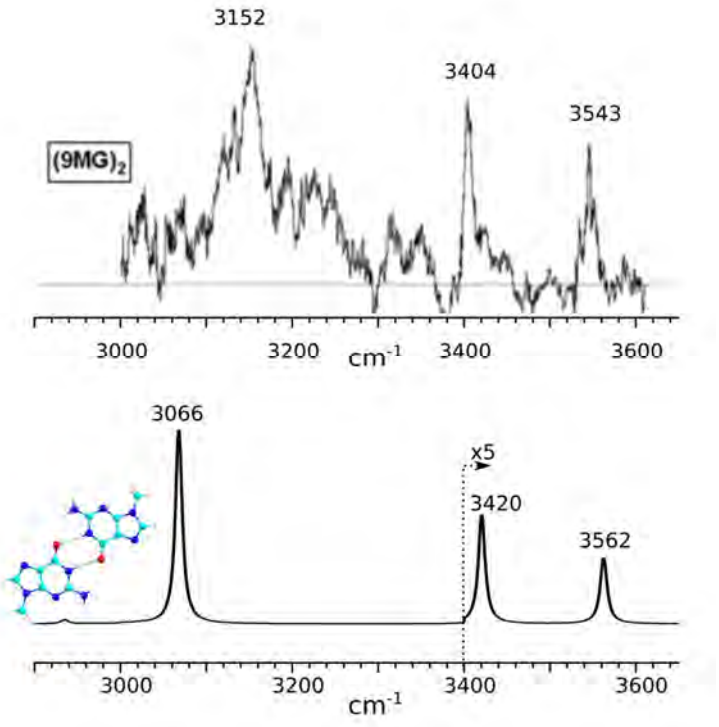


Figure 6.48 *Comparison between the mass-resolved IR/UV spectrum of 9-methylguanine dimer, reported by Urashima et al.*[80]*, and the simulated spectra of the 9m-kk1 dimer.*

6.4.3 9-methylguanine trimer

Increasing the cluster size always results in an increase in the number of possible structures, due to simple combinatorial reasons. However, in this case the third 9-methylguanine produced a modest increase, compared to its guanine homologous. Certainly, we only needed to fully optimize 54 different structures to have a good description of the conformational landscape. The variation of the Gibbs energy of the 9m-kkk trimers along the temperature may be found in **Figure 6.49**, together with the structure of the more relevant trimers and their relative energy in kJ mol^{-1}.

Following the tendency observed for guanine and 9-methylguanine, the 9m-kkk1 homotrimer dominates the conformational landscape in the whole range of temperature. The next structure, 9m-kkk2 appears above 27 kJ mol^{-1} and it continues losing relative stability with temperature, up to 30 kJ mol^{-1} at 100 K. Furthermore, the differences between stacking and H-bonded structures observed in pure 9k guanine are maintained for its methylated form. In the previous section, it was clear that H-bonded 9k9k9k guanine trimer was more stable than the first stacking structure, around 25 kJ mol^{-1}. This difference is even higher for 9m-kkk conformers, where stacked structures are separated by more than 28 kJ mol^{-1}. Nevertheless, the next 9m-kkk planar structure, 9m-kkk18, appears above 33 kJ mol^{-1}. Therefore, it is clear the extraordinary stability of 9m-kkk1 conformer.

Focusing in the structure of this aggregate, 9m-kkk1 is formed by the repetition of the same two interactions between the three constituent 9-methylguanine molecules: N_1-H$\cdots$O=C_6 and HN-H$\cdots$O=C_6.The 9m-kkk1 aggregate adopts the same disposition as 9k9k9k1 conformer. The geometries of both aggregates, together with the distances of the most relevant interactions, the NCI plots and the values of their binding energy may be found in **Figure 6.50** and the values of the electron density of the hydrogen bonds were collected in **Table 6.10**. Thus, the first hydrogen bond is slightly stronger than the second one, 0.00290 vs. 0.0238 a.u. The distances of the hydrogen bonds are quite similar to those of the most stable 9-keto guanine trimer. Nevertheless, the binding energy of 9m-kkk1 9k9k9k1 are within the computation error. In summary, the methylation of the N9-position does not alter the interaction pattern already observed in pure 9-keto trimer, which exhibits a 9m-kkk1 structure with an extraordinary stability.

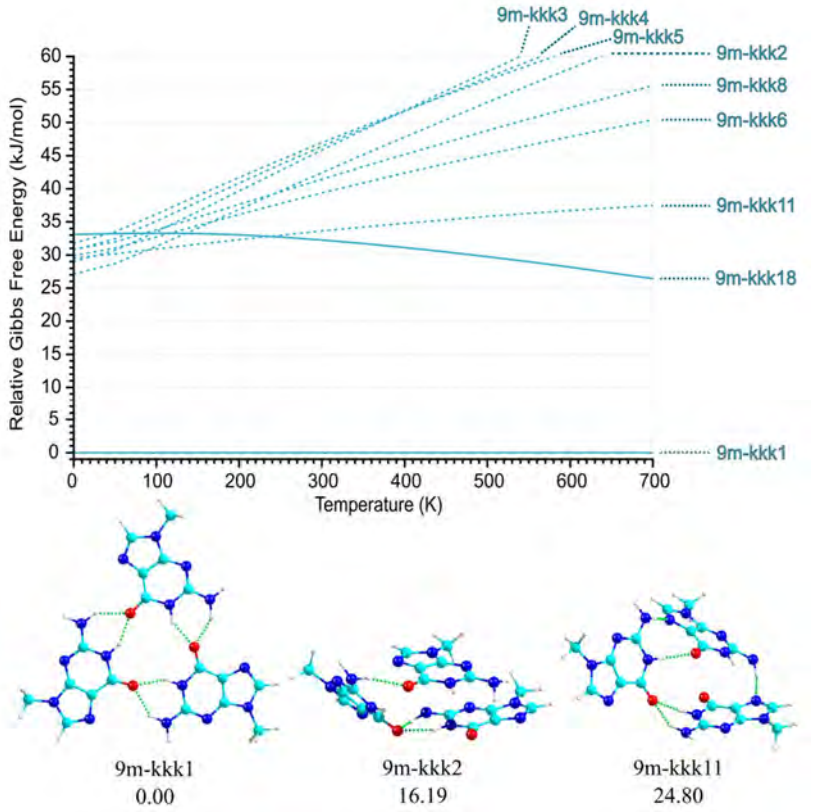


Figure 6.49 *Relative Gibbs free energy of the trimers of 9-methylguanine in the 0-700 K interval and the structure of the most relevant tautomers with their relative energy in kJ mol^{-1}. Continous line: planar structure; Dot line: stacking structure.*

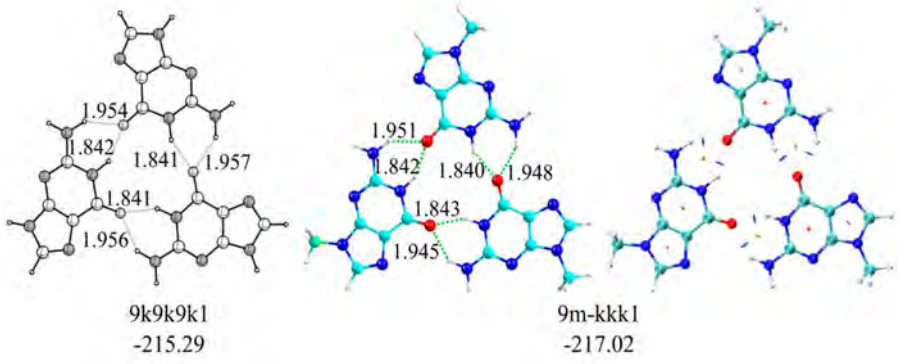


Figure 6.50 *The structure of the 9k9k9k1 guanine (left) and 9m-kkk1 with distances (Å) of the hydrogen bonds and the binding energy values in kJ mol^{-1}. The most relevant van der Waals interactions for the 9m-kk1 dimer are depicted in the NCI plot (right).*

Table 6.10. *Electron density (a.u.) of the hydrogen bonds observed for the 9k9k9k1 trimer of guanine and 9m-kkk1 trimer of 9-methylguanine.*

	Electron Density (a.u)	
	Guanine	9-methylguanine
	9k9k9k1	9m-kkk1
$N1\text{-}H^{I}\cdots O{=}C6^{II}$	0.0290	0.0290
$HN\text{-}H^{I}\cdots O{=}C6^{II}$	0.0238	0.0244
$C6{=}O^{I}\cdots H\text{-}N6^{III}$	0.0290	0.0290
$C6{=}O^{I}\cdots H\text{-}NH^{III}$	0.0238	0.0244
$N1\text{-}H^{II}\cdots O{=}C6^{III}$	0.0290	0.0290
$HN\text{-}H^{II}\cdots O{=}C6^{III}$	0.0238	0.0244

6.5 Discussion

The experimental results demonstrated that the main interaction mechanism of isolated cytosine and guanine proceeds through formation of hydrogen bonds resulting in ribbon-like structures for cytosine and planar structures for guanine, similar to those found when cytosine and guanine are deposited over metal surfaces.[52,81,82] Such structures are also similar to those reported in crystals, although the interaction with the neighboring ribbons or to the packing in the crystal can change the intermolecular distances.[83]

The tautomeric panorama is different for both systems. For cytosine, only keto-based structures were found in the aggregates, despite that the isolated enol tautomer is more stable. The loss of stability due to tautomerization into the keto form is largely compensated by the increase in interaction energy between the two molecules. This is clearly demonstrated both by the computational results and by the detection of aggregates containing exclusively keto tautomers. On the other hand, the tautomeric complexity of guanine is higher and several tautomers are present in gas expansion. Thus, 9/7-keto, 9/7-enol and 7-iminoketo guanines may be found and 9/7-enol and 7-iminoketo tautomers were detected by IR/UV spectroscopy. Despite the increase of the size of the system could complicate the conformational complexity, it seems that some combinations are preferred, all of them containing 9-keto monomer

and additional molecules of 9-keto, 7-keto or 7-iminoketo. The barriers for tautomerization from iminoketo into keto or enol are expected to be significantly high and therefore, the energy required for the process may be lost in the first instants of the cooling process. Thus, as several tautomers are necessary to explain the IR spectra, additional selective experiments, such as UV/UV hole burning or IR/UV/UV techniques, may shed some light on more accurate assignments.

There are two dominant orientations for the interaction of cytosine molecules: N1-$H^{I}\cdots N3^{II}$ // C2=$O^{I}\cdots$H-NH^{II} or the symmetric N1-$H^{I}\cdots$O=$C2^{II}$ // C2=$O^{I}\cdots$H-N1. Both are isoenergetic in the dimer, but it seems that the former is favoured "in the long run", as aggregates larger than the dimer prefer the former orientation. For guanine, there are three dominant aggregation paths: a symmetric N1-$H^{I}\cdots$O=$C6^{I}$ // C6=$O^{I}\cdots$H-$N1^{II}$ hydrogen bonds, with two HN-$H^{I}\cdots$O=$C6^{II}$ // C6=$O^{I}\cdots$H-$N1^{II}$ vdW interactions, exhibited by 9k9k1 and 9k7k2 (type I); three hydrogen bonds for the 9k7k1 and 7k7k1 structures (type II); and a double N1-HI$\cdots$N3II // C6=$O^{I}\cdots$H-$N7^{II}$ hydrogen bonds, exhibited by the 9k7ik1 and 9k7ik2 forms (type III). The difference between the two structures of each group is related with the tautomeric N9/N7 hydrogen (type I and II) or the rotation of the imino group (type III). In higher aggregates, the pattern of 9k9k1 or 9k7k1 dimers is observed in the most stable trimers.

The hydrogen bonds between cytosine and guanine monomers are extraordinarily strong, as reflected by their appearance in the spectrum as very broad absorptions, shifted to the red, to a region usually occupied by the CH stretches. Such strong interactions may compensate the stress in the DNA strands introduced by the torsions required to put in close contact the four molecules of the quadruplex. However, further experiments will be required to understand why those structures are formed only in DNA sections with a certain sequence of nucleotides.

As the aggregate grows, alternative interaction structures appear, such as formation of stacking structures or cycles. For cytosines, the former type is dominant in the tetramer from an enthalpic point of view. However, when the BSSE correction is taken into account and the temperature is introduced in the equation, ribbon-like structures become more stable. It may well be that the special conditions of the expansion favour detection of stacking structures although they are higher in energy. Certainly, stacked dimers may be formed by collision of dimers if the energy released by the stacking interaction is not large enough to dissociate the pre-existing structures or to surmount the barriers connecting them to the ribbon-like isomers. Tetramers may also be formed

by collision of a trimer with a monomer. Although the most abundant species is the monomer, dimeric aggregates are significantly more abundant than trimers. More difficult to evaluate is the collision cross section for the aggregation of a monomer + trimer compared to dimer + dimer. In any case, such mechanisms cannot be ruled out as responsible for the detection of the stacking structures, despite the fact that they are ~5 kJ mol^{-1} higher in energy at 100 K. For guanine trimers, planar H-bonded structures are favored entropically and enthalpically in comparison with stacking structures. However, a cyclic structure, 9k9k9k1 trimer, exhibits an extraordinary stability in comparison with ribbon-like planar guanine trimers.

6.5.1 Extrapolation to a more biological environment.

A deoxyribofuranose is attached to N1 of cytosine or N9 in guanosine in DNA, blocking one of the sites of the molecule to form intermolecular hydrogen bonds. This situation can be simulated adding a methyl group to cytosine in position one and to guanine in position nine, to make 1-methylcytosine or 9-methylguanine.

Computations of the aggregation preferences of 1-methylcytosine significantly increases the propensity of the base towards formation of stacked structures already in the trimer and they become both enthalpically and entropically more stable for the tetramer. However, the aggregation preferences of the 9-methylguanine does not alter the H-bond pattern and planar structures are clearly favoured in trimers. In fact, the same cyclic structure of guanine trimer is observed in its methylated derivative. The preference for the stacking conformation would be further favour in the nucleoside due to the ribose in N1 or N9 position. Previous studies using NMR on the interaction between polysaccharides[84] and in supersonic expansions using similar techniques as those employed in this work[85] show that sugars tend to form stacking structures, especially in solution, because in that way the hydrophobic surface in contact with water molecules is reduced. In good agreement, in an exploration of the aggregation preferences of adenosine dimers in jets, Asami et al.[86] found almost exclusive formation of stacked dimers. Unfortunately, the authors did not explore larger aggregates. Probably the small S/N ratio rendered by such a difficult system precluded the recording of the spectroscopy of higher-order clusters.

It is also difficult to extrapolate such observations to solution. On the one hand, in solution and at the temperature range of life, entropy certainly plays an important role, and therefore one would expect to find only ribbon-like structures for cytosines and planar structures for guanines. However, water may over compensate the entropic effect. Previous computational studies on the effect of water molecules on formation of cytosine-guanine[87] and adenine-thymine dimers[88,89] clearly demonstrated that two water molecules are enough to favour stacking structures over hydrogen bonded ones. On those studies they also observed a preference for the interaction between keto tautomers, in agreement with our own results. Thus, the studies seem to point to the existence of stacking dimers in solution.

All these evidences point to a reinforcement of the stacking interactions in cytidine compared with cytosine that combined with the effect of solvation may well tip the balance towards formation of stacked structures of dimers, present in i-motifs. For guanine aggregates, the extraordinary stability of the cyclic guanine trimer points to a different aggregation preference, guided by a cyclical planar structure, even in their methylated derivatives, which can be the reason why G-quadruplexes are found in DNA. Guanine tetramers need to be studied in order to support this final hypothesis.

6.6 Conclusions

In this chapter, the aggregation of the cytosine up to tetramer and guanine up to trimer was studied by mass-resolved laser spectroscopy and Quantum Mechanics calculations. Furthermore, the conformational landscape of their methylated derivatives, 1-methylcytosine and 9-methylguanine, were also analyzed by DFT calculations.

The experimental conditions were controlled to obtain the IR spectra from the monomer to trimer of cytosine and guanine and the tetramer of cytosine. Two regions can be perfectly distinguished: from 3700 to 3400 cm^{-1}, where narrow bands appear due to stretching of free hydrogen of NH groups (except for guanine trimer, whose band at 3503 is broad due to a combination of several bands) and the from 2700 to 3400 cm^{-1}, where broad peaks appear due to the stretching of hydrogens involved in hydrogen bonds.

The tautomeric complexity of these systems complicates highly the aggregation study. It has been observed that several tautomeric forms can

coexist in our experimental conditions. This makes the aggregation process sometimes extremely hard to address, especially for the guanine, which exhibits five different possible tautomers. Despite such amount of possibilities, it has been observed that keto forms are the preferred ones to form the aggregates, both cytosine and guanine. Furthermore, the methylation of the N-position to form the nucleoside derivatives decreases substantially that complexity. However, keto tautomers of methylated DNA bases are still the preferred ones to form the aggregates. Therefore, the biological tautomers of the studied DNA bases are the preferred structures to form aggregates, which suggest why nature selects these species for the storage of the genetic information.

The aggregation process of the DNA bases is guided by a hydrogen bond pattern. This pattern is different for the two DNA bases. On the one hand, cytosine adopts ribbon-like structures, being assembled by N_1-H$\cdots$$N_3$ and C_2=O$\cdots$H-NH hydrogen bonds, the structure of the most stable cytosine dimer. Thus, the structure of the dimer is observed in the assigned trimers and tetramers. On the other hand, 9-keto guanine dimers adopt almost planar structure, throughout two symmetrical N_1-H$\cdots$O=C_6 hydrogen bonds and two HN-H$\cdots$O=C_6 van der Waals interactions. The addition of the third molecule opens the dimer to stablish a triple symmetrical N_1-H$\cdots$O=C_6 and HN-H$\cdots$O=C_6 hydrogen bonds, adopting a planar cyclic structure. Furthermore, the aggregation process of their methylated derivatives exhibits several differences between them. On the one hand, the methylation of the N_1 position of the 1-methylcytosine eliminates an important point of contact observed in cytosine aggregates and the most stable structure adopts the structure of the kk3 cytosine dimer, with two symmetrical HN-H$\cdots$$N_3$ hydrogen bonds. Moreover, 1-mehtylcytosine trimer and tetramer adopted again the same structure observed in dimers. However, the methylation of the N9 position of the guanine does not alter the aggregation pattern observed in the 9k-guanine aggregates and the upper aggregates of 9-methylguanine maintain the geometries of 9k-guanine.

The competition between stacking and planar structures has been analyzed for DNA bases. Again, the results are different for the two DNA bases. Both DNA bases exhibit planar dimers and the difference with stacking structures is high. However, when the size of the aggregate is increased, some differences are observed. In the case of the cytosine, the difference between stacking and planar structures is reduced in trimers and stacking tetramers are more stable than planar ones. However, planar tetramers are favored entropically, what suggest a competition between both tendencies in the tetramers. This competition is confirmed by the

experiments, where the IR bands are explained by the combination of planar and stacking tetramers. Moreover, its methylated derivative exhibits the same pattern, even sooner than in cytosine. Thus, this competition appears in 1-methylcytosine trimer and tetramer exhibited the opposite tendency, where stacking structures are favored entalpically and entropically. On the other hand, guanine and its methylated form do not follow this pattern. Planar structures are favored and the difference is not reduced with the increase of the system.

Finally, it is difficult to extrapolate these observations to the biological real systems (nucleosides) or with water solvent. Despite this effect has been emulated with a methyl group, the natural tendency of the sugars to form stacked structures[84,85] could reinforce the competition between stacking and planar aggregates. On the other hand, the presence of water solvent also tend to hamper the formation of planar structures and more experiments can be done in this direction in order to shed some light. However, the importance of the hydrogen bond as the intrinsic force in the aggregation of the DNA bases has undoubtedly proven.

6.5 References:

1 D. L. Nelson and M. M. Cox, *Principles of Biochemistry*, Omega, New York, 4th edn., 2000.

2 B. Alberts, A. Johnson, D. Morgan, M. Raff and J. Lewis, *Molecular Biology of the Cell*, Taylor & Francis Inc, New York, 2014.

3 L. Stryer, J. M. Berg and J. L. Tymoczko, *Biochemistry*, Reverté, New York, 5th edn., 2002.

4 E. T. Kool, J. C. Morales and K. M. Guckian, *Angew. Chemie - Int. Ed.*, 2000, **39**, 990–1009.

5 S. Lobsiger, M. Etinski, S. Blaser, H. M. Frey, C. Marian and S. Leutwyler, *J. Chem. Phys.*, 2015, **143**, 0–12.

6 C. G. Triandafillou and S. Matsika, *J. Phys. Chem. A*, 2013, **117**, 12165–12174.

7 D. Mishra and S. Pal, *J. Mol. Struct.-THEOCHEM*, 2009, **902**, 96–102.

8 O. Kostko, K. Bravaya, A. Krylov and M. Ahmed, *Phys. Chem. Chem. Phys.*, 2010, **12**, 2860–2872.

9 A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, *Proc. Natl. Acad. Sci.*, 2005, **102**, 20–23.

10 K. Kleinermanns, D. Nachtigallová and M. S. De Vries, *Int. Rev. Phys. Chem.*, 2013, **32**, 308–342.

11 C. E. Crespo-Hernández, L. Martínez-Fernández, C. Rauer, C. Reichardt, S. Mai, M. Pollum, P. Marquetand, L. González and I. Corral, *J. Am. Chem. Soc.*, 2015, **137**, 4368–4381.

12 C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández and B. Kohler, *Annu. Rev. Phys. Chem.*, 2009, **60**, 217–239.

13 E. Mburu and S. Matsika, *J. Phys. Chem. A*, 2008, **112**, 12485–12491.

14 Z. Gengeliczki, M. P. Callahan, N. Svadlenak, C. I. Pongor, B. Sztáray, L. Meerts, D. Nachtigallová, P. Hobza, M. Barbatti, H. Lischka and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5375–5388.

15 Z. Y. Kan, Y. Lin, F. Wang, X. Y. Zhuang, Y. Zhao, D. W. Pang, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2007, **35**, 3646–3653.

16 H. Han and L. H. Hurley, *Trends Pharmacol. Sci.*, 2000, **21**, 136–142.

17 V. Setnička, J. Nový, S. Böhm, N. Sreenivasachary, M. Urbanová and K. Volka, *Langmuir*, 2008, **24**, 7520–7527.

18 Q. Wang, J. Q. Liu, Z. Chen, K. W. Zheng, C. Y. Chen, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2011, **39**, 6229–6237.

19 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–80.

20 M. Rueda, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2006, **128**, 3608–3619.

21 V. Sychrovský, Z. Sochorová Vokáčová and L. Trantírek, *J. Phys. Chem. A*, 2012, **116**, 4144–4151.

22 M. Guéron and J. L. Leroy, *Curr. Opin. Struct. Biol.*, 2000, **10**, 326–331.

23 H. A. Day, P. Pavlou and Z. A. E. Waller, *Bioorganic Med. Chem.*, 2014, **22**, 4407–4418.

24 A. T. Phan and J.-L. Mergny, *Nucleic Acids Res.*, 2002, **30**, 4618–4625.

25 S. Benabou, R. Ferreira, A. Aviño, C. González, S. Lyonnais, M. Solà, R. Eritja, J. Jaumot and R. Gargallo, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, **1840**, 41–52.

26 J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 2440–2450.

27 L. Cai, L. Chen, S. Raghavan, R. Ratliff, R. Moyzis and A. Rich, *Nucleic Acids Res.*, 1998, **26**, 4696–4705.

28 P. K. Patel, N. S. Bhavesh and R. V Hosur, *Biochem. Biophys. Res. Commun.*, 2000, **270**, 967–71.

29 B. Pan, Y. Xiong, K. Shi, J. Deng and M. Sundaralingam, *Structure*, 2003, **11**, 815–823.

30 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. De Vries, *J. Chem. Phys.*, 2001, **115**, 4604–4611.

31 E. Nir, I. Hunig, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2003, **5**, 4780.

32 E. Nir, C. Plützer, K. Kleinermanns and M. De Vries, *Eur. Phys. J. D*, 2002, **20**, 317–329.

33 E. Nir, L. Grace, B. Brauer and M. S. De Vries, *J. Am. Chem. Soc.*, 1999, **121**, 4896–4897.

34 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2002, **4**, 740–750.

35 E. Nir, I. Hünig, K. Kleinermanns and M. S. De Vries, *ChemPhysChem*, 2004, **5**, 131–137.

36 C. Plützer, I. Hünig, K. Kleinermanns, E. Nir and M. S. De Vries, *ChemPhysChem*, 2003, **4**, 838–842.

37 J. M. Bakker, I. Compagnon, G. Meijer, G. von Helden, M. Kabelac, P. Hobza and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2810.

38 M. S. de Vries, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, eds. A. M. Rijs and J. Oomens, Springer International Publishing, Cham, 2015, pp. 271–297.

39 E. Nir, M. Müller, L. I. Grace and M. S. De Vries, *Chem. Phys. Lett.*, 2002, **355**, 59–64.

40 E. Nir, P. Imhof, K. Kleinermanns and M. S. De Vries, *J. Am. Chem. Soc.*, 2000, **122**, 8091–8092.

41 K. Seefeld, R. Brause, T. Häber and K. Kleinermanns, *J. Phys. Chem. A*, 2007, **111**, 6217–6221.

42 E. Nir, K. Kleinermanns and M. S. de Vries, *Nature*, 2000, **408**, 949–951.

43 J. Sponer, J. Leszczynski and P. Hobza, *J. Comput. Chem.*, 1996, **17**, 841–850.

44 P. Mignon, S. Loverix, J. Steyaert and P. Geerlings, *Nucleic Acids Res.*, 2005, **33**, 1779–1789.

45 R. Parthasarathi and V. Subramanian, *Struct. Chem.*, 2005, **16**, 243–255.

46 Z. Yang and M. T. Rodgers, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2749–2757.

47 Y. Zhao and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2701–2705.

48 D. Sivanesan, K. Babu, S. R. Gadre, V. Subramanian and T. Ramasami, *J. Phys. Chem. A*, 2000, **104**, 10887–10894.

49 Z. Czyznikowska and R. Zaleśny, *Biophys. Chem.*, 2009, **139**, 137–143.

50 G. Hill, G. Forde, N. Hill, W. A. Lester, W. A. Sokalski and J. Leszczynski, *Chem. Phys. Lett.*, 2003, **381**, 729–732.

51 P. Jurečka, J. Šponer and P. Hobza, *J. Phys. Chem. B*, 2004, **108**, 5466–5471.

52 A. Manukyan and A. Tekin, *Phys. Chem. Chem. Phys.*, 2015, **17**, 14685–701.

53 J. L. Alonso, V. Vaquero, I. Peña, J. C. López, S. Mata and W. Caminati, *Angew. Chemie - Int. Ed.*, 2013, **52**, 2331–2334.

54 Y. Valadbeigi, M. Soleiman-Beigi and R. Sahraei, *Chem. Phys. Lett.*, 2015, **629**, 1–7.

55 I. Leon, I. Usabiaga, J. Millan, E. J. Cocinero, A. Lesarri and J. A. Fernandez, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16968–16975.

56 P. M. Keane, M. Wojdyla, G. W. Doorley, G. W. Watson, I. P. Clark, G. M. Greetham, A. W. Parker, M. Towrie, J. M. Kelly and S. J. Quinn, *J. Am. Chem. Soc.*, 2011, **133**, 4212–4215.

57 J. A. Frey, P. Ottiger and S. Leutwyler, *J. Phys. Chem. B*, 2014, **118**, 682–691.

58 M. Mons, I. Dimicoli, F. Piuzzi, B. Tardivel and M. Elhanine, *J. Phys. Chem. A*, 2002, **106**, 5088–5094.

59 W. Chin, M. Mons, I. Dimicoli, F. Piuzzi, B. Tardivel and M. Elhanine, *Eur. Phys. J. D*, 2002, **20**, 347–355.

60 F. Piuzzi, M. Mons, I. Dimicoli, B. Tardivel and Q. Zhao, *Chem. Phys.*, 2001, **270**, 205–214.

61 M. Mons, F. Piuzzi, I. Dimicoli, L. Gorb and J. Leszczynski, *J. Phys. Chem. A*, 2006, **110**, 10921–10924.

62 J. L. Alonso, I. Peña, J. C. López and V. Vaquero, *Angew. Chemie - Int. Ed.*, 2009, **48**, 6141–6143.

63 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2002, **4**, 740–750.

64 C. M. Marian, *J. Phys. Chem. A*, 2007, **111**, 1545–1553.

65 M. Y. Choi and R. E. Miller, *J. Am. Chem. Soc.*, 2006, **128**, 7320–7328.

66 W. Liang, H. Li, X. Hu and S. Han, *Chem. Phys.*, 2006, **328**, 93–102.

67 I. R. Gould, M. A. Vincent and I. H. Hillier, *Spectrochim. Acta Part A Mol. Spectrosc.*, 1993, **49**, 1727–1734.

68 M. Piacenza and S. Grimme, *J. Comput. Chem.*, 2004, **25**, 83–98.

69 M. K. Shukla and J. Leszczynski, *Chem. Phys. Lett.*, 2006, **429**, 261–265.

70 N. Gardner, D. Magers and G. Hill, *Struct. Chem.*, 2014, **25**, 347–353.

71 I. Leon, E. J. Cocinero, J. Millán, S. Jaeqx, A. M. Rijs, A. Lesarri, F. Castaño and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2012, **14**, 4398–4409.

72 I. León, J. Millán, E. J. Cocinero, A. Lesarri and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23301–23307.

73 I. León, P. F. Arnáiz, I. Usabiaga and J. A. Fernández, *Phys. Chem. Chem. Phys.*, 2016, **18**, 27336–27341.

74 A. Lüchow, D. Spangenberg, C. Janzen, A. Jansen, M. Gerhards and K. Kleinermanns, *Phys. Chem. Chem. Phys.*, 2001, **3**, 2771–2780.

75 C. Janzen, D. Spangenberg, W. Roth and K. Kleinermanns, *J. Chem. Phys.*, 1999, **110**, 9898–9907.

76 M. S. de Vries, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Springer International Publishing, Cham, 1st edn., 2015, vol. 11, pp. 271–297.

77 E. Nir, K. Kleinermanns, L. Grace and M. S. De Vries, *J. Phys. Chem. A*, 2001, **105**, 5106–5110.

78 E. Nir, C. Plützer, K. Kleinermanns and M. De Vries, *Eur. Phys. J. D*, 2002, **20**, 317–329.

79 U. Saigusa, S. H. Urashima and H. Asami, *J. Phys. Chem. A*, 2009, **113**, 3455–3462.

80 S. H. Urashima, H. Asami, M. Ohba and H. Saigusa, *J. Phys. Chem. A*, 2010, **114**, 11231–11237.

81 R. E. A. Kelly, M. Lukas, L. N. Kantorovich, R. Otero, W. Xu, M. Mura, E. Lægsgaard, I. Stensgaard and F. Besenbacher, *J. Chem. Phys.*, 2008, **129**, 184707.

82 M. Furukawa, H. Tanaka and T. Kawai, *J. Chem. Phys.*, 2001, **115**, 3419–3423.

83 D. L. Barker and R. E. Marsh, *Acta Crystallogr.*, 1964, **17**, 1581–1587.

84 A. G. Santana, E. Jiménez-Moreno, A. M. Gómez, F. Corzana, C. González, G. Jiménez-Oses, J. Jiménez-Barbero and J. L. Asensio, *J. Am. Chem. Soc.*, 2013, **135**, 3347–3350.

85 I. Usabiaga, J. González, I. León, P. F. Arnaiz, E. J. Cocinero and J. A. Fernández, *J. Phys. Chem. Lett.*, 2017, **8**, 1147–1151.

86 H. Asami, K. Yagi, M. Ohba, S. H. Urashima and H. Saigusa, *Chem. Phys.*, 2013, **419**, 84–89.

87 T. Zelený, P. Hobza and M. Kabelác, *Phys. Chem. Chem. Phys.*, 2009, **11**, 3430–5.

88 M. Kabeláč, F. Ryjáček and P. Hobza, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4906–9.

89 D. Sivanesan, I. Sumathi and W. J. Welsh, *Chem. Phys. Lett.*, 2003, **367**, 351–360.

CHAPTER 7

CONCLUSIONS AND FUTURE PERSPECTIVES

7. - CONCLUSIONS AND FUTURE PERSPECTIVES

The conformational complexity and aggregation process of the components of two biological building blocks, proteins and DNA, has been analyzed in depth. The study started with the computational analysis of the intramolecular interactions of isolated amino acids and DNA bases. Thus, hydrogen bonds and van der Waals dispersive forces control the orientation of the non-polar (alanine, valine and isoleucine), polar (asparagine, glutamine and arginine) and aromatic (phenylalanine, tryptophan and tyrosine) amino acids, for which we perform a detailed conformational map. Furthermore, the side chain also plays its role in the final conformation. Thus, for non-polar amino acids, the lateral chain stays away from the backbone, whereas, for polar amino acids, the lateral chain plays a main role in the stability of the system. The same applies to the aromatic amino acids, which exhibit interactions between the backbone and the aromaticity of its rings in their lateral chain. The analysis of the most relevant non-covalent intramolecular interactions found for the two most stable conformations shows that they are always cyclic structures, having a ring of 7 (γ_L or γ_D) or 5 (β_L) members. Both conformations are of similar stability, with differences in most cases within the computation error in all the range of temperatures investigated (0-700 K). On the other hand, DNA bases exhibit only weak destabilizing interactions due to 5 and 6-member strains of their aromatic rings. Finally, the study also permitted to select the suitable computational method for the calculation of the aggregates: M06-2X, which takes into account dispersive forces, was selected for the rest of the study with a relatively low computational cost.

Once these intramolecular interactions were characterized, DNA bases and amino acids were combined in order to investigate the intermolecular interactions which guide the aggregation process of the DNA-histone superstructure. So, two kinds of intermolecular interactions, moderate hydrogen bonds and van der Waals forces, govern the aggregation process of the clusters. However, it has been demonstrated that hydrogen bond is the leading force of the aggregation. The DNA base – amino acid aggregation follows an interaction pattern, which is different depending on the nature of the amino acid's side chain and the base. On the one hand, focusing in the side chain of the amino acids, three different trends are observed. Non-polar amino acids establish hydrogen bonds between the peptide bond and the DNA base, placing their side chain far from the base. Polar amino acid's side chain tries to maximize the contacts with the base, playing a leading role in the aggregation and

forcing to the whole molecule to accommodate its geometry. Finally, aromatic amino acids with try to get the DNA base closer to the aromatic side chain, to stablish π···π interactions, approaching the peptidic chain to the base to form hydrogen bonds. On the other hand, the nature of the DNA base is also relevant in the aggregation process. So, complementary DNA bases exhibit a similar trend. The dispersive contribution for adenine and thymine aggregates is higher than for cytosine and guanine aggregates, which is traduced in lower binding energy values. Moreover, the balance between hydrogen bond and van der Waals interactions and the binding energies values of amino acid-DNA base pairs are satisfactorily accounted in our study.

Once the intermolecular interactions in the aggregates were analyzed, the analysis about statistical studies of the contacts between amino acids and the DNA bases was attempted. Thus, the mean binding energy of the most stable aggregate of each DNA base – amino acid complex was compared with statistical observations found in literature. There is a good correlation for aliphatic side chain, which exhibits both a lower number of contacts and smaller dissociation values. Furthermore, the increase of the number of contacts for polar amino acids and DNA bases is correlated with their binding energy values, except for asparagine, which exhibited a lower dissociation value than observed in statistical study. Finally, the comparison for the binding energy values of aromatic amino acid - DNA base aggregates is complicated due to the over-stabilization of the binding energy value related with the π···π interactions.

Therefore, it is observed that the interaction between the superstructures of histones and DNA is primarily controlled by the individual amino acid-DNA base pair interactions. However, regarding that some of the proposed points of contact of the aggregates are involved in the complementarity of the DNA bases, the model was improved to reproduce a more realistic biological system. DNA bases were attached to its complementary base and faced to the amino acids to evaluate the changes observed in the aggregation process. As the study shows, the improvement of the model does not make meaningful changes in the conclusions (despite the correlation of the aromatic binding energy values is slightly better) and underlines the noticeable correlation between our results for the binding energies and the propensities protein-DNA base found in literature.

In summary, the computational data provide chemical insight into the dominant interactions representing the DNA-histone interactions,

offering a chemical tool for the prediction of aggregation properties of DNA and highlighting that the sequence of the histones is designed to present to DNA those amino acids which present higher affinity to the bases. Such specificity may well be related to their role in the regulation of DNA expression.

Finally, the self-aggregation of guanine and cytosine was also tested experimentally and computationally. For its biological derivatives only a computationally study was done for the intrinsic tendency in their aggregation process. The experimental conditions were controlled to obtain the IR spectra from the monomer to trimer of cytosine and guanine and the tetramer of cytosine. However, the tautomeric complexity of these systems complicates highly the assignation. Nevertheless, it has been observed that keto tautomers are the preferred ones to form the aggregates in cytosine, guanine and their biological methylated derivatives, which may be point the reason why these tautomeric forms are selected for the storage of the genetic information. However, IR/UV/UV experiments could be done to reduce the amount of possible conformers in the assignations. Furthermore, the aggregation process of the DNA bases is governed by a hydrogen bond pattern, different for both DNA bases. On the one hand, cytosine adopts ribbon-like structures assembled by N_1-H$\cdots$$N_3$ and C_2=O$\cdots$H-NH hydrogen bonds. However, the methylation of the N1 position of the cytosine does not permit this assemblage and a double HN-H$\cdots$$N_3$ hydrogen bond pattern appeared for 1-methylcytosine aggregates. On the other hand, 9-keto guanine adopts almost planar structure, the most stable one exhibited a pattern of N_1-H$\cdots$O=C_6 and HN-H$\cdots$O=C_6 hydrogen bonds. On the other hand, the tautomeric complexity complicates the assignation of the IR spectra, and other planar structures with different points of contacts are observed. 9-methylguanine maintains the geometries of the most favored 9k-guanine. Finally, the competition between stacking and planar structures also shows different results for the two DNA bases. Both exhibit planar dimers with a high difference in energy with stacking structures. However, whereas the increase of the size does not alter the differences between planar and stacking structures in guanine aggregates, in cytosine this difference is reduced in trimers and the stability is inverted in tetramers. So, stacking structures are favored enthalpically but planar tetramers are favored entropically, what suggests a competition between both tendencies in cytosine tetramers, observed in the IR spectrum. Furthermore, methylated derivatives exhibit the same pattern than for cytosine and guanine, even sooner in 1-methylcytosine, where the competition observed for cytosine tetramers appears in the trimer.

Future perspectives

As this thesis is an opening study in the aggregation of the biological building blocks, it is likely to carry out future studies in several directions:

- In the DNA-protein aggregation study, it can be seen that the proposed model explains reasonably well the shape and the interactions present in histone-DNA complexes for amino acids with hydrophobic and hydrophilic side chain, but is must be improved for aromatic amino acids. The future work should be focus in enhancing the model in order to reproduce the propensities of the aromatic amino acids. Thus, a more realistic model with several DNA base pairs placed by stacking is required to reproduce the steric constraints. In fact, we are currently working on a model, shown in **Figure 7.1**, of the interaction between a tripeptide and a section of a DNA formed by six DNA base pairs. Such a bigger system requires of sophisticated computational strategies that combines MM and DFT calculations. The model is still in the first steps of the development.

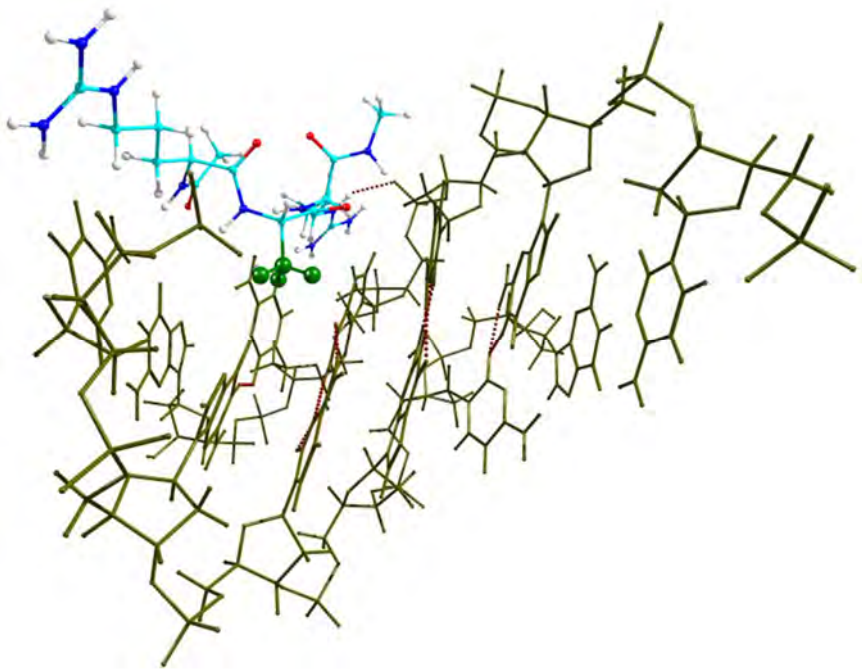


Figure 7.1 *Representation of the model of six DNA base pairs with the sugar backbone (brown) and the tripeptide aR-aA-aR.*

- In the study of the aggregation of DNA bases, different paths can be addressed.

From the experimental point of view, the increase of the size of the aggregates is difficult, especially for cytosine aggregates, due to the low S/N ratio. In the case of guanine, the signal is much more intense and the experiments can be focus in obtaining of the guanine tetramers to compare with the biological G-quadruplexes. In fact, REMPI spectrum of the tetramer of guanine has been already collected and may be found in **Figure 7.2**.

Moreover, in biological systems, cytosine and guanine are commonly paired forming dimers. Thus, the combination of this two DNA bases should be tackled in order to compare the competition between self-aggregation and mixed aggregation. In fact, REMPI spectra from 32900 to 33800 cm^{-1} have been already collected and may also be found in **Figure 7.2**. However, IR spectra are not collected yet and the experiments should be focused in that direction.

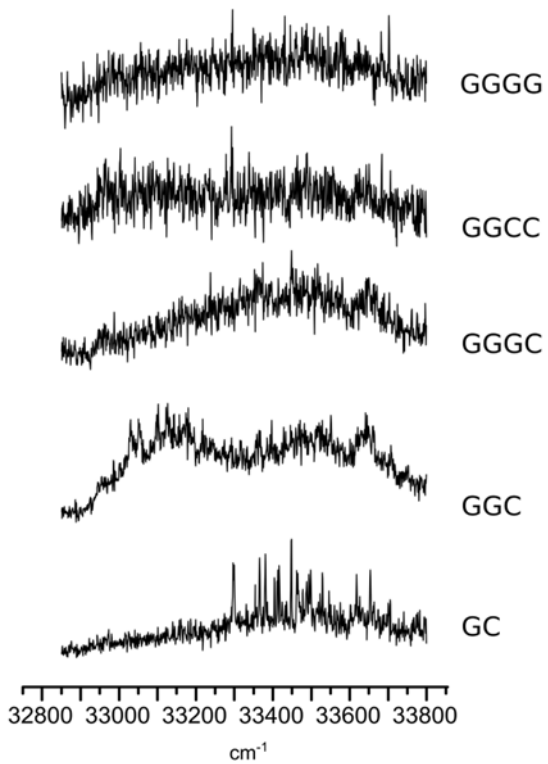


Figure 7.2 *REMPI spectra of guanine – cytosine (GC) dimer, GGC trimer and GGGC, GGCC and GGGG tetramers, scanned at 32900-33800* cm^{-1}.

Finally, two more DNA bases, adenine and thymine, are also biologically relevant. Despite no experiments have been presented in this work for these bases, the experience obtained with cytosine and guanine makes those systems an exciting new pathway to be studied.

From the theoretical point of view, the increase of the aggregate has been already addressed for keto cytosine and 9-keto guanine. The pentamers of cytosine and tetramers of guanine has also been calculated at M06-2X/6-31+G(d) level. Some relevant structures of both systems may be found in **Figure 7.3**.

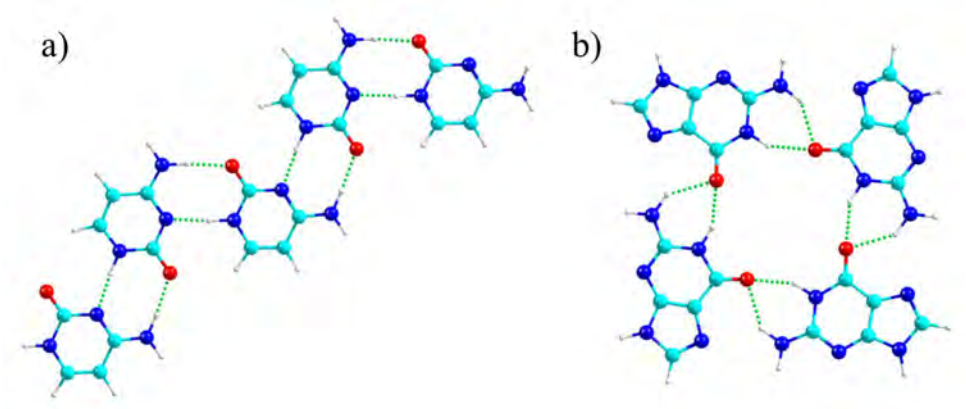


Figure 7.3 *Relevant structure of the pentamers of cytosine (a) and tetramers of guanine (b) found in the conformational analysis.*

On the other hand, the conformational analysis of guanine – cytosine aggregates has also been started and the structure of the 10 different tautomeric dimers is in progress. With the results of the dimers and the tautomeric selectivity, upper aggregates should be carried out.

Finally, the substituted amino acids pointed to the early stacking behavior compared to their nude precursors. Thus, analyzing the conformational behavior of aggregation of nucleosides in comparison with their methylated forms can also be an interesting study to be approached.

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

Este trabajo no habría sido posible sin la dedicación de mis dos directores: Jose Andrés Fernández y Judith Millán. Gracias Jose Andrés por acoger a ese chaval logroñés en tu grupo de investigación y por tus innumerables consejos y recomendaciones, siempre para mejorar aunque no lo percibiera en ese momento. Y a ti, Judith, por transmitirme todos y cada uno de los conocimientos de química computacional que tengo, por las horas que has invertido en mi desarrollo y por estar ahí siempre que lo he necesitado.

Me gustaría agradecer a los Drs. Asier Longarte, Rodrigo Martínez, Iker León y Raúl Montero, personas a las que realmente admiro, por toda la ayuda que me habéis ofrecido a lo largo de esta etapa y por permitirme adquirir un pedazo de toda la sabiduría que desprendéis.

Por supuesto, he de agradecer a mis dos compañeros Imanol Usabiaga y Pedro Felipe Arnáiz, dos personas excepcionales que me han brindado su ayuda siempre que lo he necesitado y con los que he compartido risas, discusiones y grandes momentos. Y al resto de miembros del grupo de Espectroscopía y Espectrometría de masas: Carol, Francisco, Rafa, Virginia, Montse, Iciar, Emilio, Estibaliz, Marta, Rober, Jone, Ander, Aran… por hacer que mi estancia en Bilbao haya sido maravillosa y por permitirme disfrutar de vuestra amistad. De la Universidad de la Rioja me gustaría agradecer a las Drs. Purificación Fernandez, Jose Miguel Avizcuri y Luis Vaquero por sus consejos y los grandes momentos vividos a lo largo de este trabajo.

Por otro lado, he de agradecer a la Dr. Anouk Rijs por acogerme durante tres meses en su grupo de investigación y a Sjors Bakkels y Daniël Bakker por hacer de la estancia en el FELIX una experiencia inolvidable. Y a la Dr. Julia Contreras por enseñarme todos los entresijos del NCIplot.

Además, he de agradecer a los Drs. Jose María Mercero y Edu Ogando, por el inestimable apoyo del Servicio de Computación (IZO/SGIker) de la UPV/EHU y resolver las n-mil dudas y problemas surgidos, siempre diligentemente y con una educación excepcional.

Y por supuesto he de dedicárselo a mi familia. Esta tesis es vuestra. Porque habéis estado desde el primer minuto apoyándome en los días buenos y sobre todo, en los malos. Porque me habéis educado desde pequeño en que la perseverancia y la tenacidad dan sus frutos y este trabajo es el vivo retrato de ello. Y a ti, Carla, por estar a mi lado en los buenos y malos momentos y haber sido un apoyo vital, brindándome con mucha paciencia, todo tu amor y cariño. Por todo ello, muchas gracias.

Finalmente, me gustaría agradecer a la Universidad del País Vasco (UPV/EHU) por la beca de investigación predoctoral, al servicio de Computación (SGI/IZO) de la UPV/EHU por sus recursos humanos y computacionales y a todos los miembros del laboratorio FELIX por su trabajo y sus instalaciones.

CHAPTER 8

APPENDIX

8.1 Amino acids and DNA bases

Figure 8.1.1 *Structure of the 11 most stable structures of aA at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

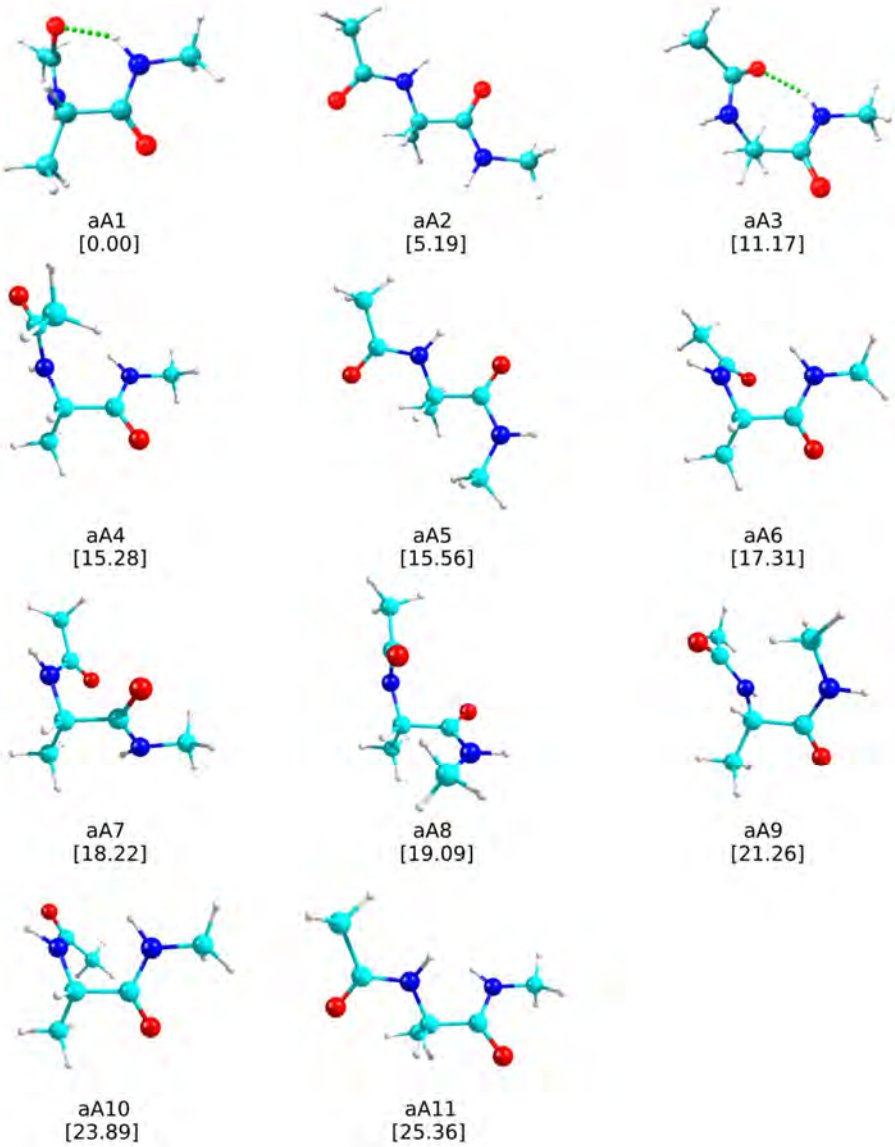


Figure 8.1.2 *Structure of the 15 most stable structures of aI at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

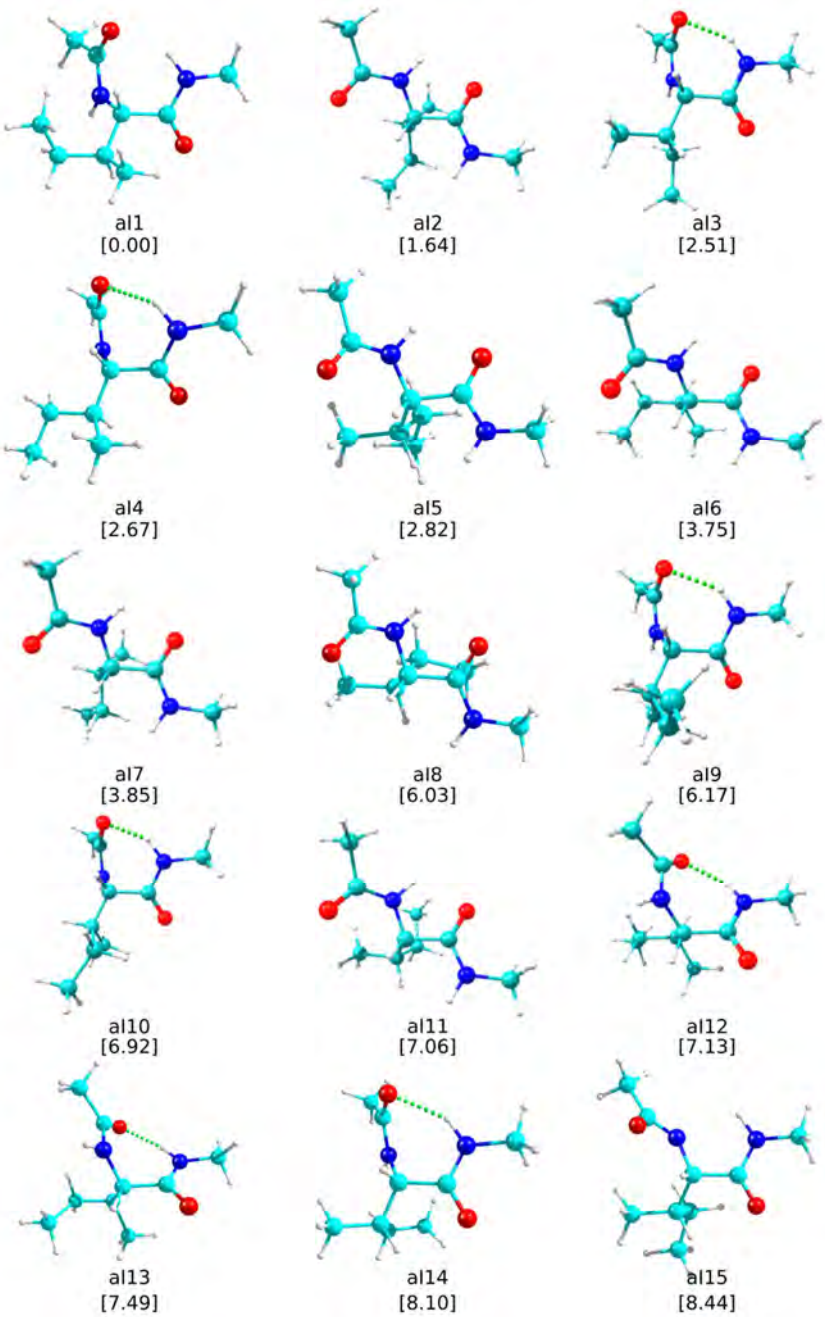


Figure 8.1.3 *Structure of the 15 most stable structures of aV at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

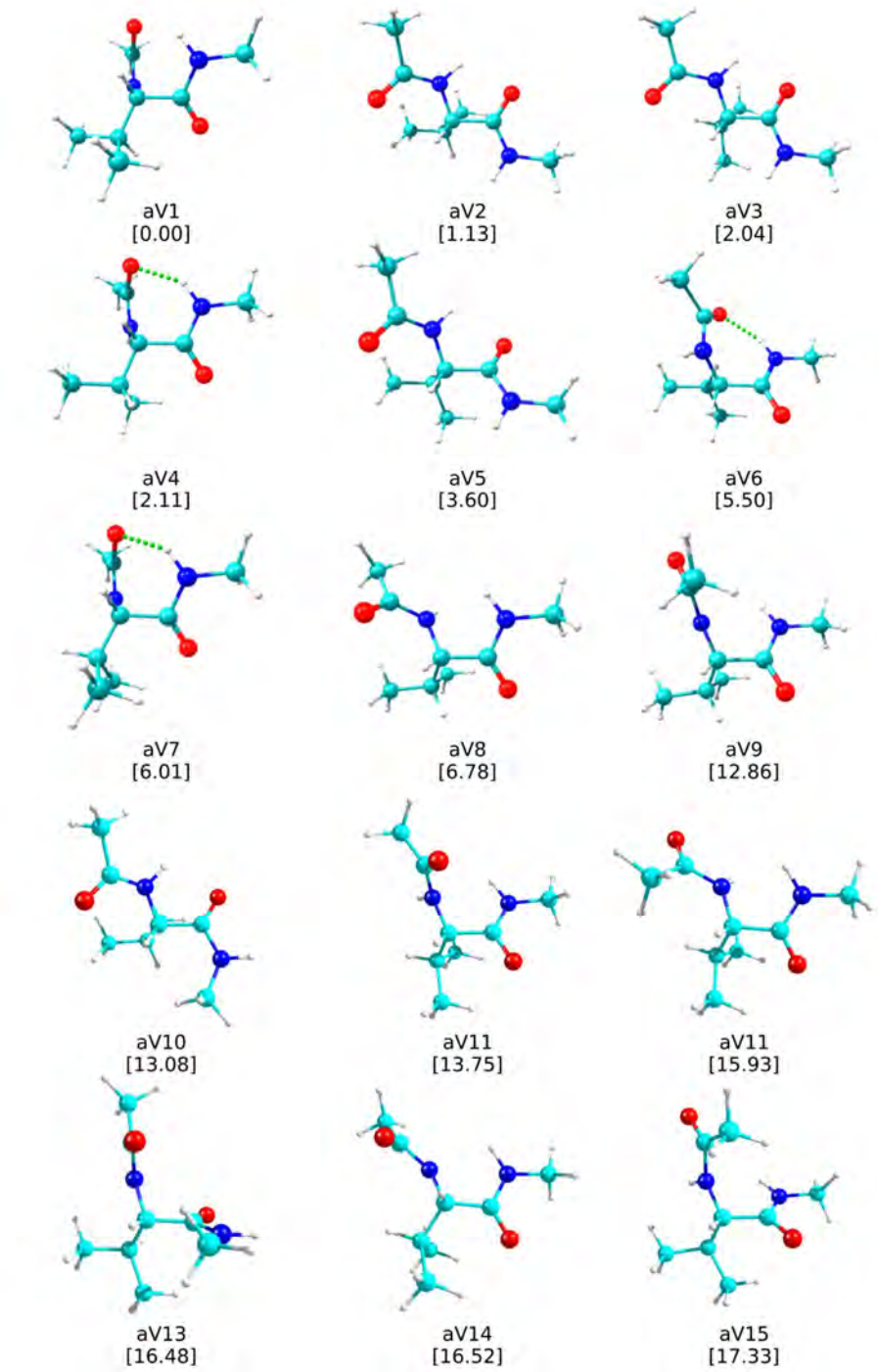


Figure 8.1.4 *Structure of the 15 most stable structures of aN at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

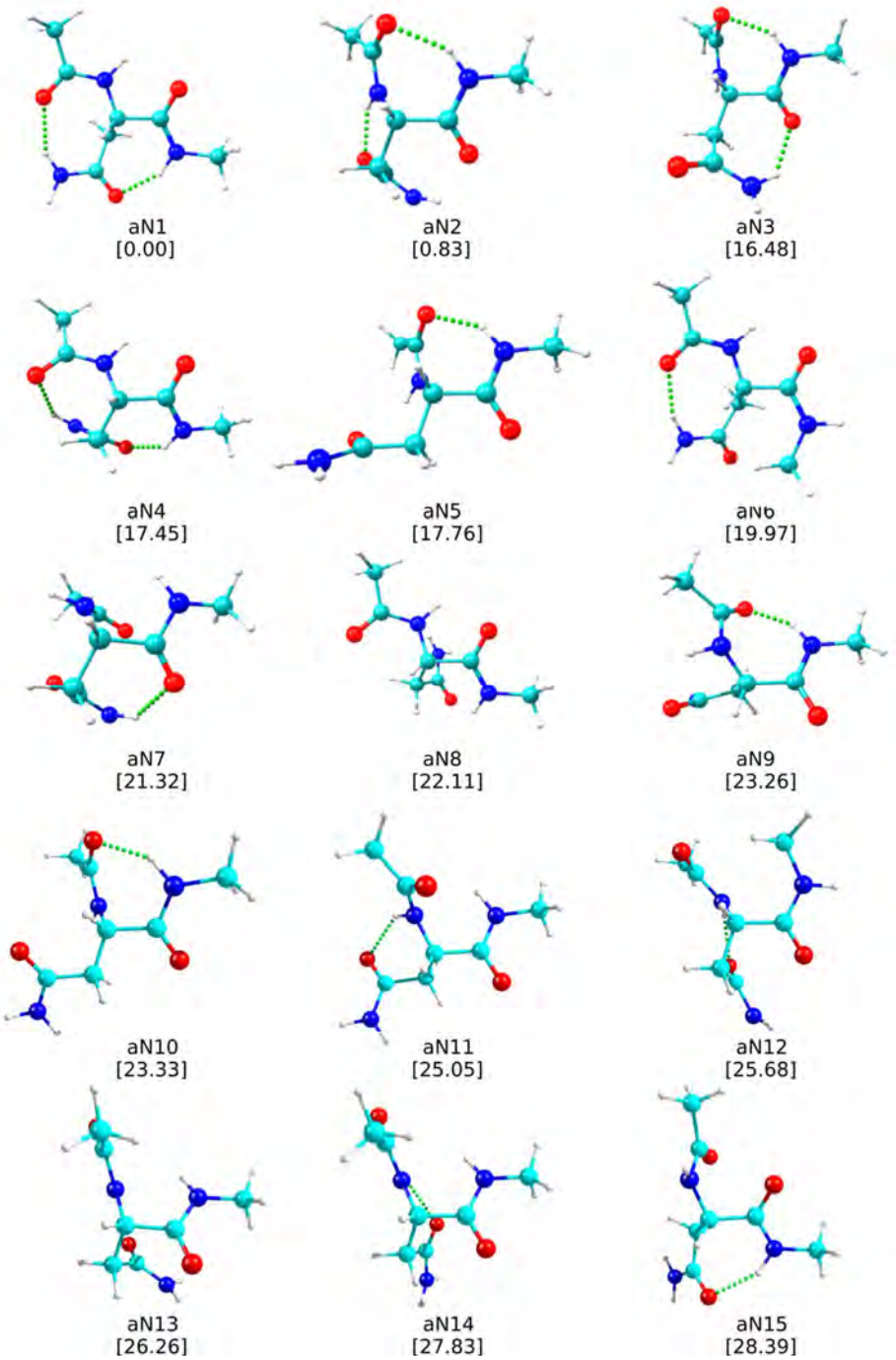


Figure 8.1.5 *Structure of the 15 most stable structures of aQ at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

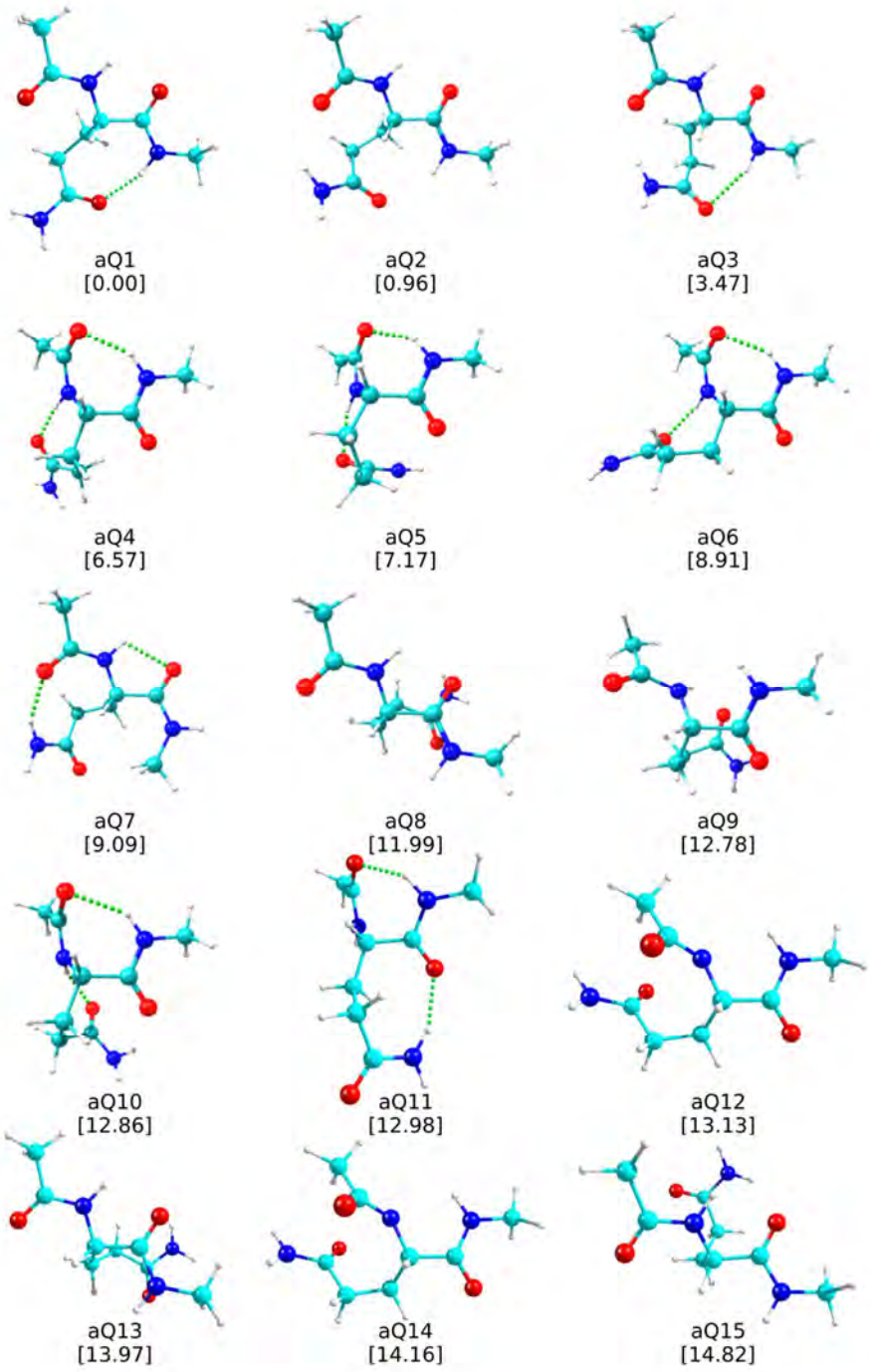


Figure 8.1.6 *Structure of the 15 most stable structures of aR at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol-1) in brackets. ZPE correction was applied to the energy values.*

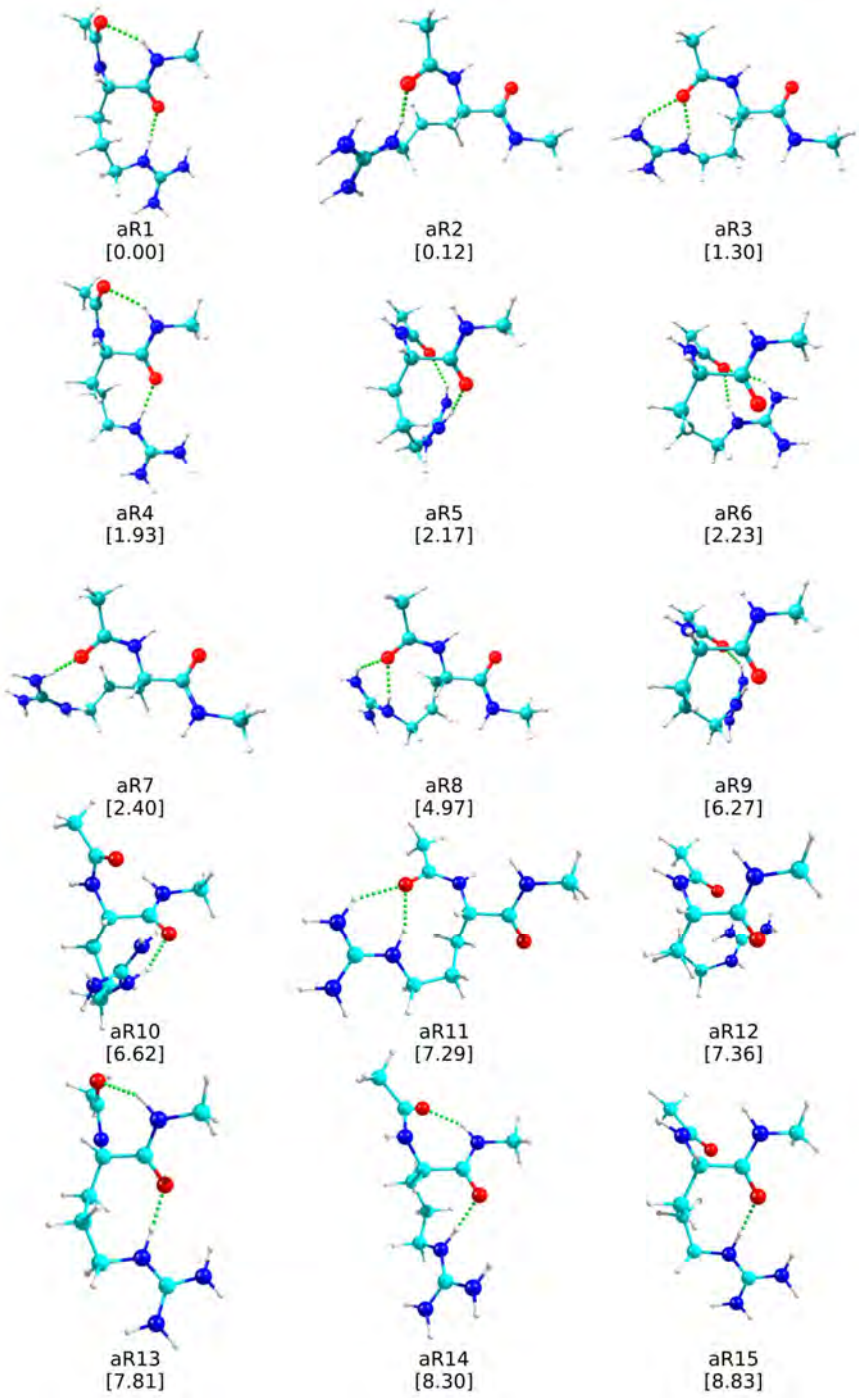


Figure 8.1.7 *Structure of the 15 most stable structures of aF at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

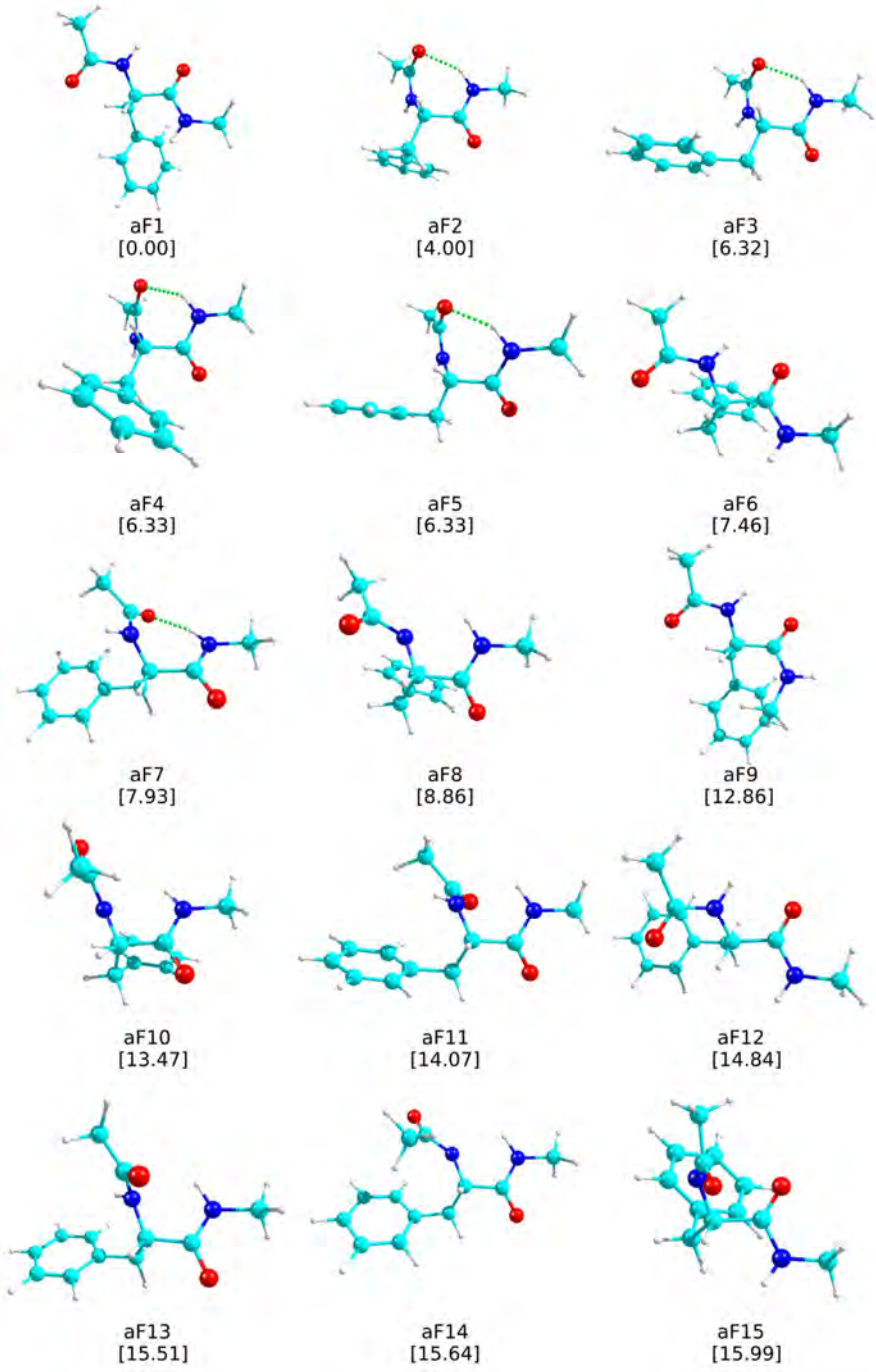


Figure 8.1.8 *Structure of the 15 most stable structures of aW at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

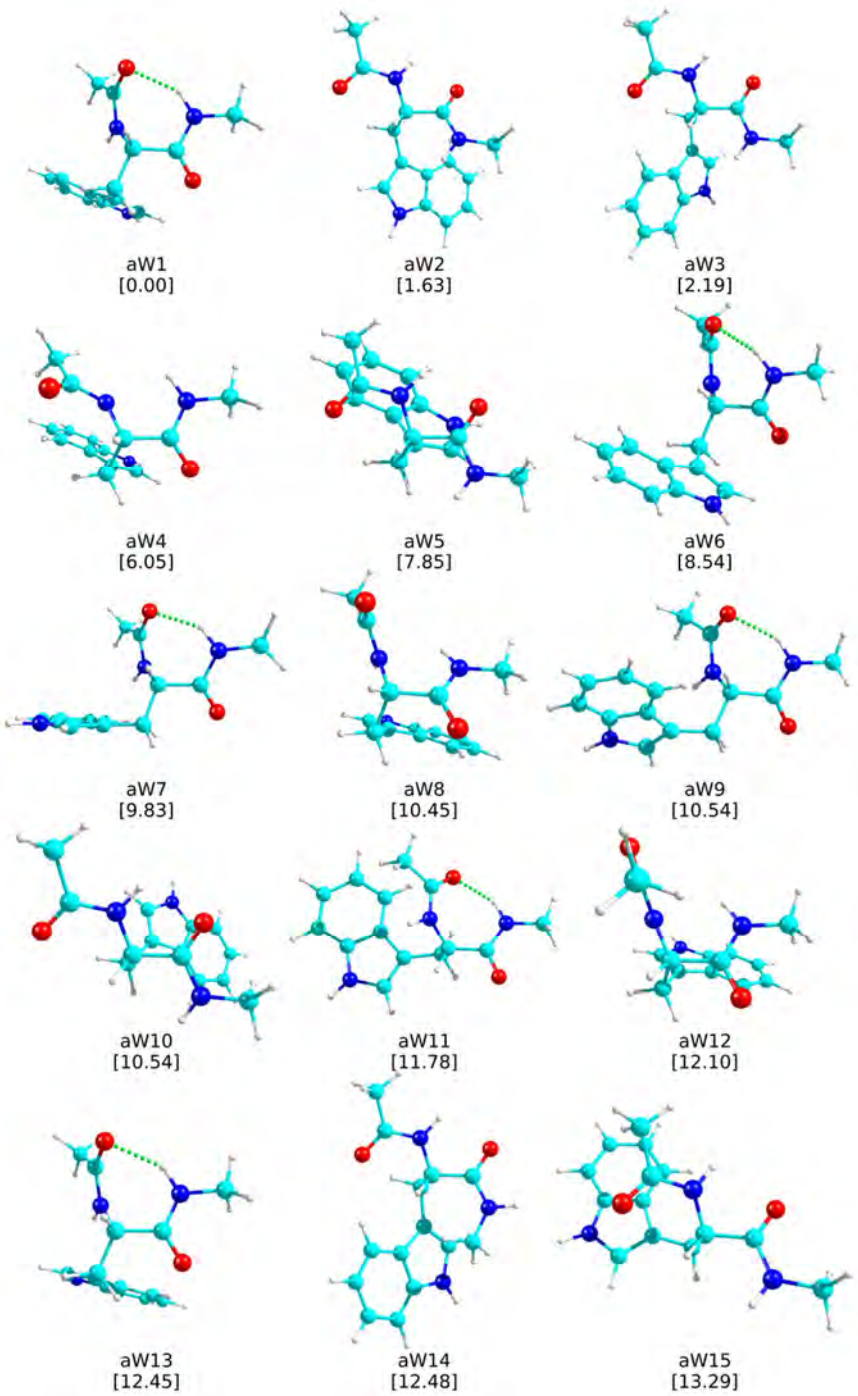


Figure 8.1.9 *Structure of the 15 most stable structures of aY at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

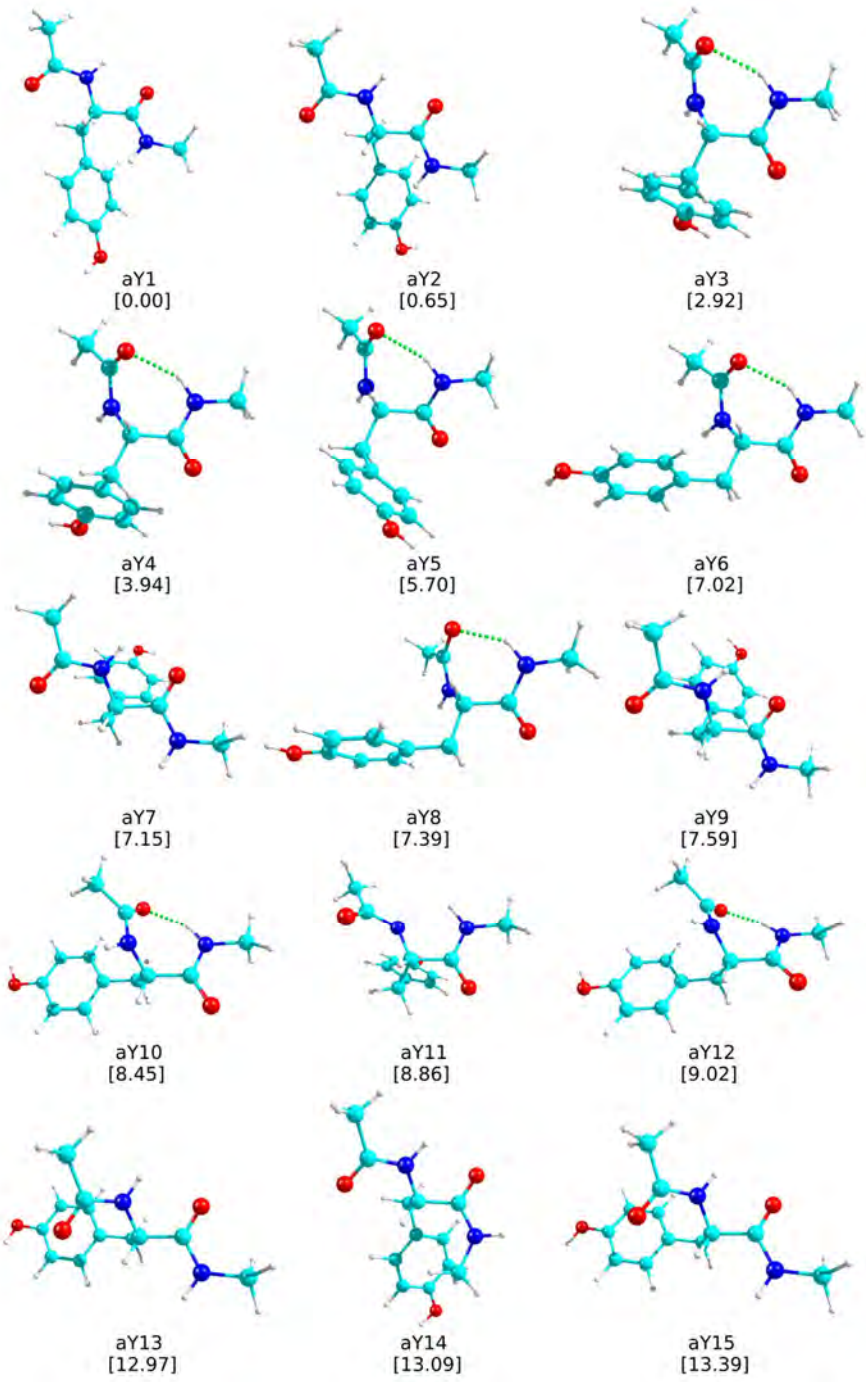


Figure 8.1.10 *2DNCI plots (s vs. ρ*sign(λ_2)) for the two most stable conformations of aA, aI and aV.*

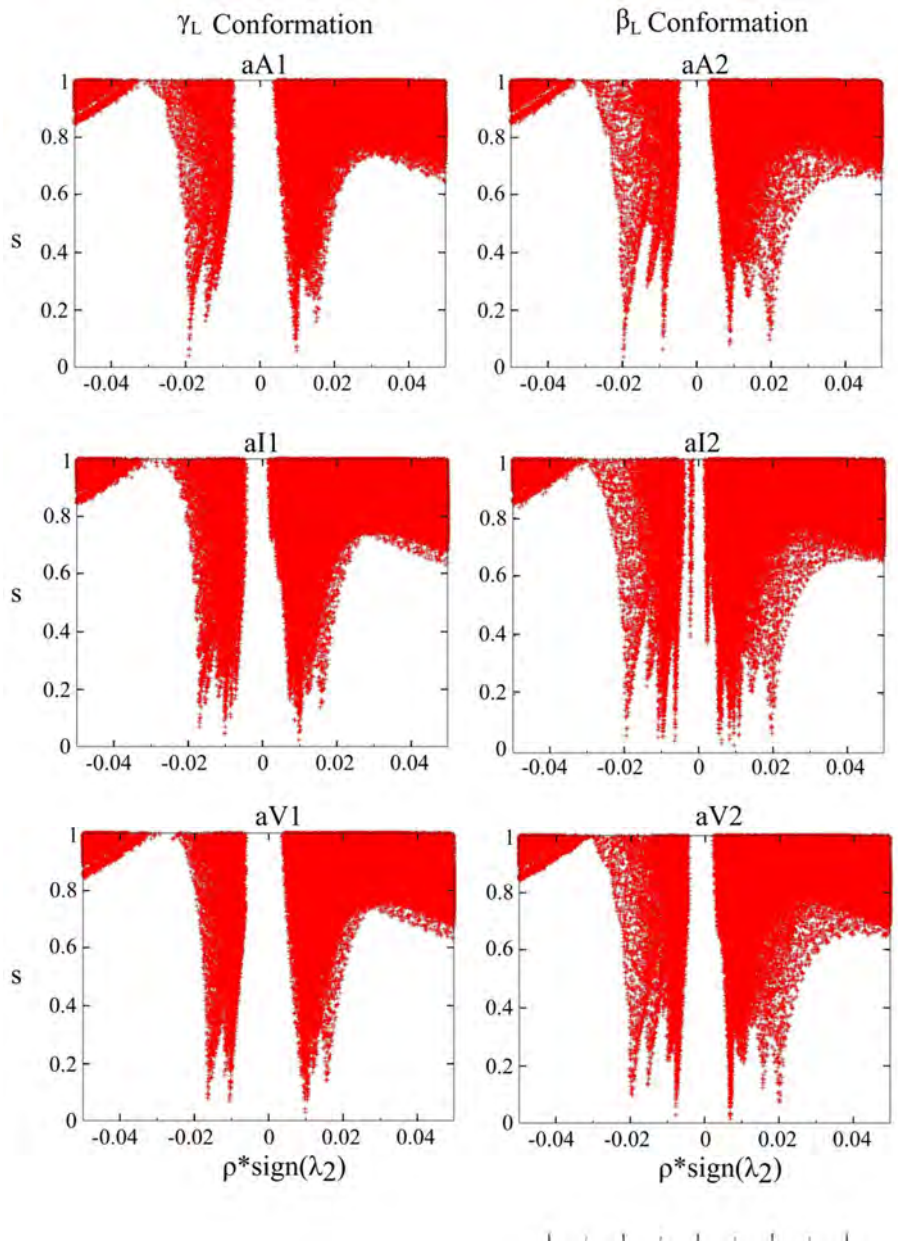


Figure 8.1.11 *2D-NCI plots (s vs. ρ*sign(λ2)) for the two most stable conformations of aN, aQ and aR.*

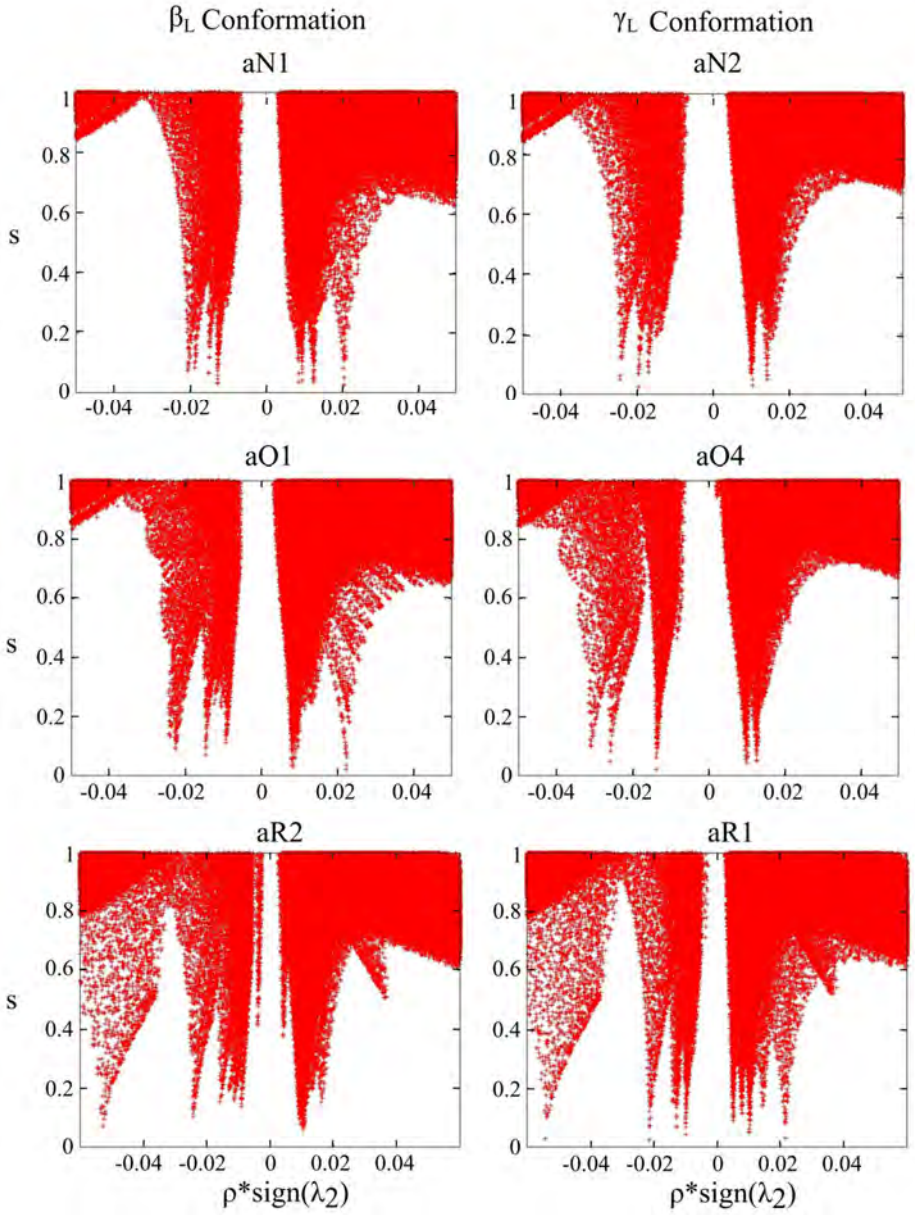
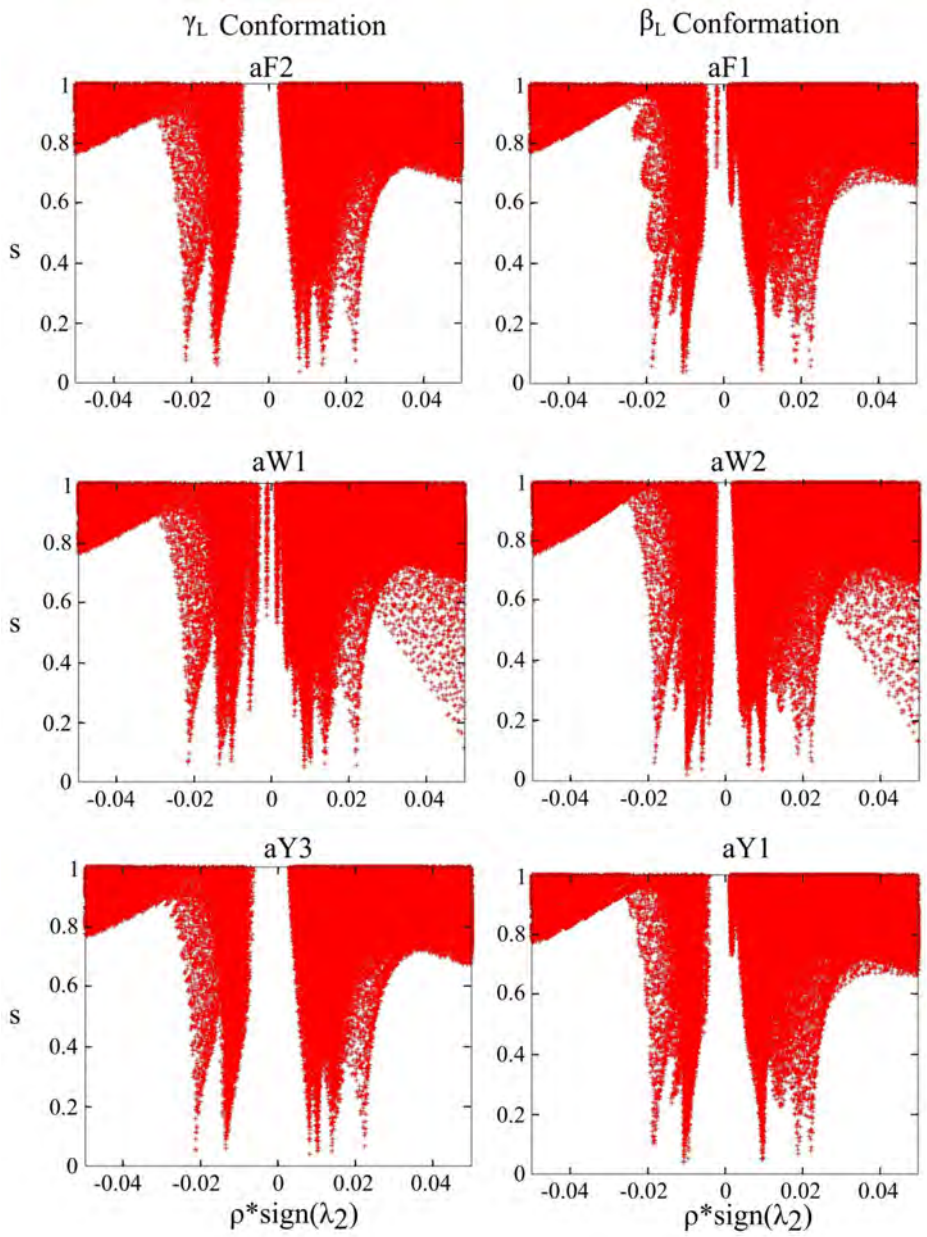
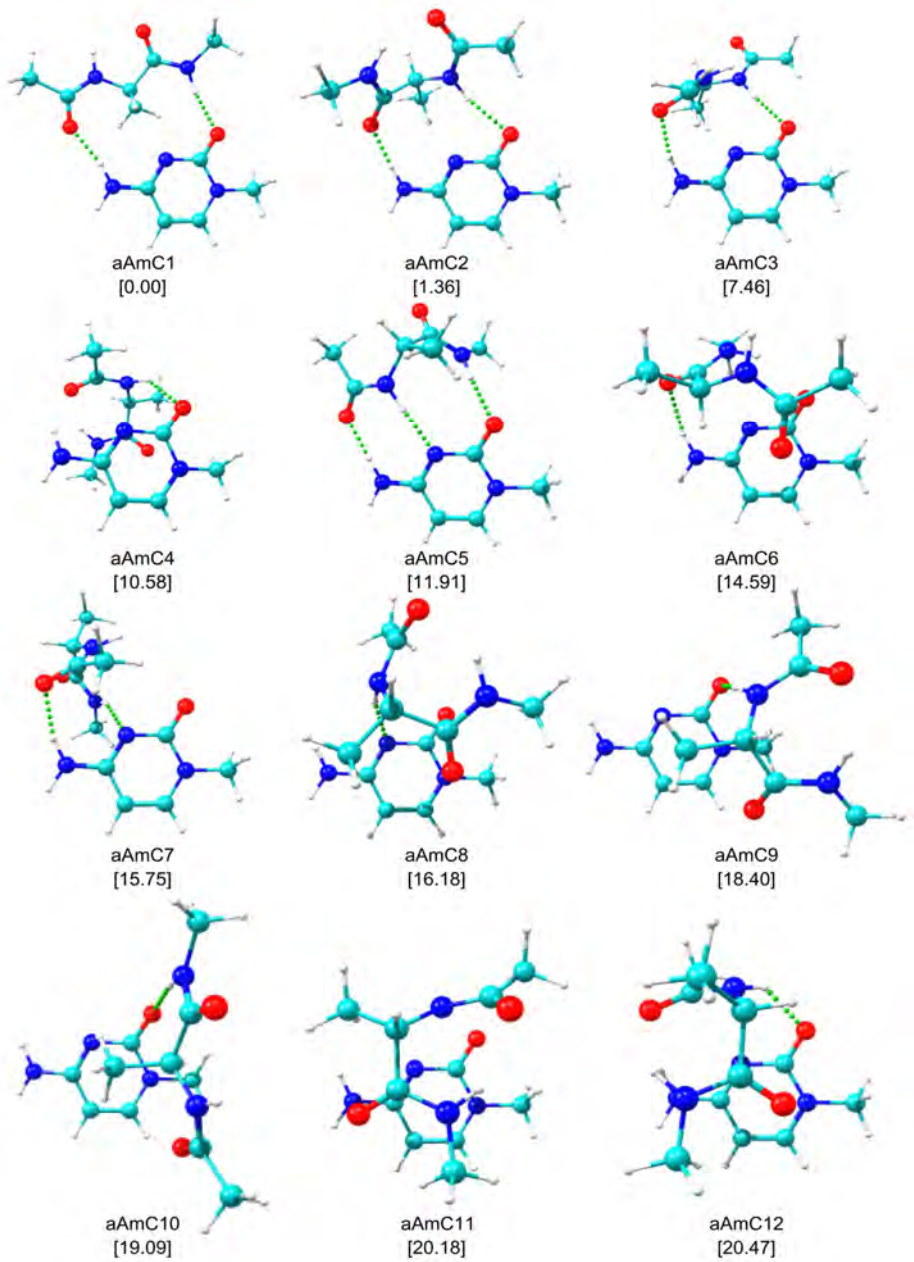


Figure 8.1.12 *2D-NCI plots (s vs. ρ*sign(λ2)) for the two most stable conformations of aF, aW and aY.*

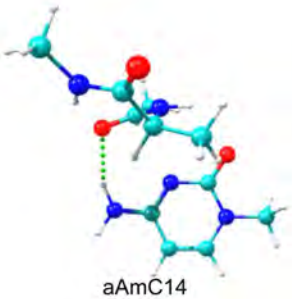


8.2 DNA – protein aggregation

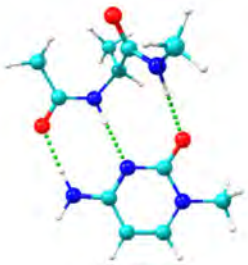
Figure 8.2.1 *The first 20 calculated structures for aAmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



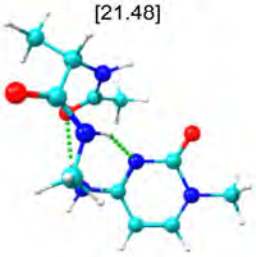
aAmC13
[21.48]



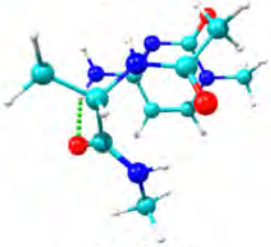
aAmC14
[22.24]



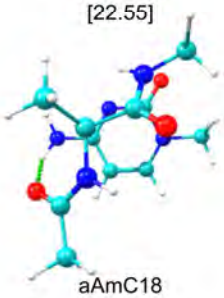
aAmC15
[22.55]



aAmC16
[22.80]

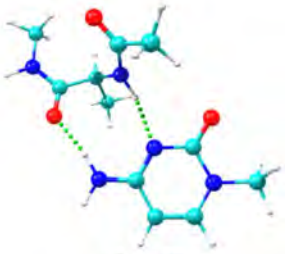


aAmC17
[23.07]



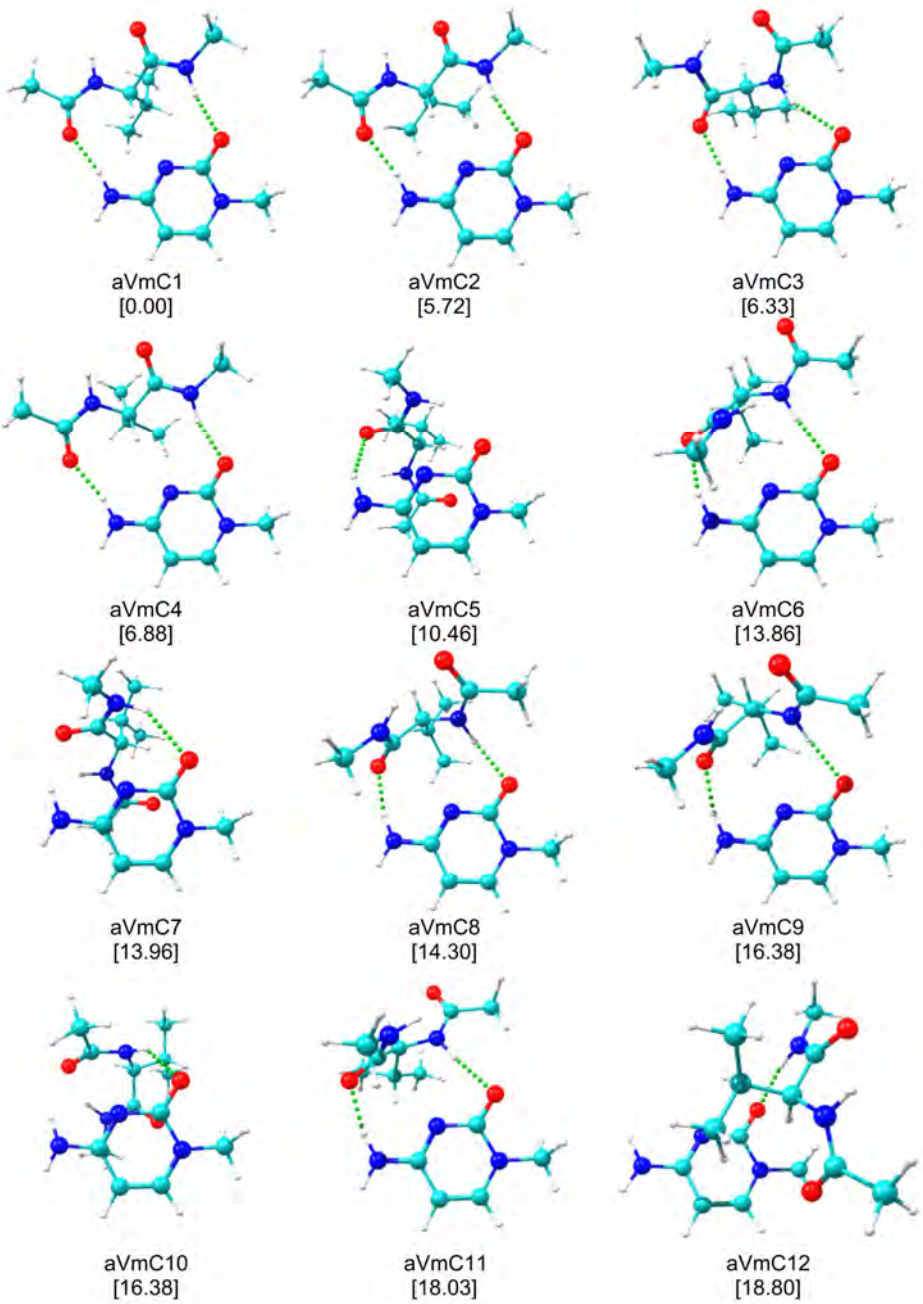
aAmC18
[23.17]

aAmC19
[24.18]



aAmC20
[26.22]

Figure 8.2.2 *The first 20 calculated structures for aVmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



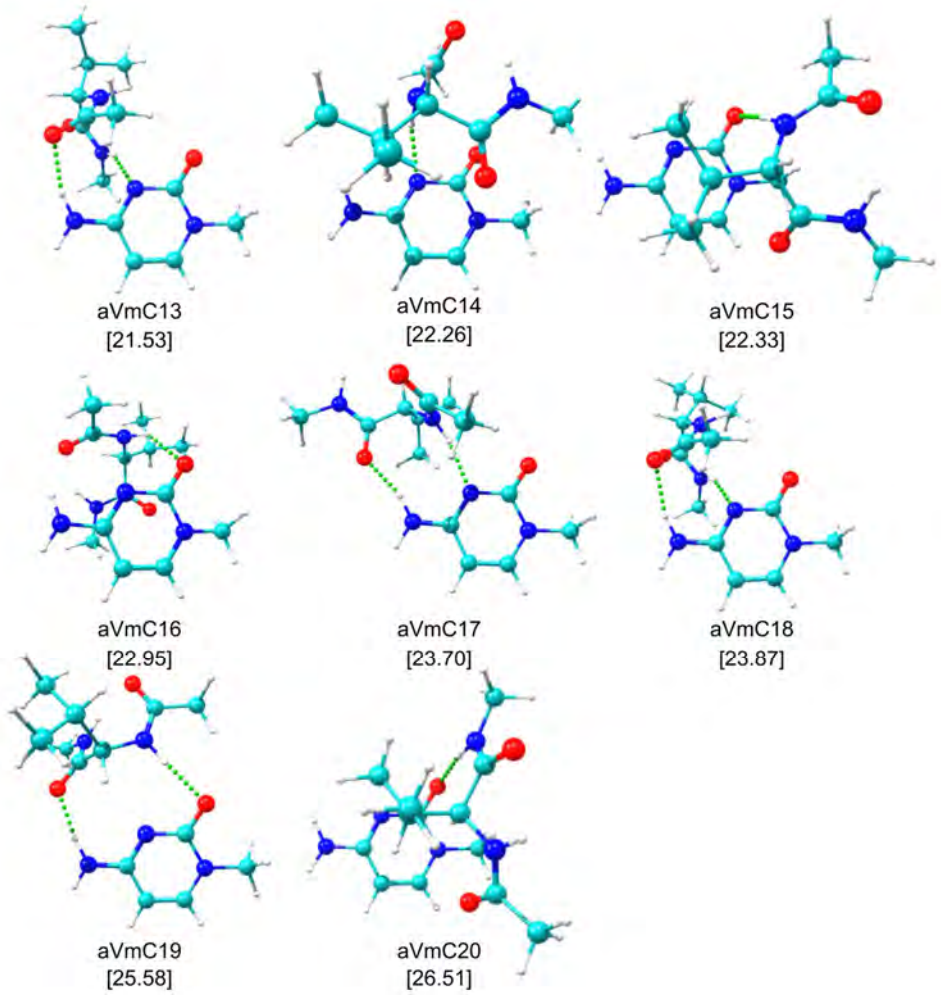
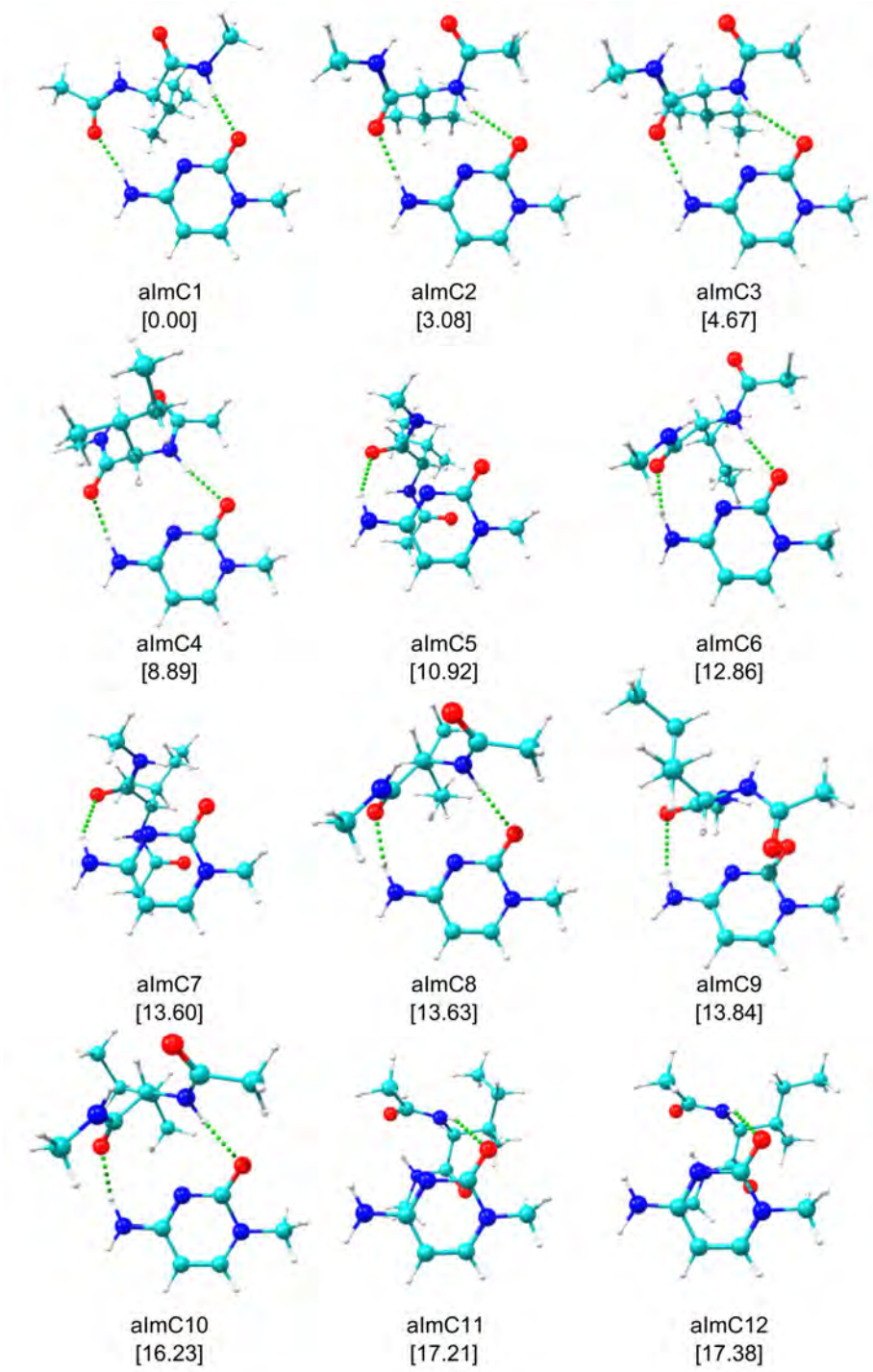
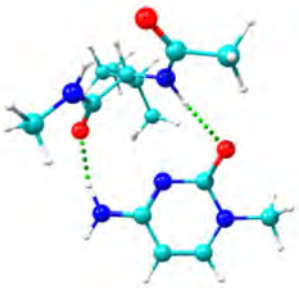
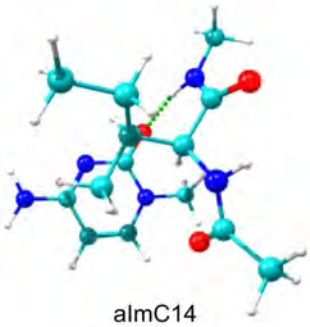

aVmC13
[21.53]
aVmC14
[22.26]
aVmC15
[22.33]
aVmC16
[22.95]
aVmC17
[23.70]
aVmC18
[23.87]
aVmC19
[25.58]
aVmC20
[26.51]

Figure 8.2.3 *The first 20 calculated structures for aImC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

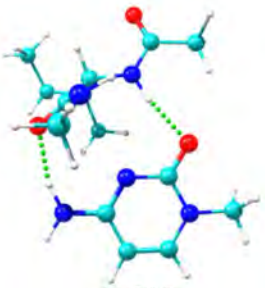




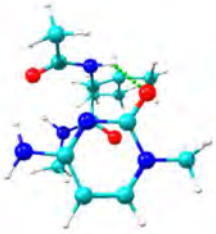
almC13
[17.60]



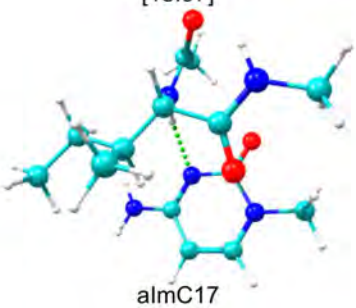
almC14
[18.97]



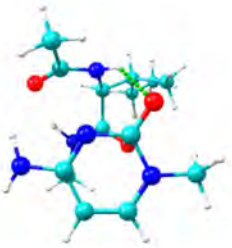
almC15
[19.34]



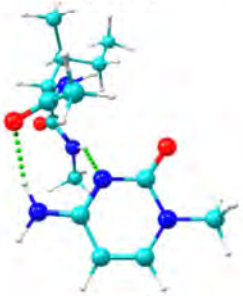
almC16
[19.66]



almC17
[19.70]



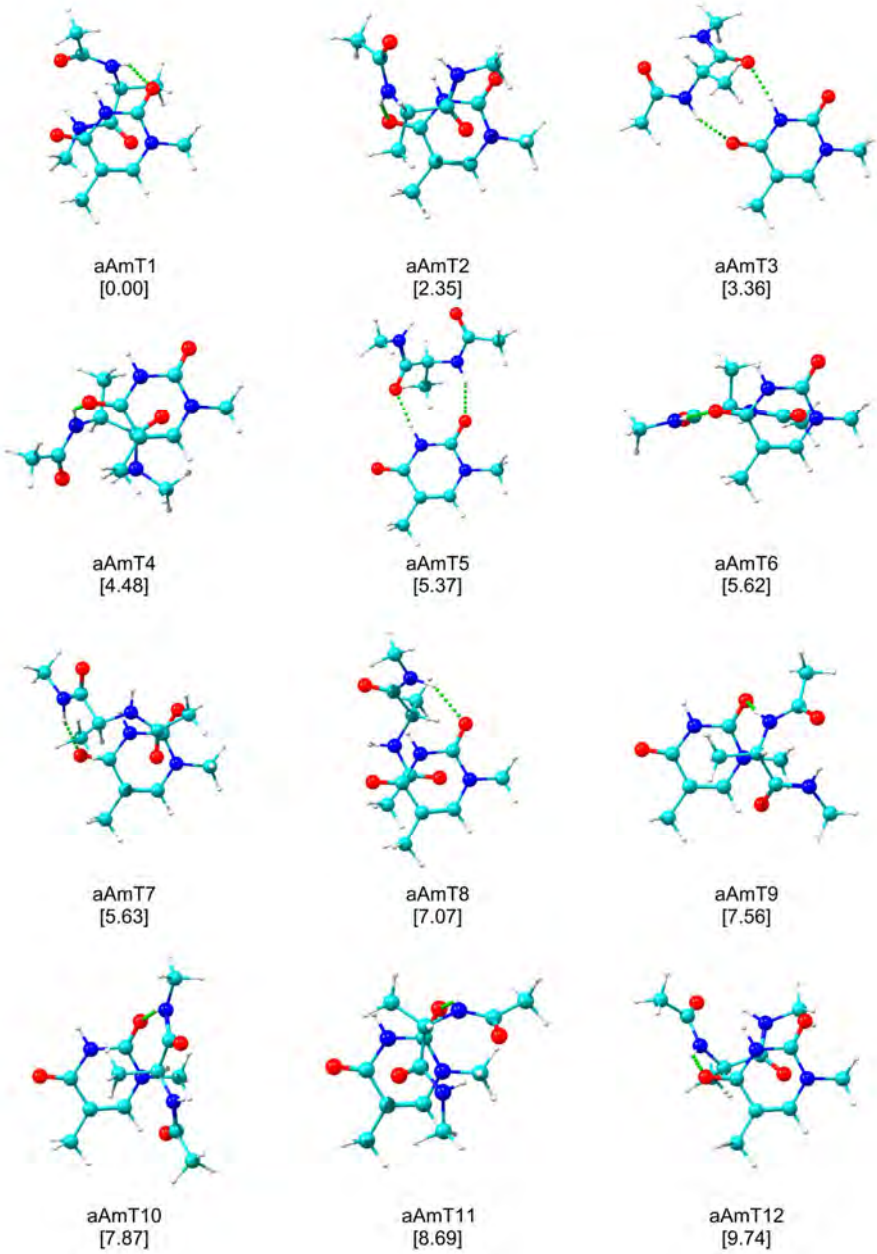
almC18
[20.38]



almC19
[20.43]

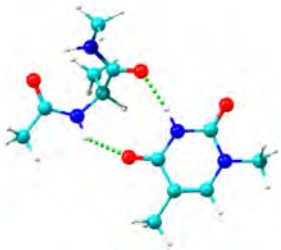
almC20
[21.06]

Figure 8.2.4 *The first 20 calculated structures for aAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

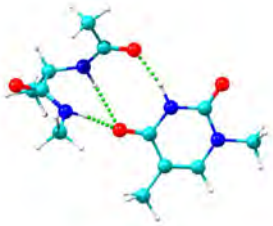


aAmT13
[11.44]

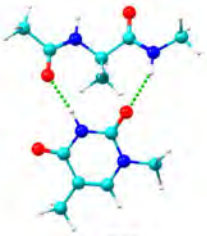
aAmT14
[11.50]



aAmT15
[13.53]

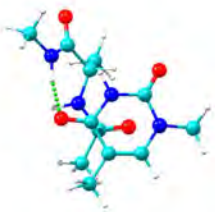


aAmT16
[14.50]



aAmT17
[14.75]

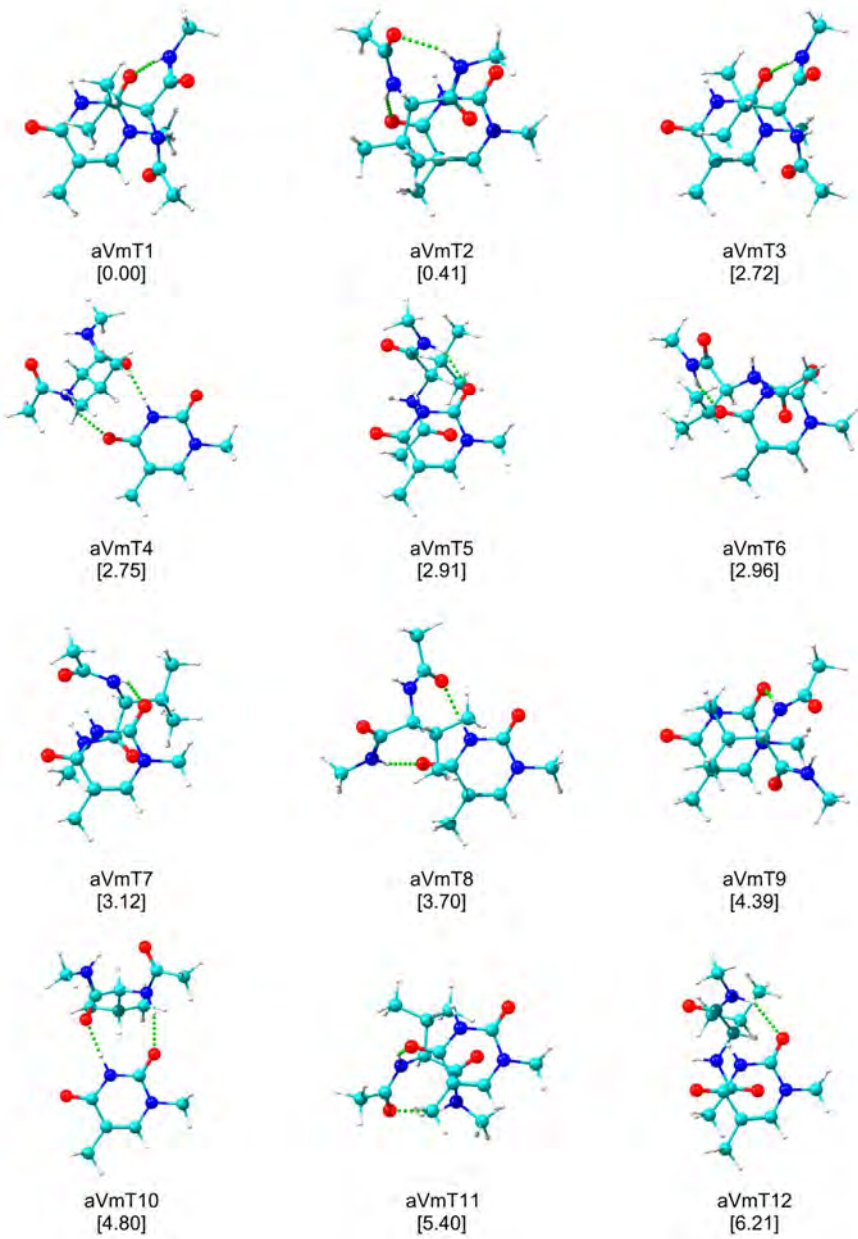
aAmT18
[14.92]



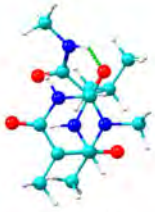
aAmT19
[15.01]

aAmT20
[15.15]

Figure 8.2.5 *The first 20 calculated structures for aVmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

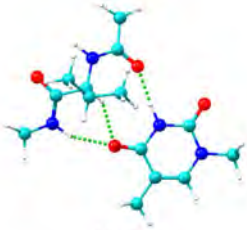


aVmT13
[6.77]

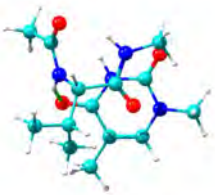


aVmT14
[6.91]

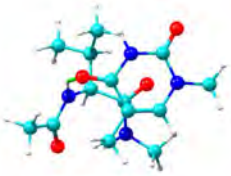
aVmT15
[7.13]



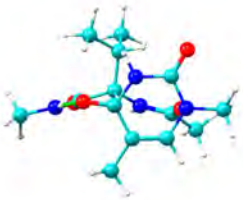
aVmT16
[7.24]



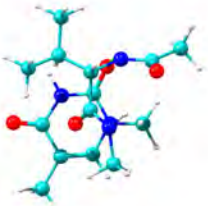
aVmT17
[7.97]



aVmT18
[8.02]

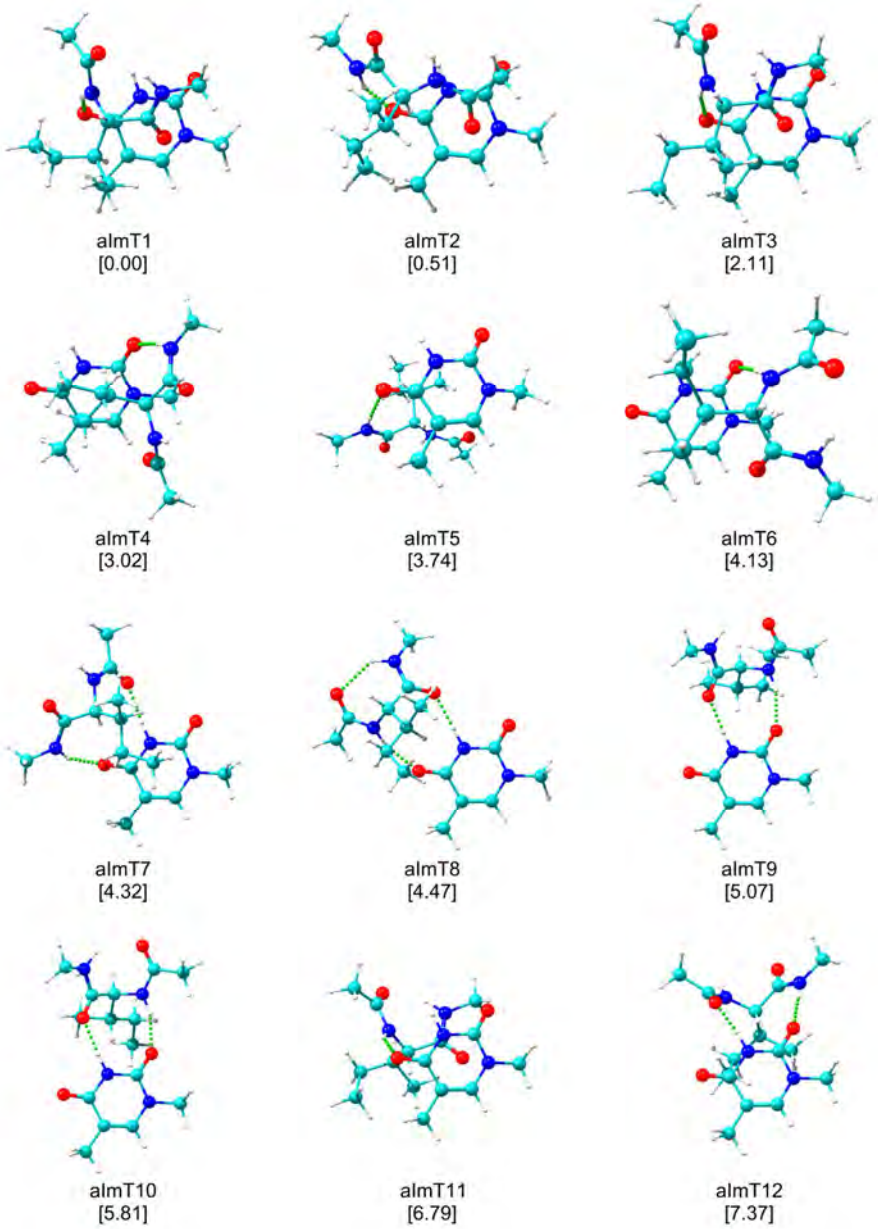


aVmT19
[10.15]

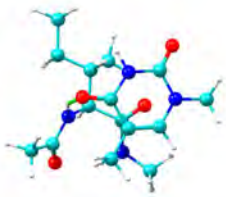


aVmT20
[10.16]

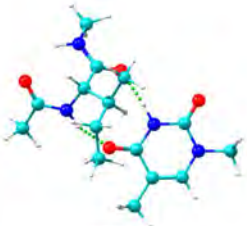
Figure 8.2.6 *The first 20 calculated structures for aImT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



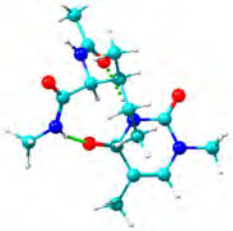
almT13
[8.28]



almT14
[8.71]



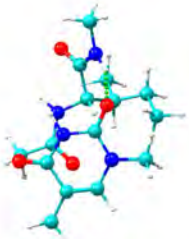
almT15
[8.82]



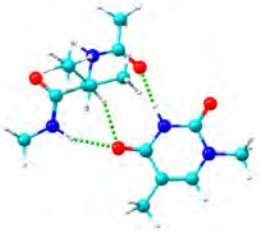
almT16
[8.87]

almT17
[8.90]

almT18
[9.47]

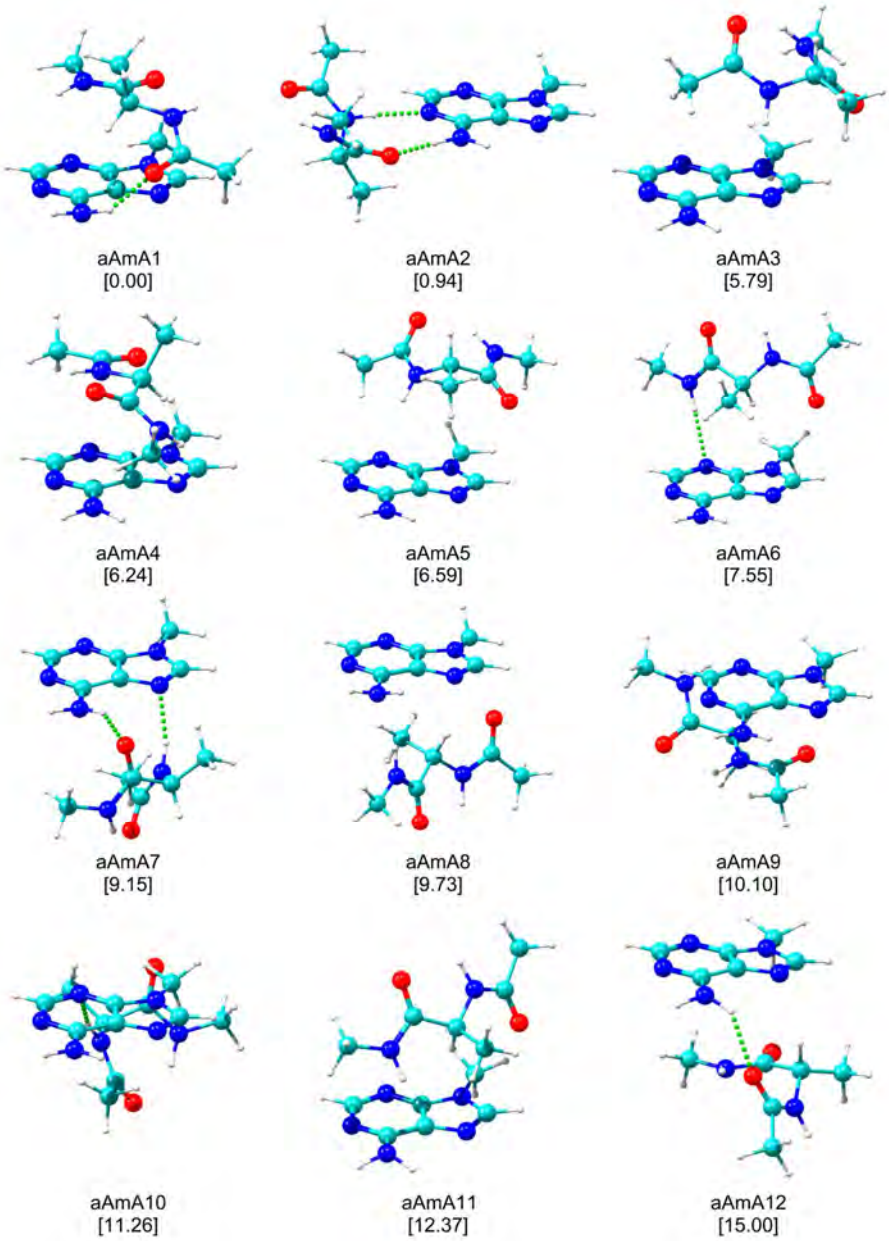


almT19
[9.64]

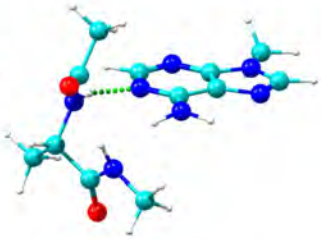


almT20
[9.89]

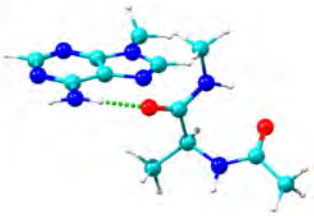
Figure 8.2.7 *The first 20 calculated structures for aAmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



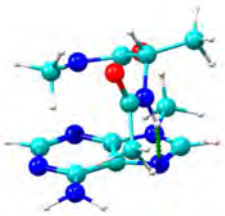
aAmA13
[15.22]



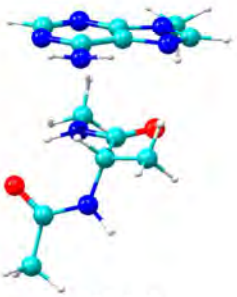
aAmA14
[17.13]



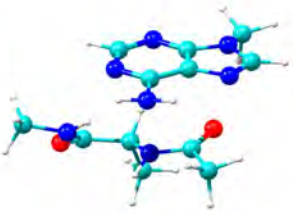
aAmA15
[18.60]



aAmA16
[18.85]

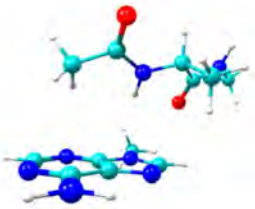


aAmA17
[19.46]



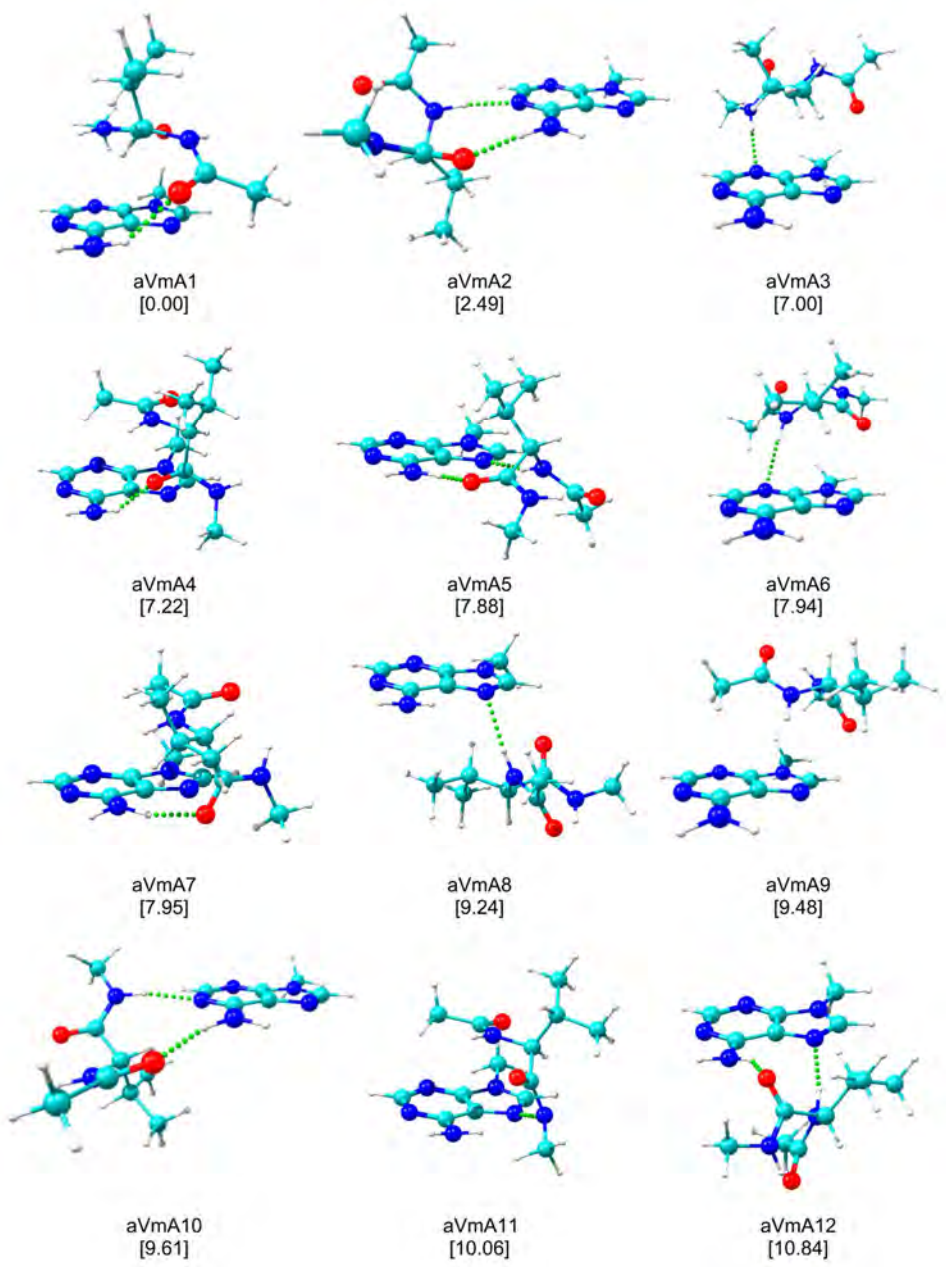
aAmA18
[21.04]

aAmA19
[22.32]



aAmA20
[22.63]

Figure 8.2.8 *The first 20 calculated structures for aVmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



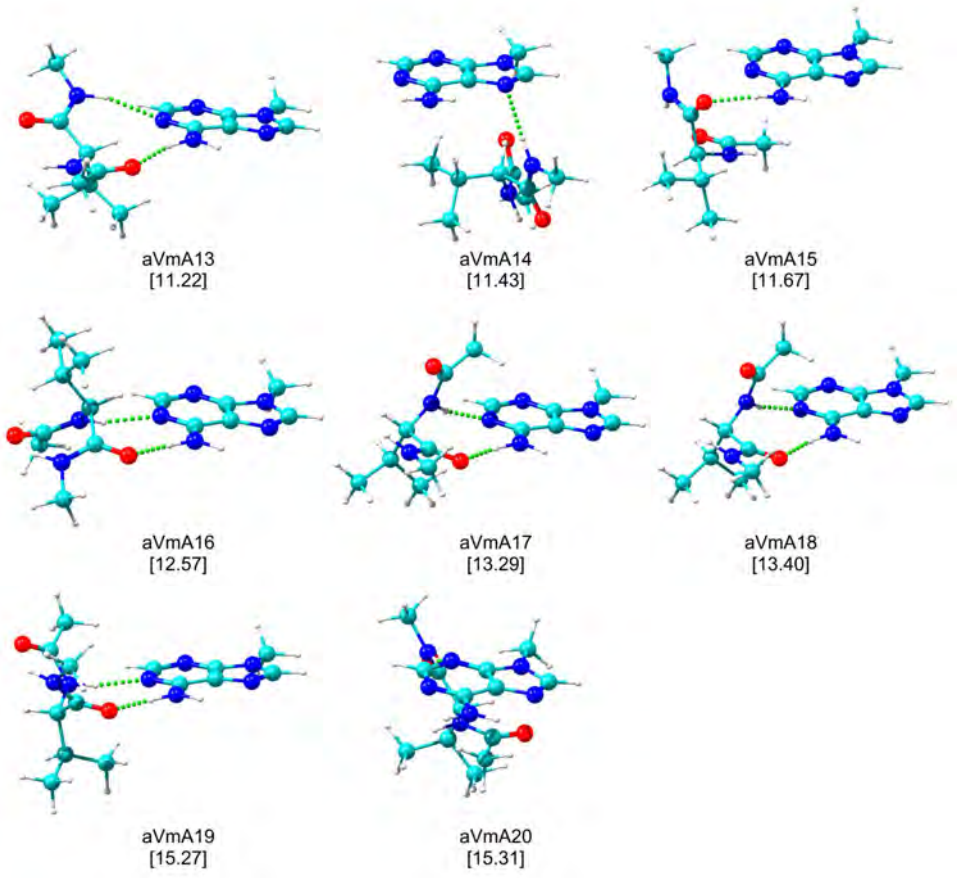
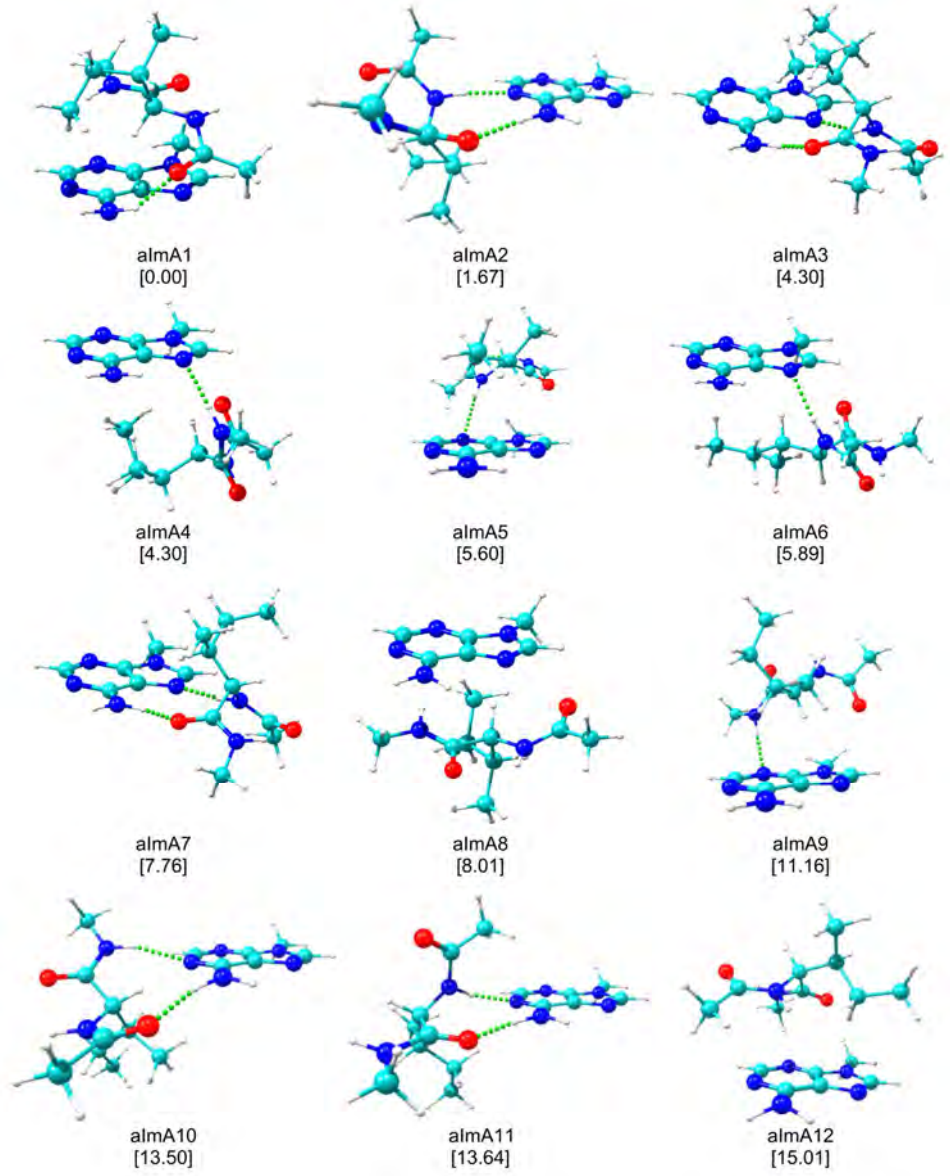
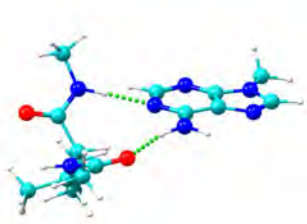
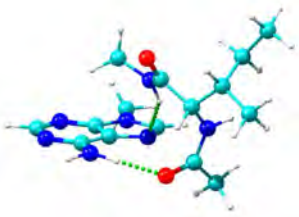

aVmA13
[11.22]
aVmA14
[11.43]
aVmA15
[11.67]
aVmA16
[12.57]
aVmA17
[13.29]
aVmA18
[13.40]
aVmA19
[15.27]
aVmA20
[15.31]

Figure 8.2.9 *The first 20 calculated structures for aImA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

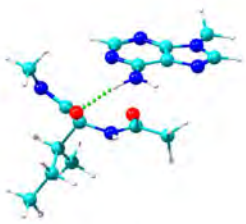




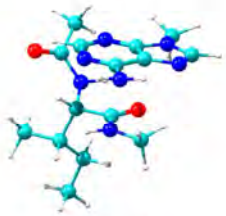
almA13
[15.88]



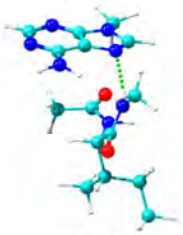
almA14
[16.05]



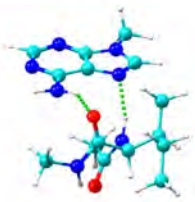
almA15
[17.50]



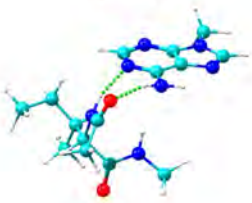
almA16
[17.77]



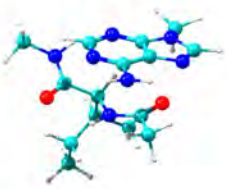
almA17
[18.69]



almA18
[20.04]

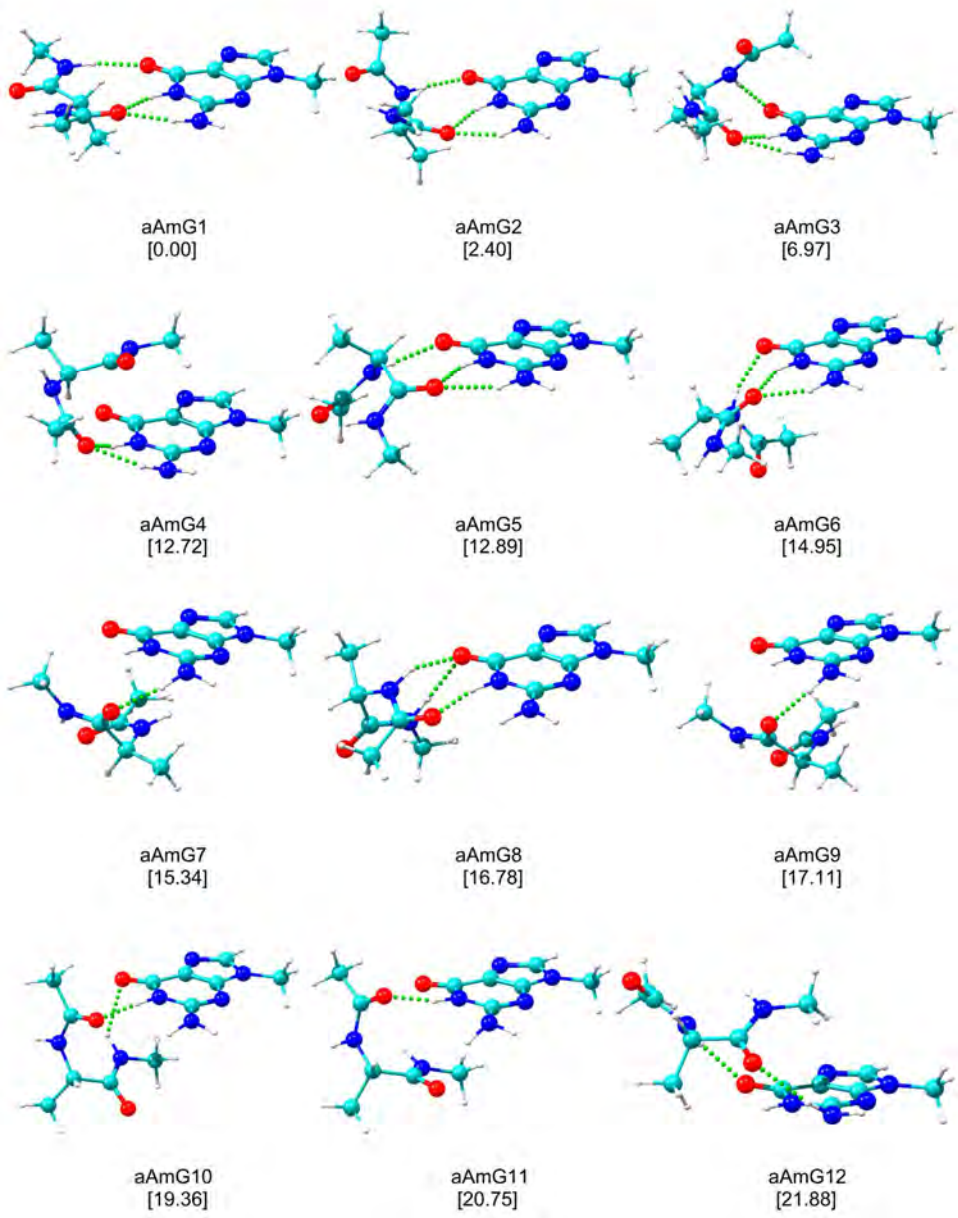


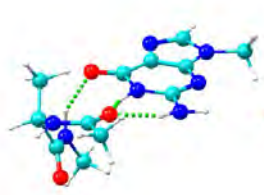
almA19
[20.92]



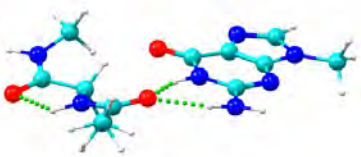
almA20
[21.20]

Figure 8.2.10 *The first 20 calculated structures for aAmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

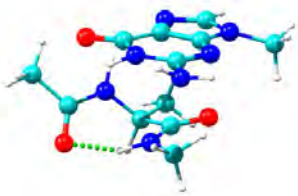




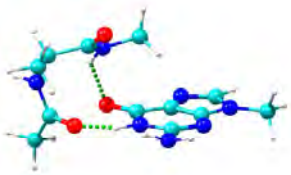
aAmG13
[22.20]



aAmG14
[23.35]



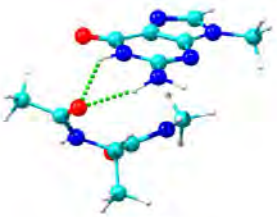
aAmG15
[24.37]



aAmG16
[24.51]

aAmG17
[25.94]

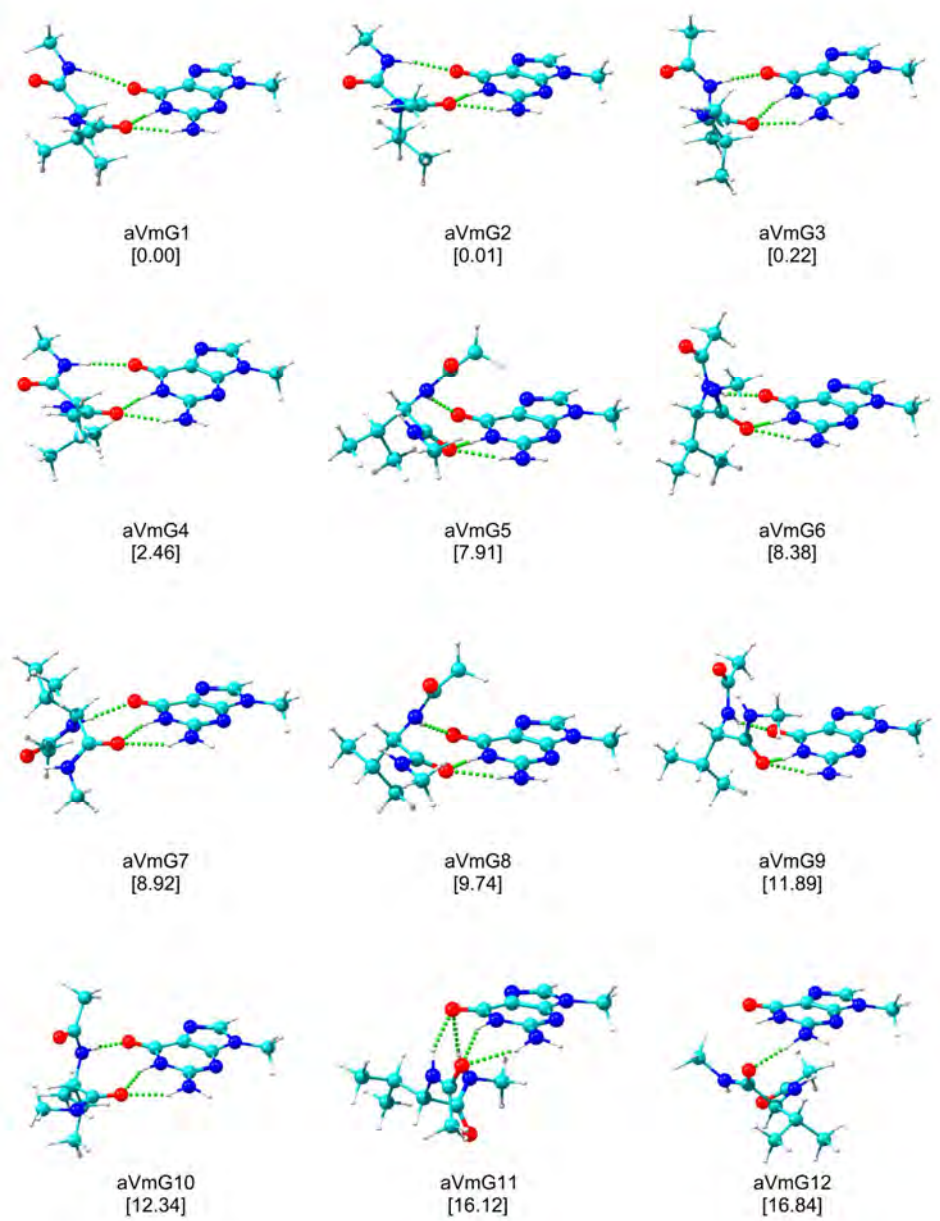
aAmG18
[26.58]

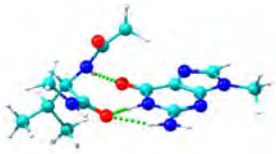


aAmG19
[27.70]

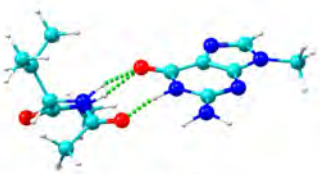
aAmG20
[29.55]

Figure 8.2.11 *The first 20 calculated structures for aVmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

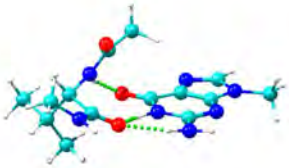




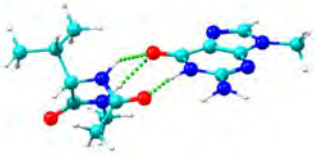
aVmG13
[16.93]



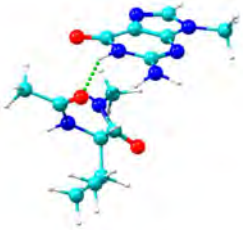
aVmG14
[17.21]



aVmG15
[17.45]

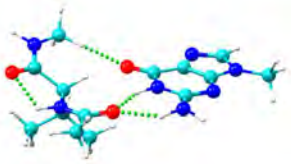


aVmG16
[19.34]

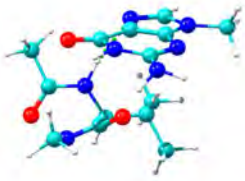


aVmG17
[19.96]

aVmG18
[21.40]

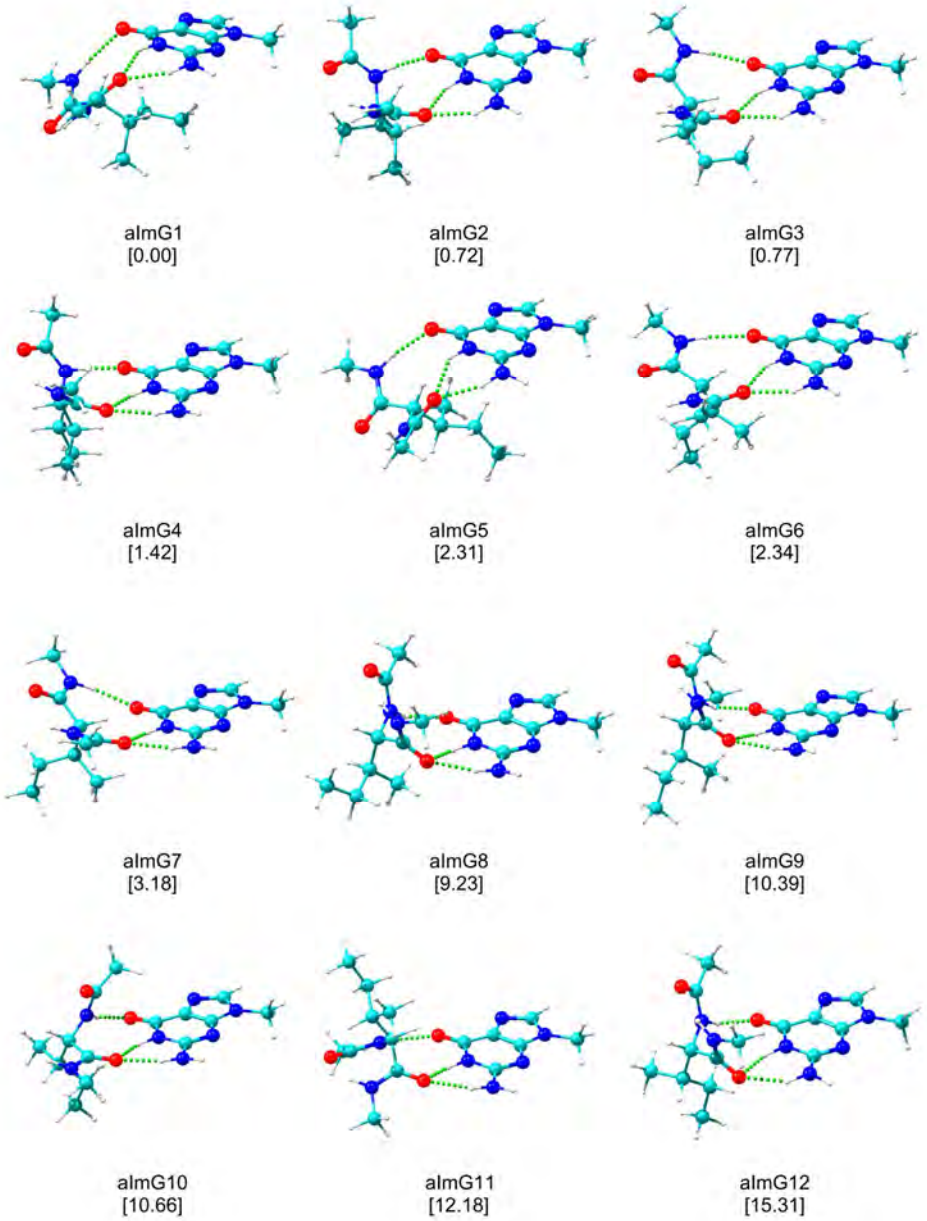


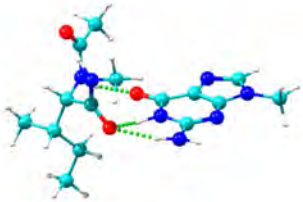
aVmG19
[21.70]



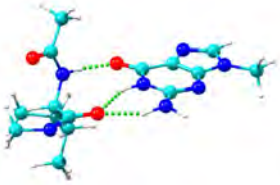
aVmG20
[24.69]

Figure 8.2.12 *The first 20 calculated structures for aImG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

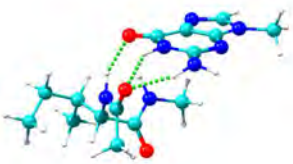




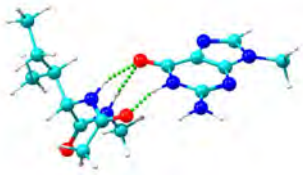
almG13
[15.65]



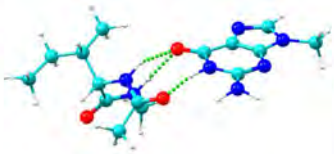
almG14
[15.81]



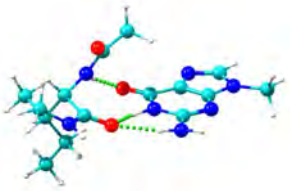
almG15
[17.05]



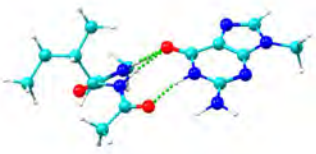
almG16
[17.76]



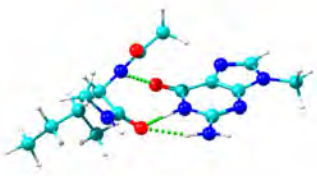
almG17
[19.17]



almG18
[19.74]

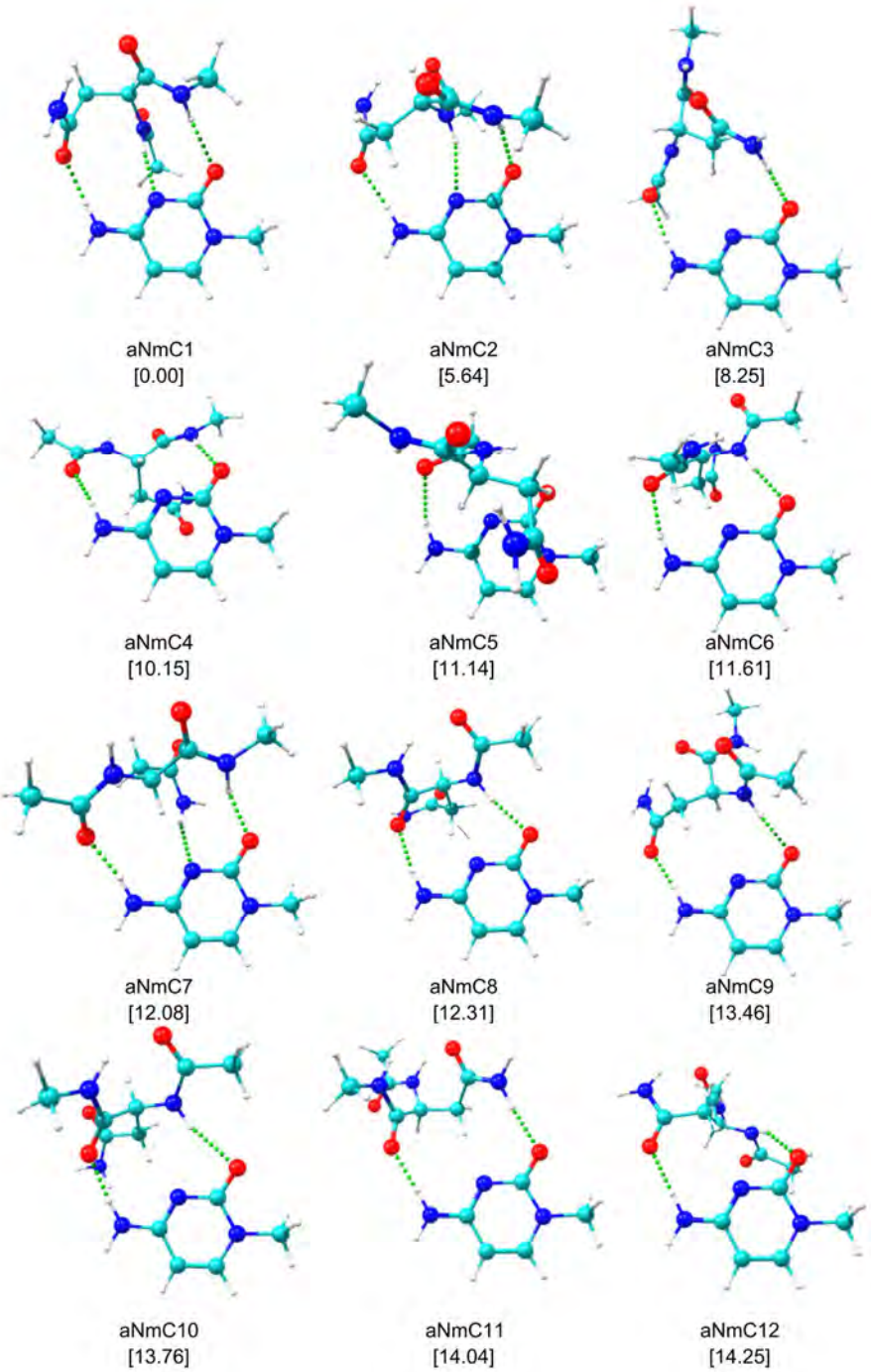


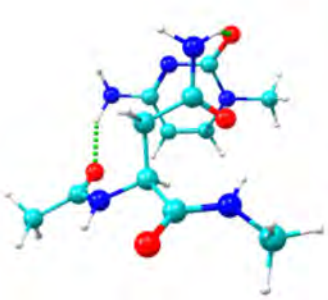
almG19
[20.27]



almG20
[20.40]

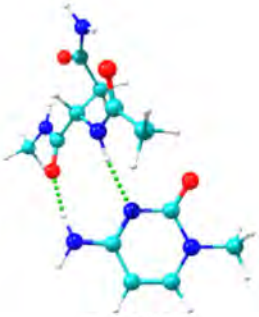
Figure 8.2.13 *The first 20 calculated structures for aNmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



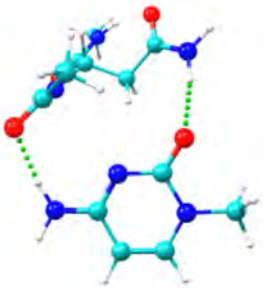


aNmC13
[15.33]

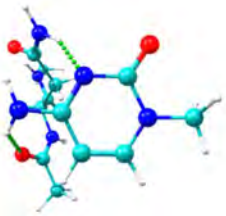
aNmC14
[15.38]



aNmC15
[15.91]

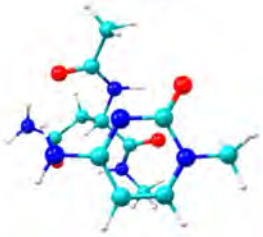


aNmC16
[17.26]



aNmC17
[17.93]

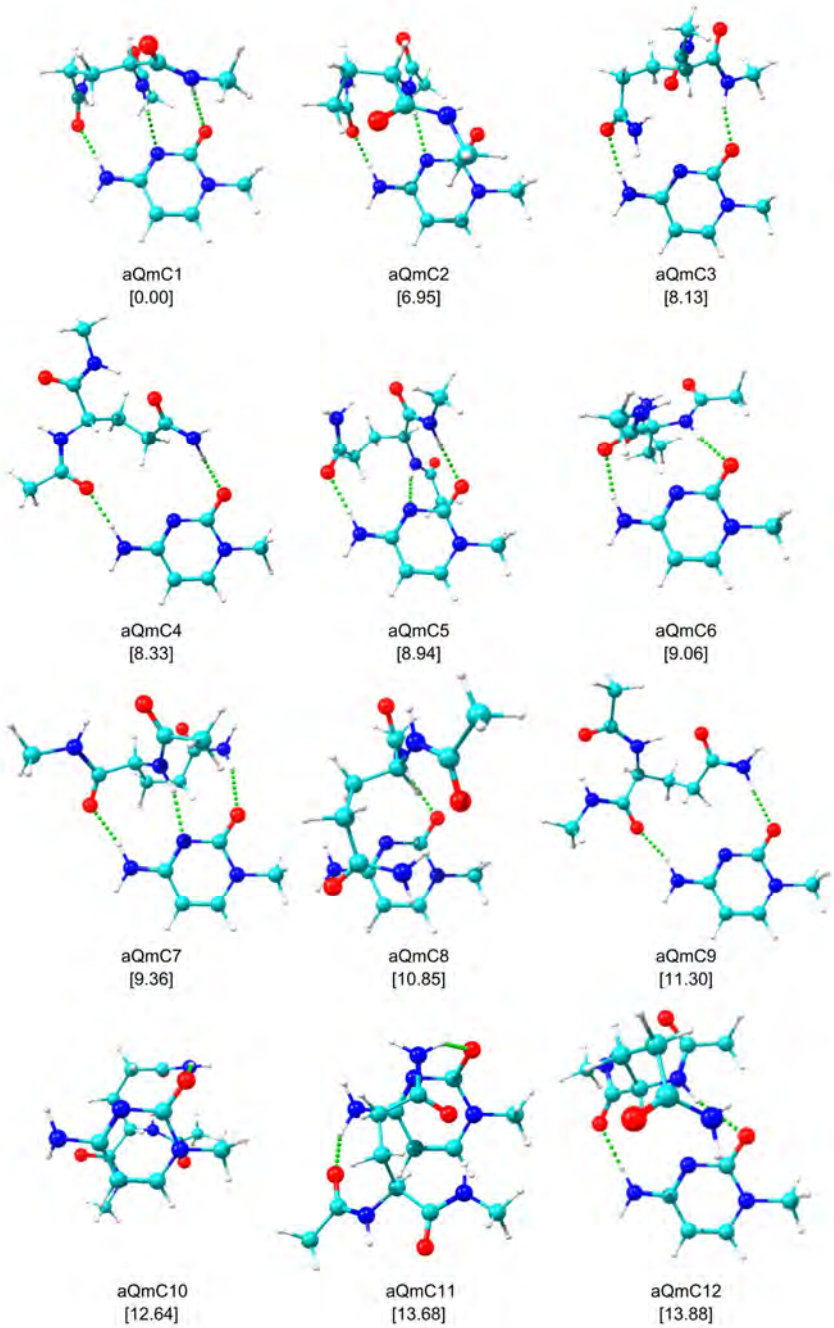
aNmC18
[18.10]

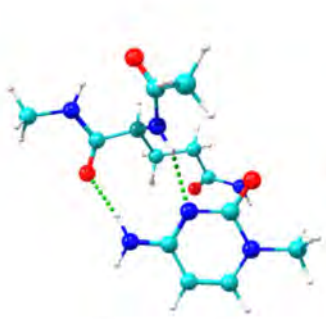


aNmC19
[18.25]

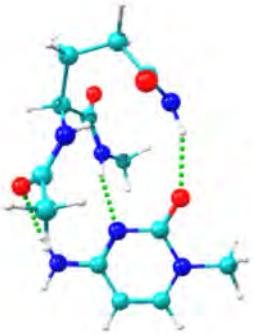
aNmC20
[20.14]

Figure 8.2.14 *The first 20 calculated structures for aQmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

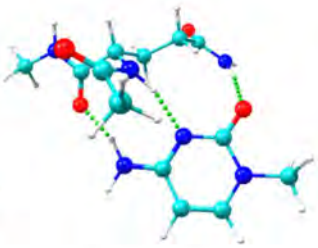




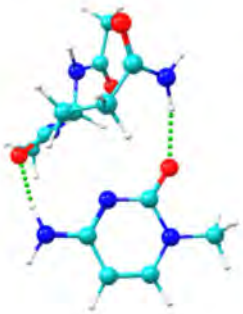
aQmC13
[14.09]



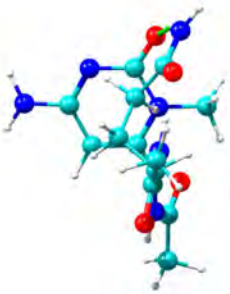
aQmC14
[14.29]



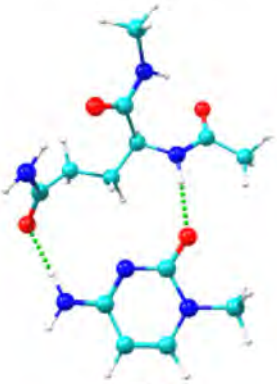
aQmC15
[15.01]



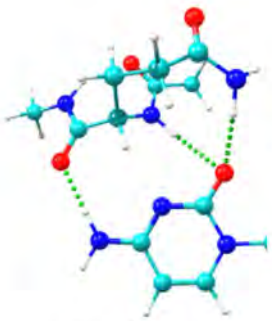
aQmC16
[16.56]



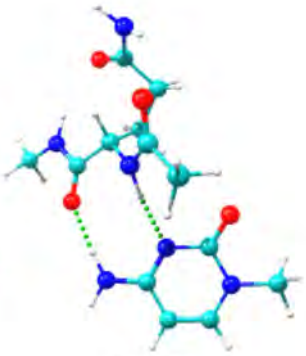
aQmC17
[17.31]



aQmC18
[17.56]

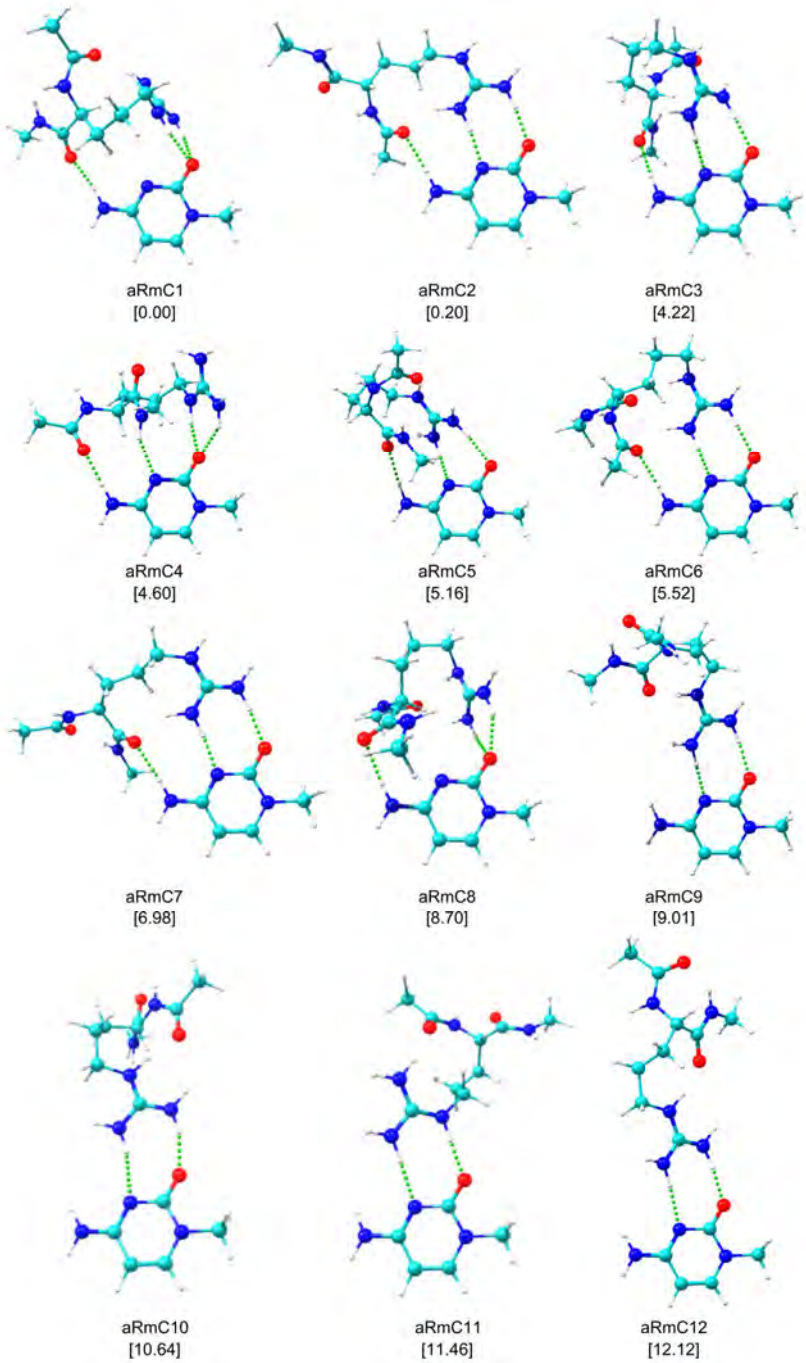


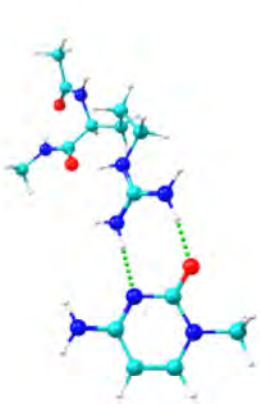
aQmC19
[18.28]



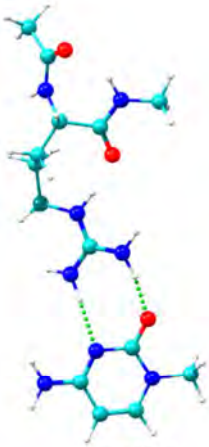
aQmC20
[18.42]

Figure 8.2.15 *The first 20 calculated structures for aRmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

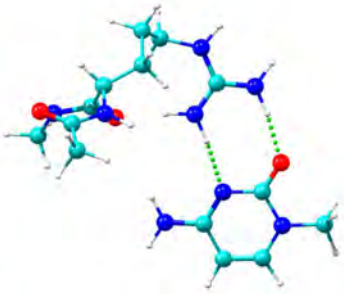




aRmC13
[12.65]

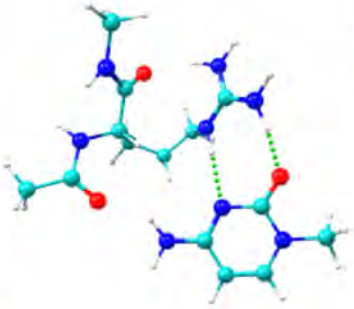


aRmC14
[12.76]

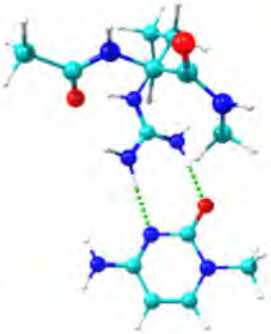


aRmC15
[13.44]

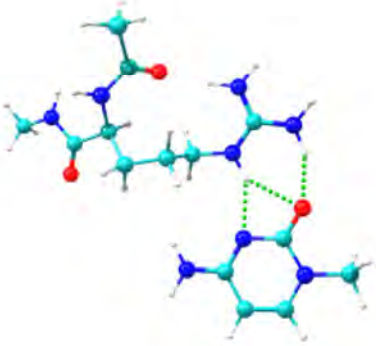
aRmC16
[13.94]



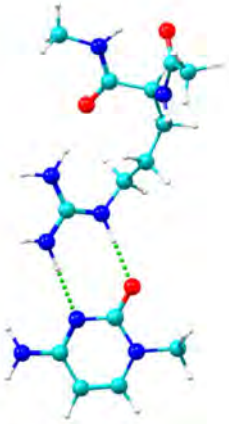
aRmC17
[14.23]



aRmC18
[16.21]

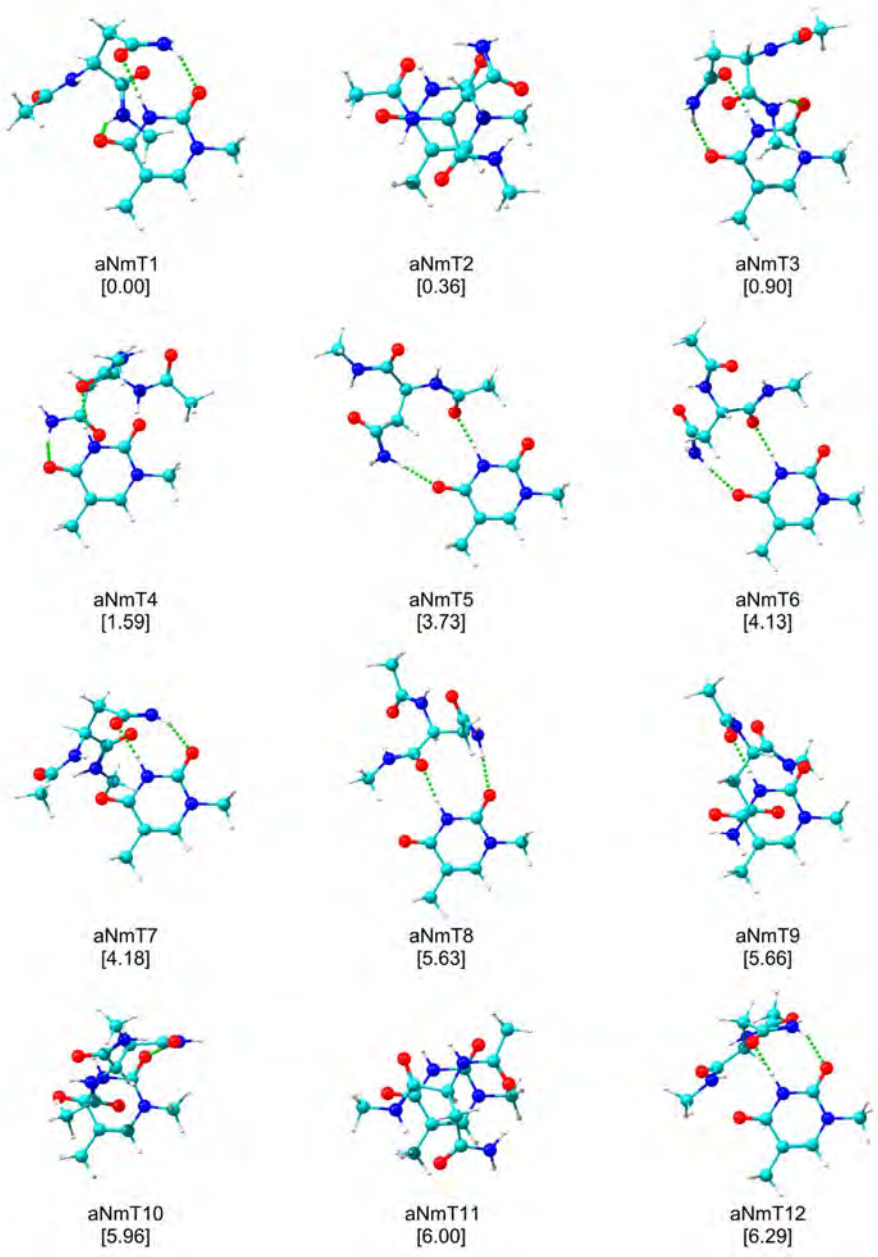


aRmC19
[16.30]



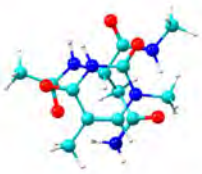
aRmC20
[16.80]

Figure 8.2.16 *The first 20 calculated structures for aNmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



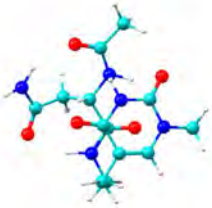
aNmT13
[6.37]

aNmT14
[6.44]

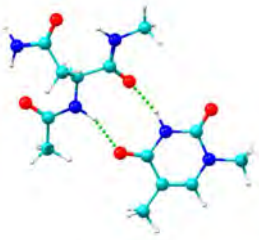


aNmT15
[7.15]

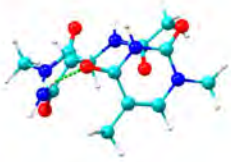
aNmT16
[7.43]



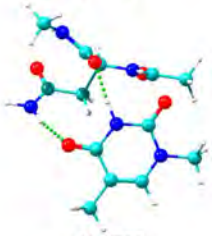
aNmT17
[8.04]



aNmT18
[8.30]

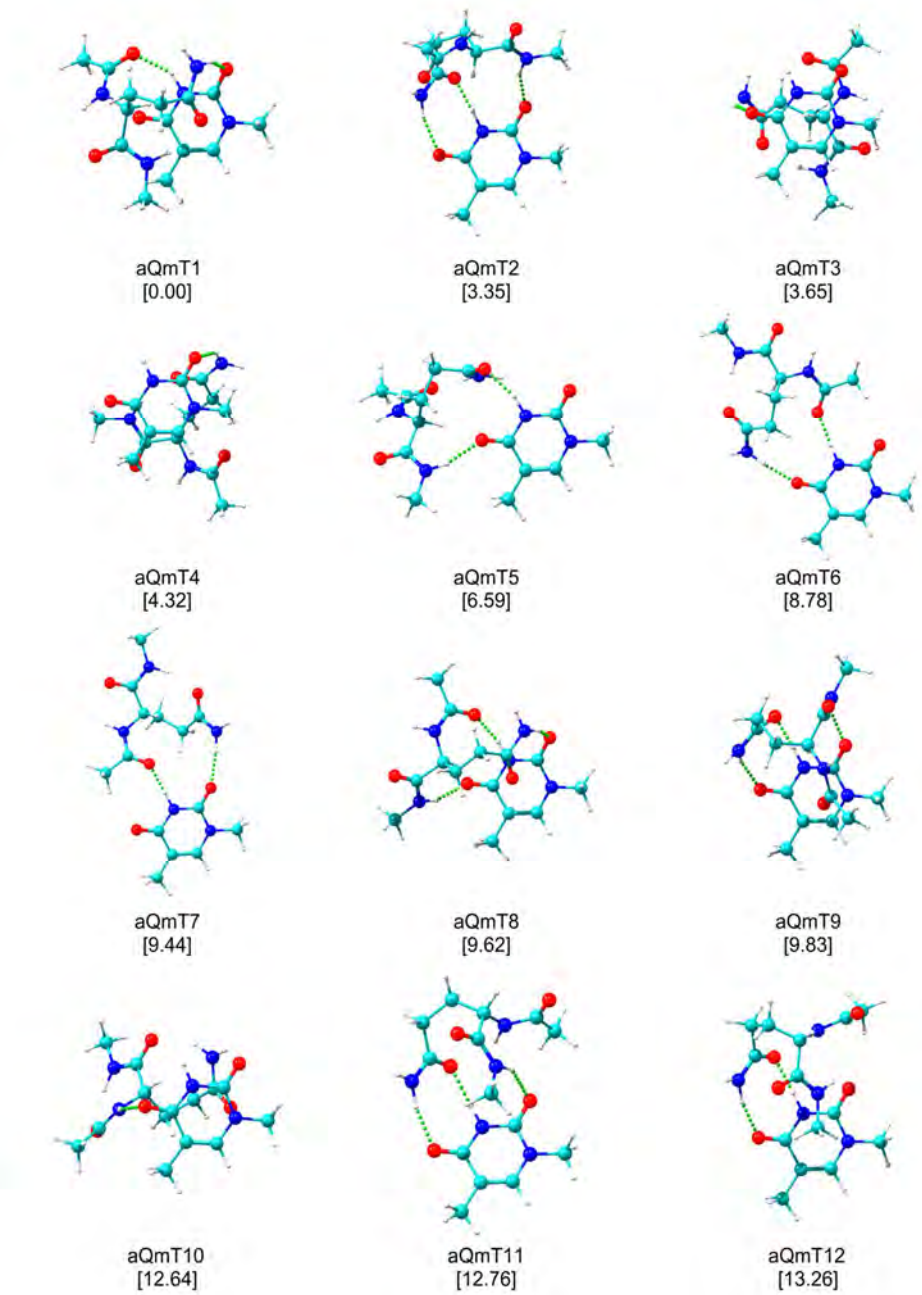


aNmT19
[8.79]

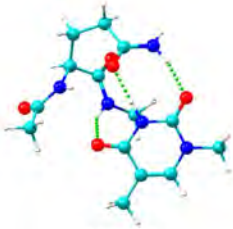


aNmT20
[9.75]

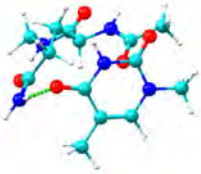
Figure 8.2.17 *The first 20 calculated structures for aQmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



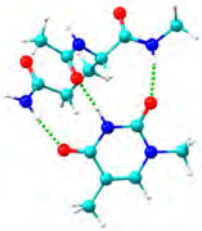
aQmT13
[13.73]



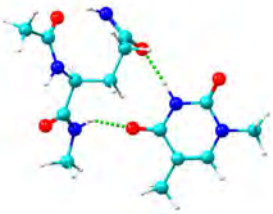
aQmT14
[13.79]



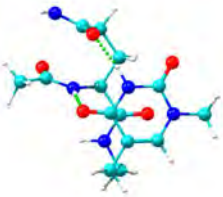
aQmT15
[14.33]



aQmT16
[14.84]

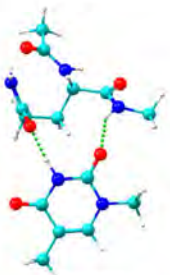


aQmT17
[14.96]



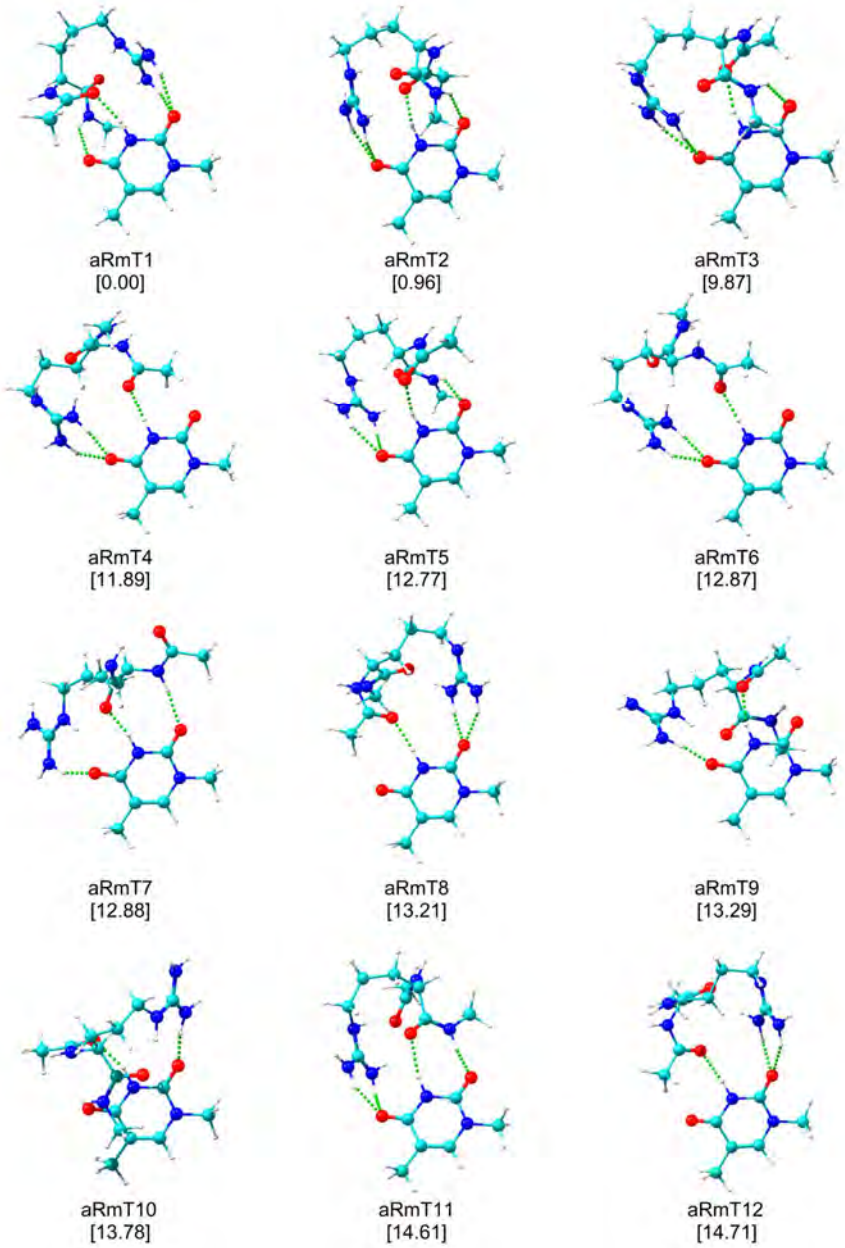
aQmT18
[15.12]

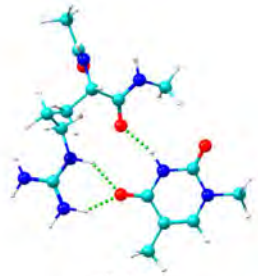
aQmT19
[16.20]



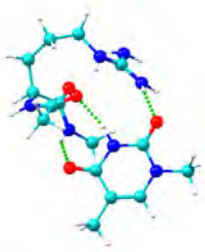
aQmT20
[17.19]

Figure 8.2.18 *The first 20 calculated structures for aRmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

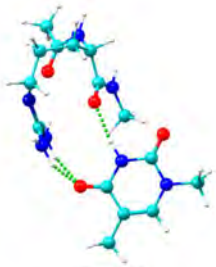




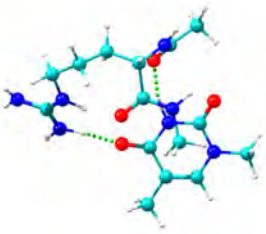
aRmT13
[15.02]



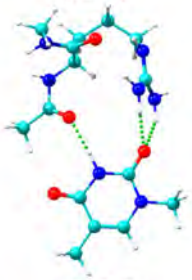
aRmT14
[15.15]



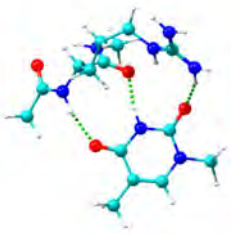
aRmT15
[15.34]



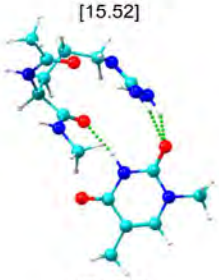
aRmT16
[15.52]



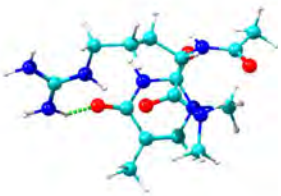
aRmT17
[16.35]



aRmT18
[17.27]

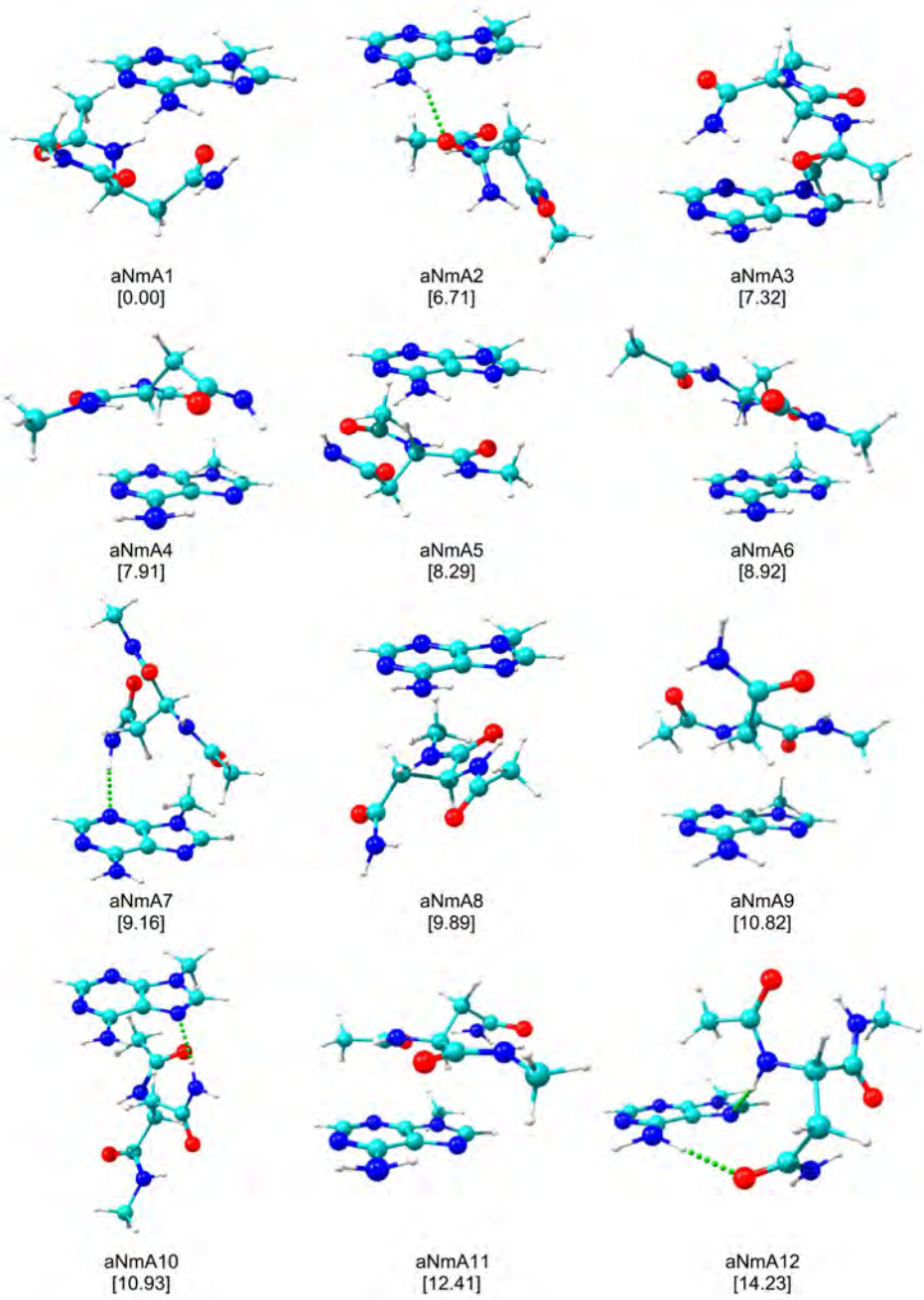


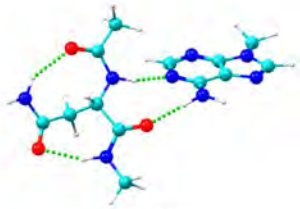
aRmT19
[18.52]



aRmT20
[20.42]

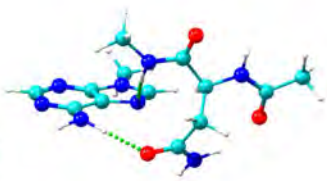
Figure 8.2.19 *The first 20 calculated structures for aNmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



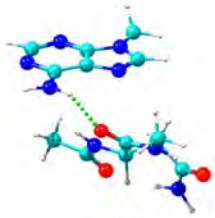


aNmA13
[16.12]

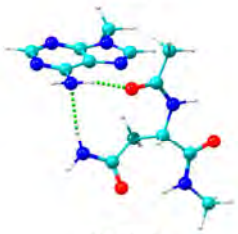
aNmA14
[17.16]



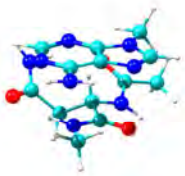
aNmA15
[17.97]



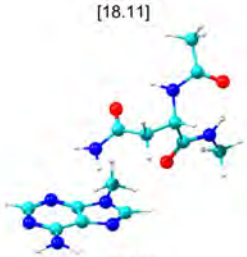
aNmA16
[18.11]



aNmA17
[18.83]



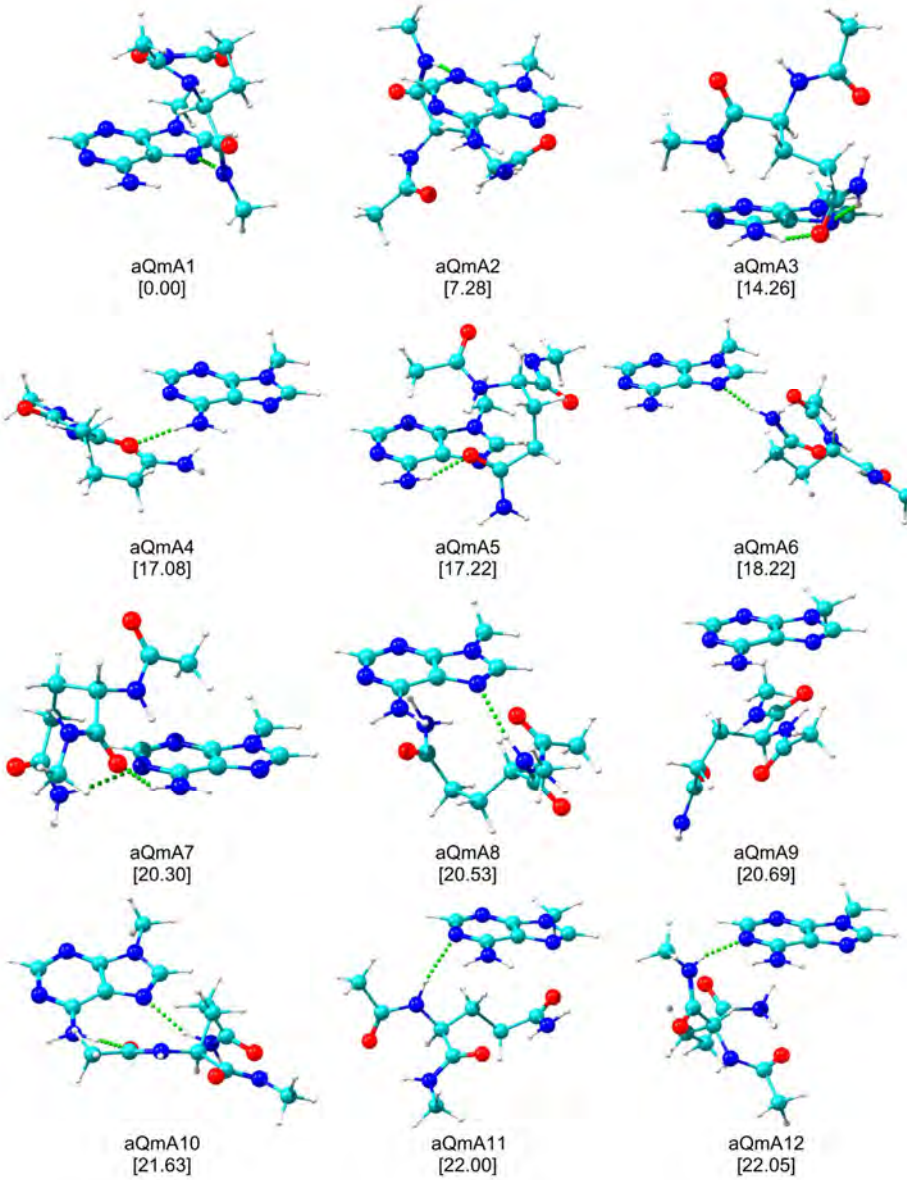
aNmA18
[19.53]

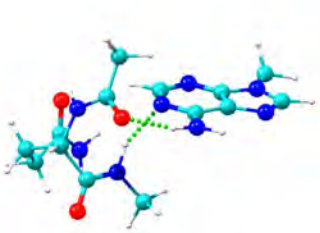


aNmA19
[22.77]

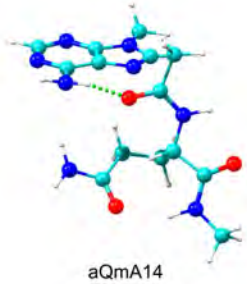
aNmA20
[24.81]

Figure 8.2.20 *The first 20 calculated structures for aQmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



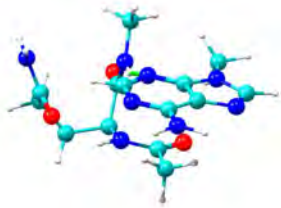


aQmA13
[22.13]



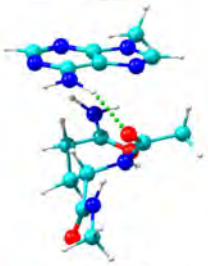
aQmA14
[22.69]

aQmA15
[22.77]



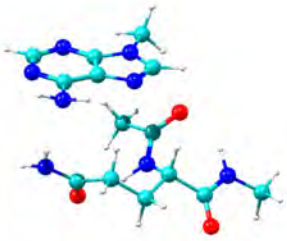
aQmA16
[23.17]

aQmA17
[23.52]



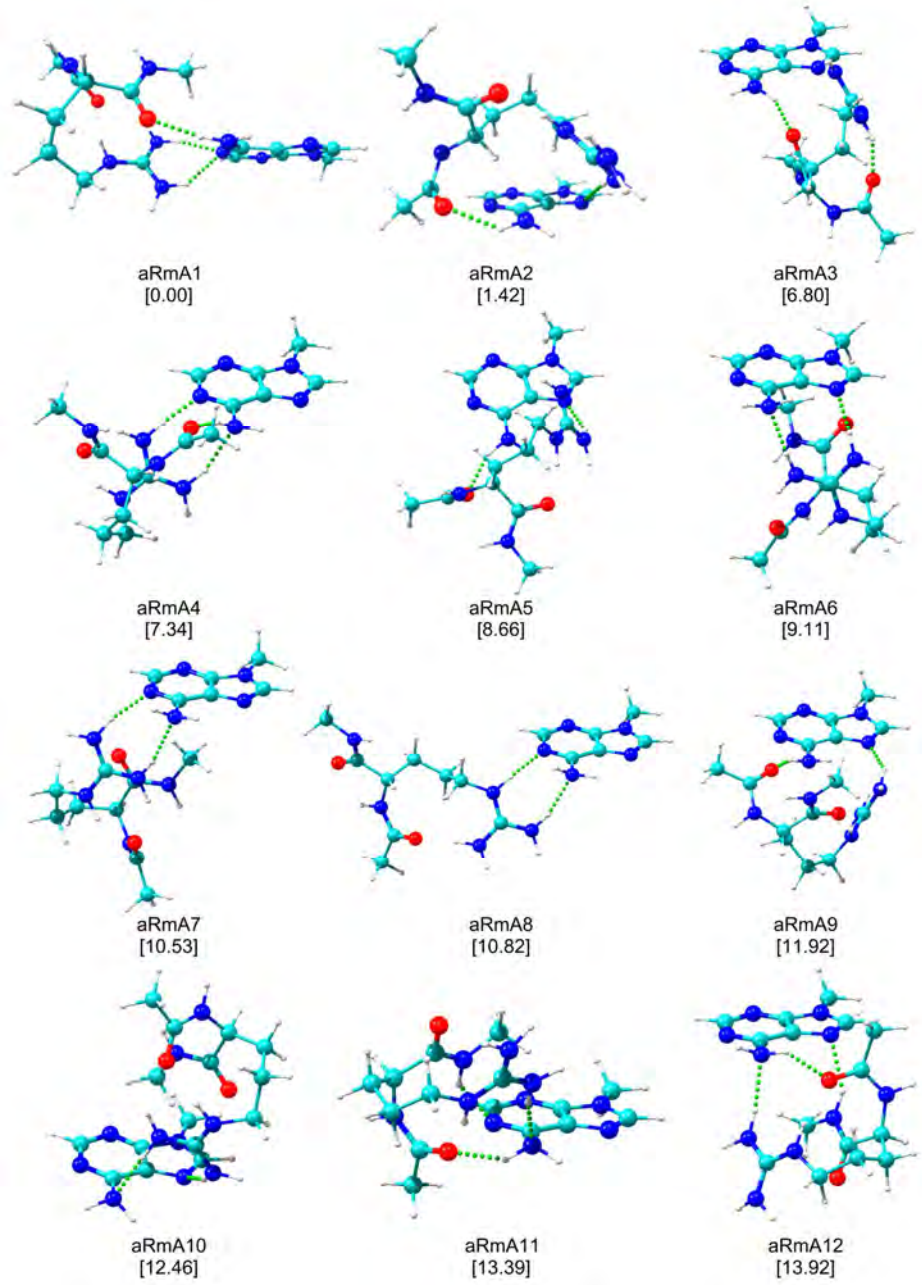
aQmA18
[23.92]

aQmA19
[24.03]



aQmA20
[24.57]

Figure 8.2.21 *The first 20 calculated structures for aRmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



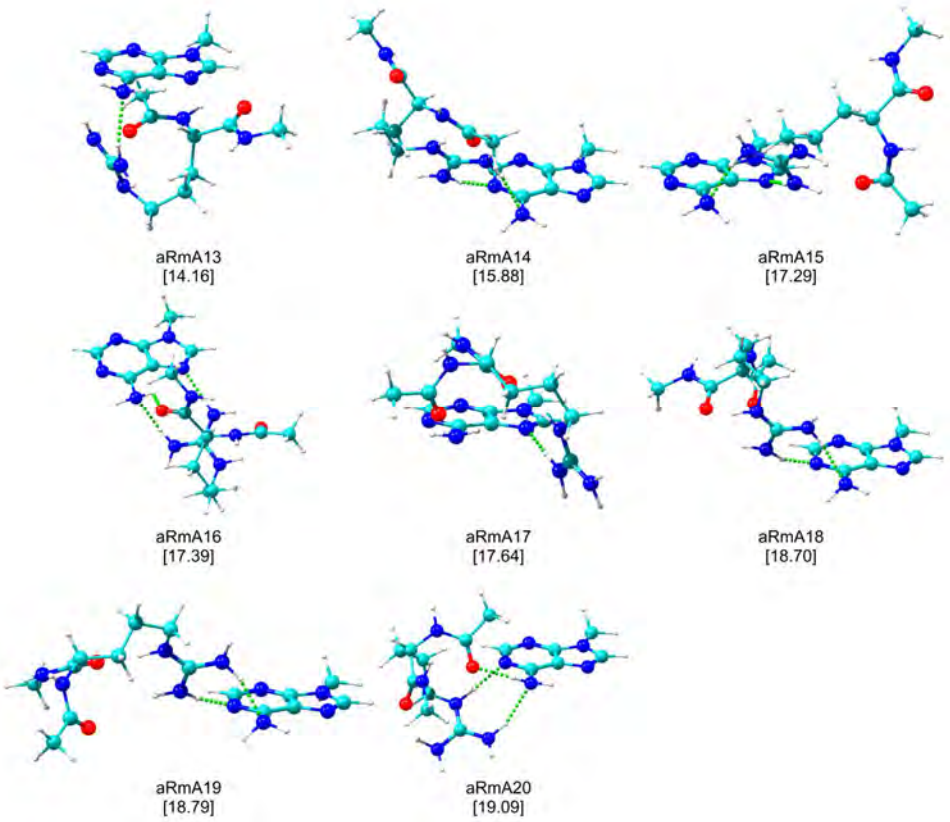
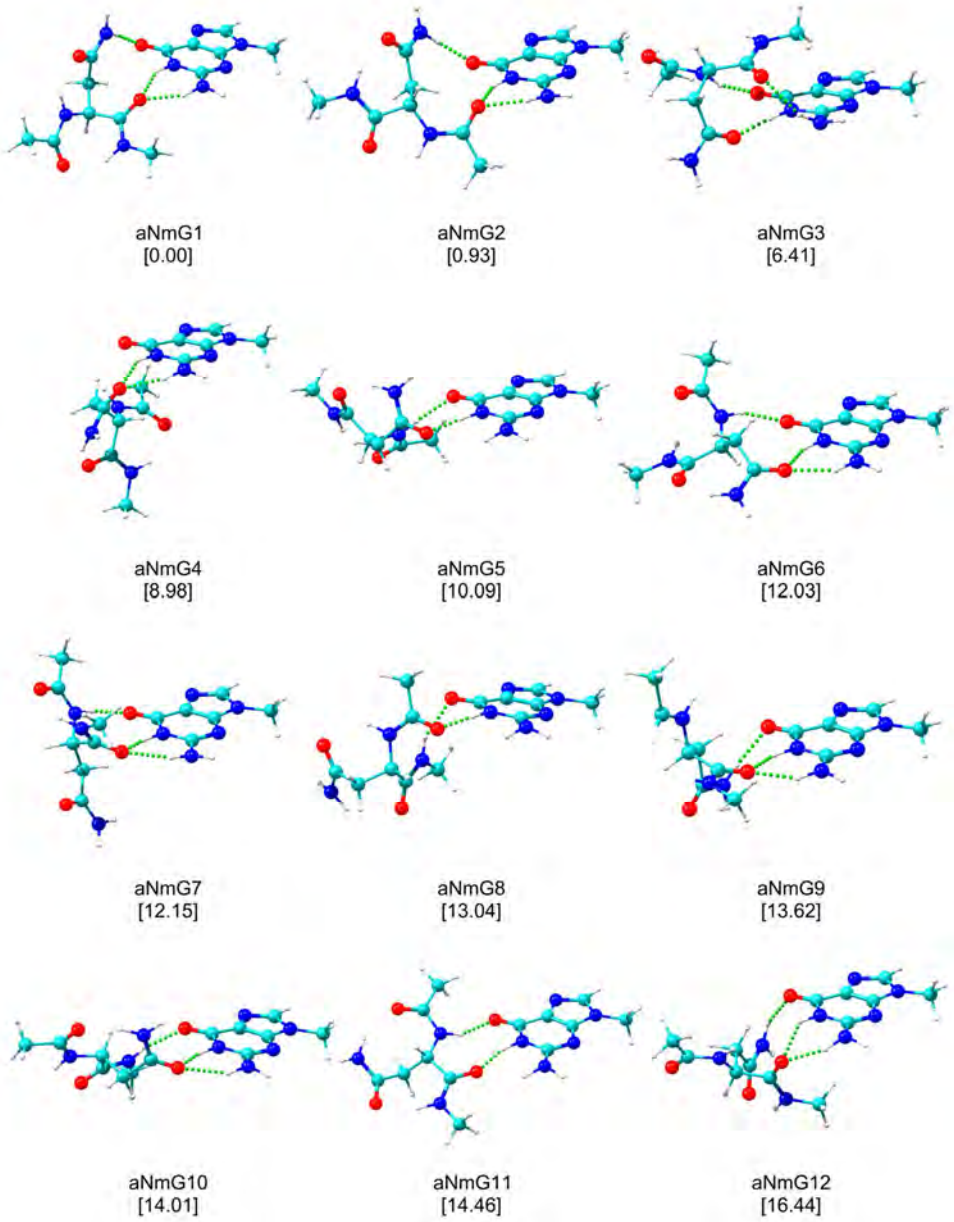
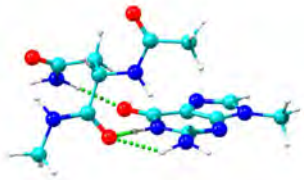
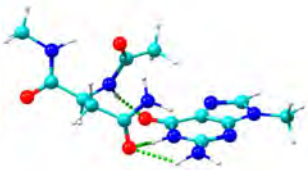

aRmA13
[14.16]
aRmA14
[15.88]
aRmA15
[17.29]
aRmA16
[17.39]
aRmA17
[17.64]
aRmA18
[18.70]
aRmA19
[18.79]
aRmA20
[19.09]

Figure 8.2.22 *The first 20 calculated structures for aNmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

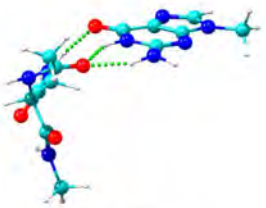




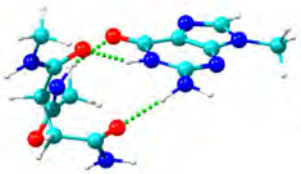
aNmG13
[16.65]



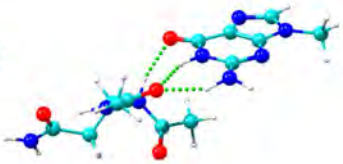
aNmG14
[16.86]



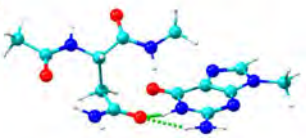
aNmG15
[17.45]



aNmG16
[19.91]

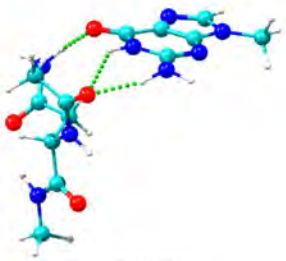


aNmG17
[19.98]



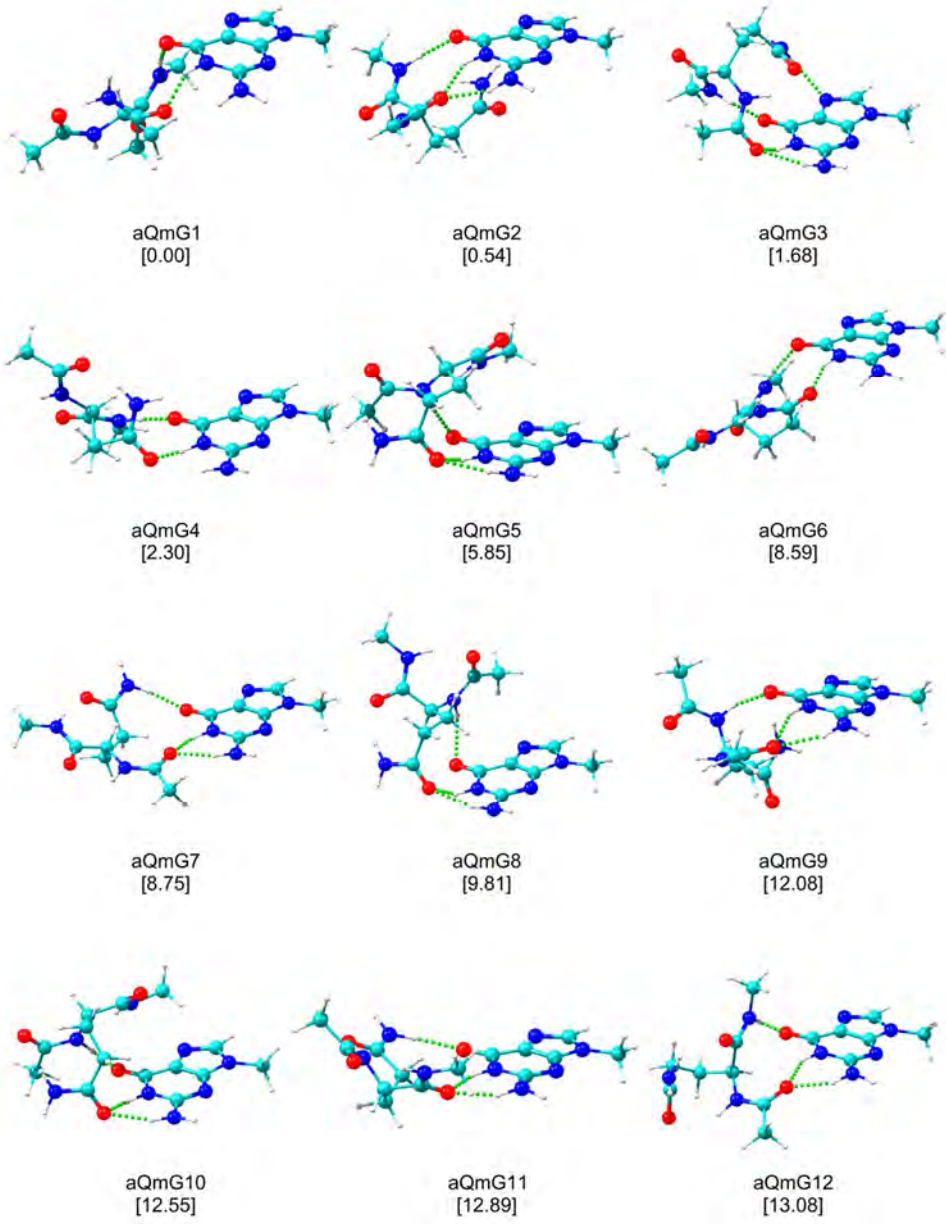
aNmG18
[20.35]

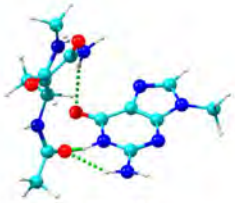
aNmG19
[22.70]



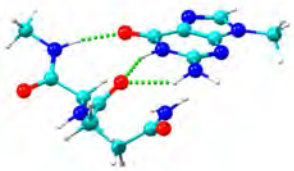
aNmG20
[23.11]

Figure 8.2.23 *The first 20 calculated structures for aQmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

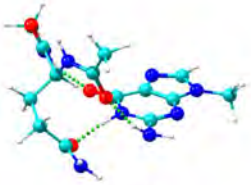




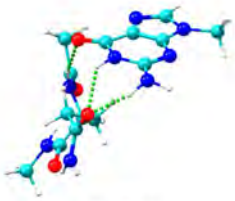
aQmG13
[13.11]



aQmG14
[15.00]

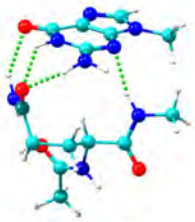


aQmG15
[15.35]

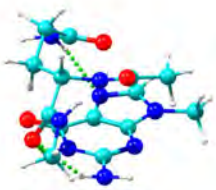


aQmG16
[15.90]

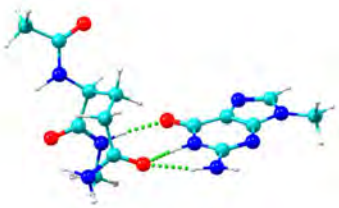
aQmG17
[16.71]



aQmG18
[17.00]

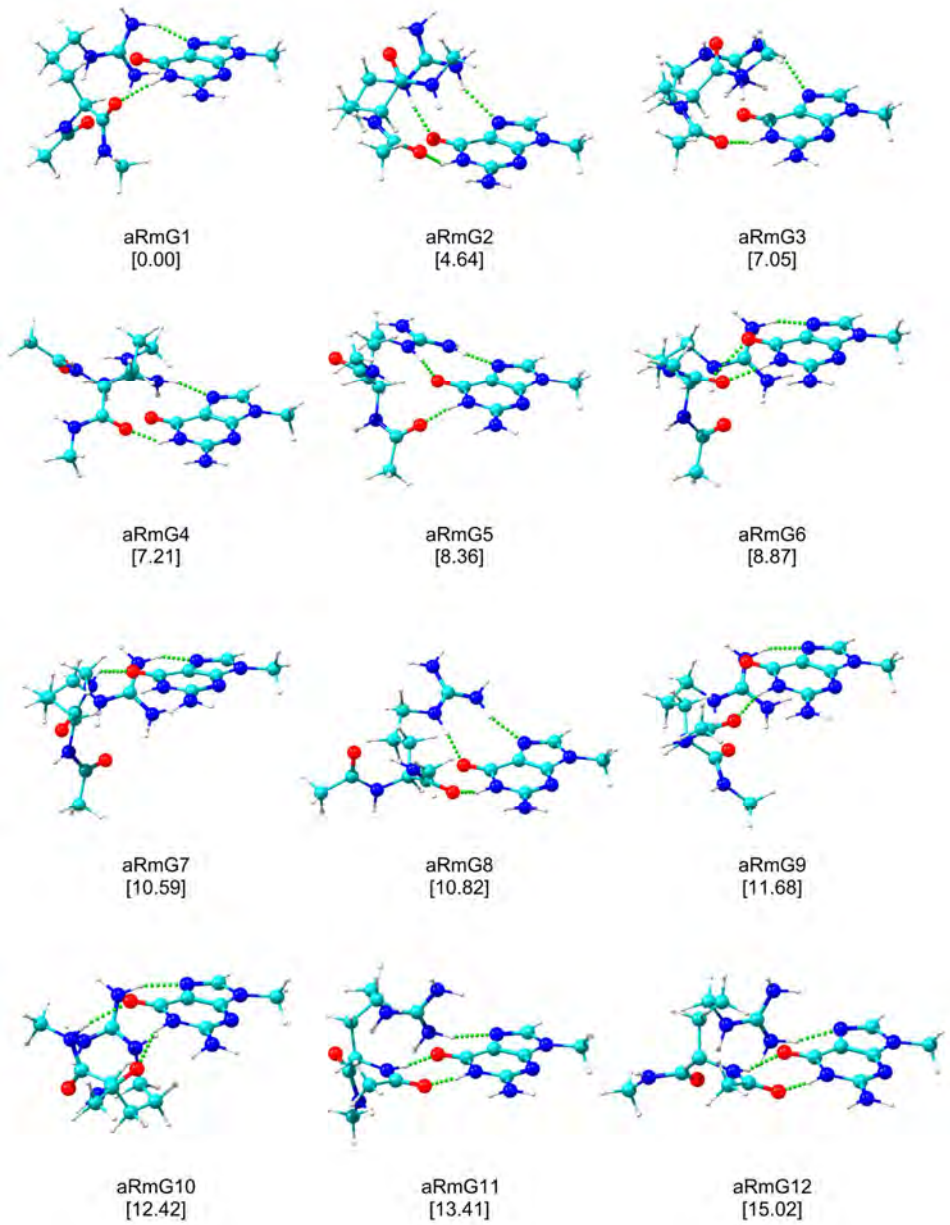


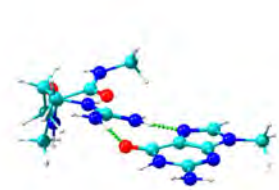
aQmG19
[17.47]



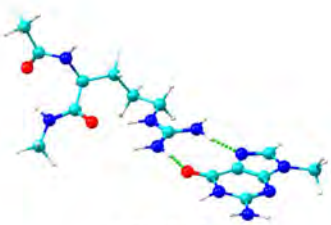
aQmG20
[17.63]

Figure 8.2.24 *The first 20 calculated structures for aRmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

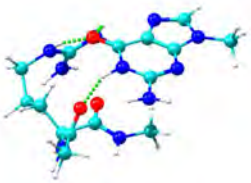




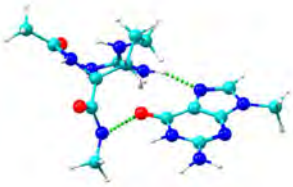
aRmG13
[17.93]



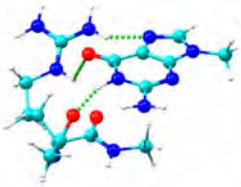
aRmG14
[17.93]



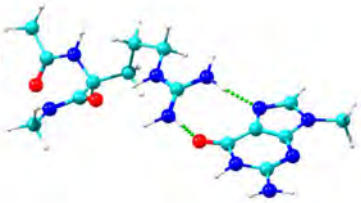
aRmG15
[17.95]



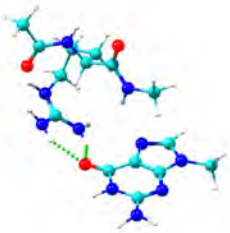
aRmG16
[18.45]



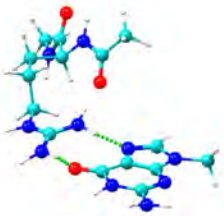
aRmG17
[18.91]



aRmG18
[19.89]

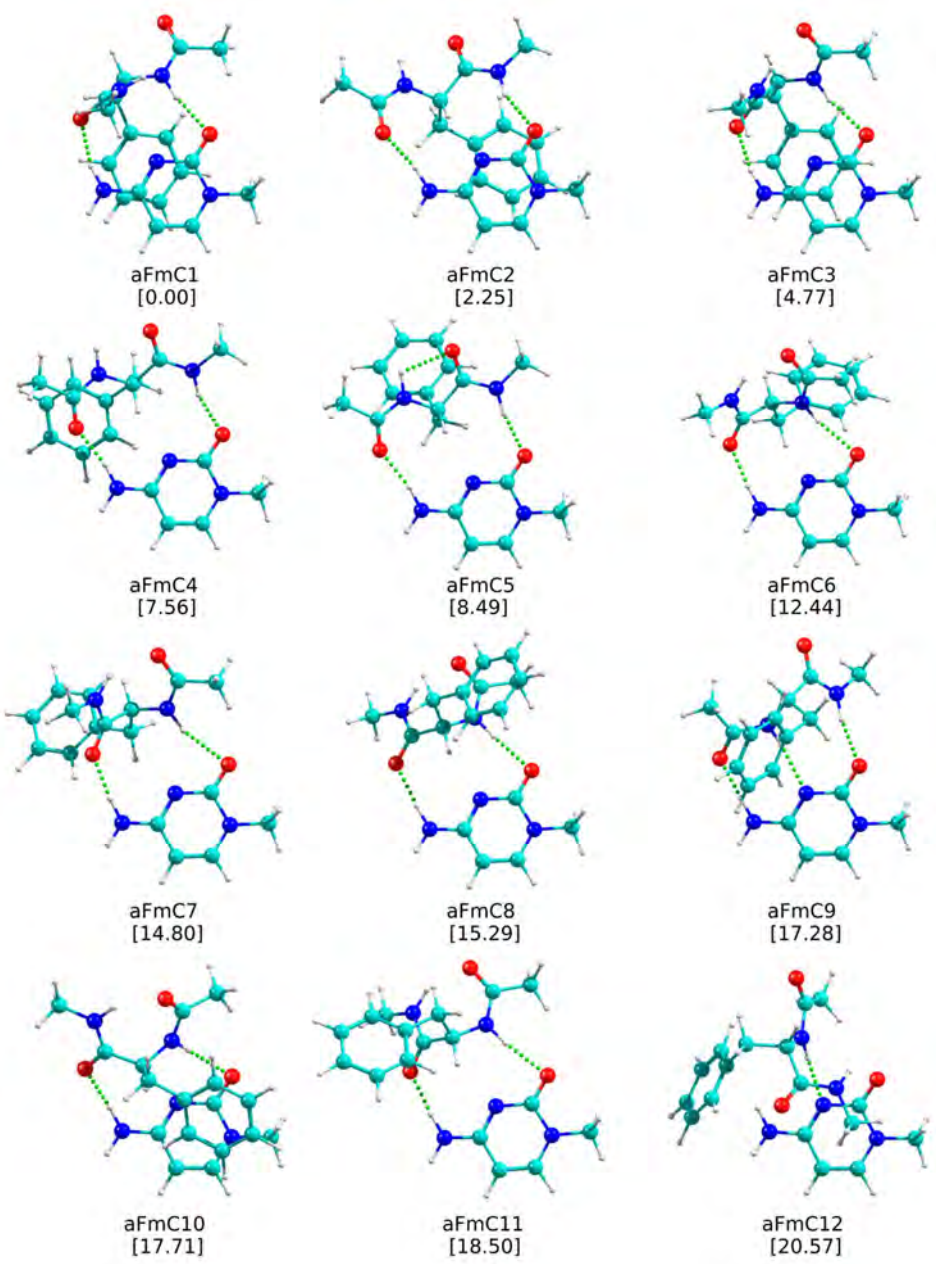


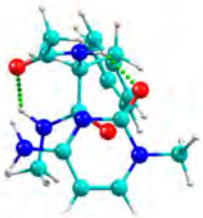
aRmG19
[19.96]



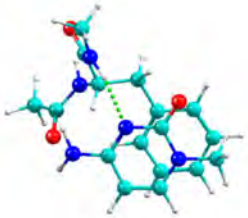
aRmG20
[20.42]

Figure 8.2.25 *The first 20 calculated structures for aFmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

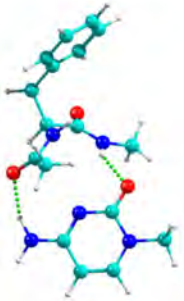




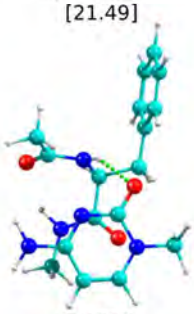
aFmC13
[21.49]



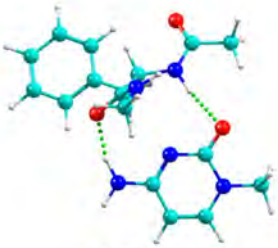
aFmC14
[23.32]



aFmC15
[24.38]



aFmC16
[24.41]



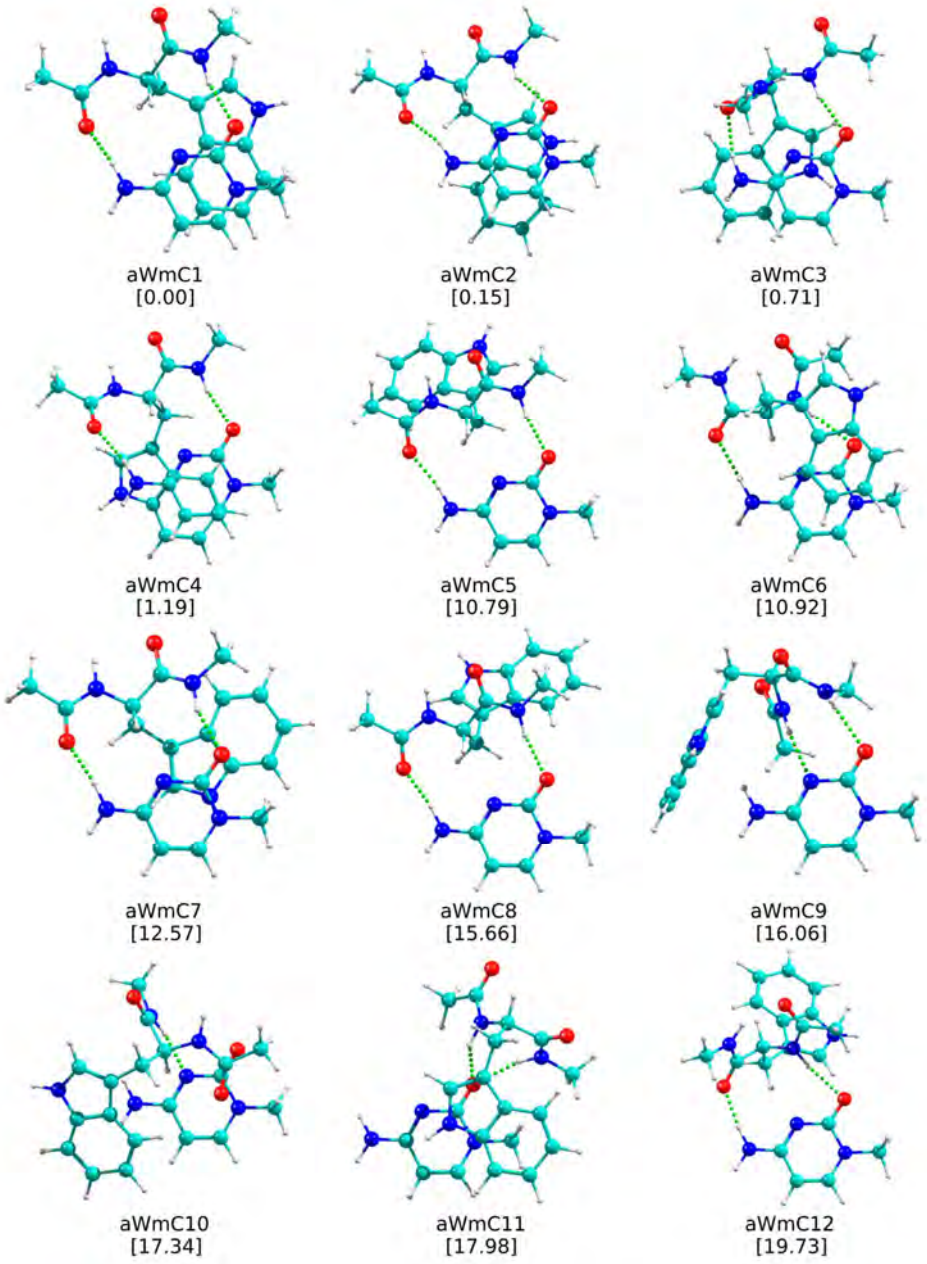
aFmC17
[25.24]

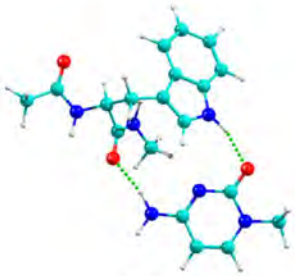
aFmC18
[25.56]

aFmC19
[27.21]

aFmC20
[29.22]

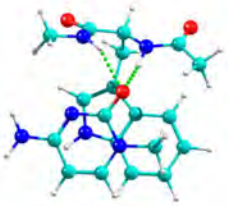
Figure 8.2.26 *The first 20 calculated structures for aWmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



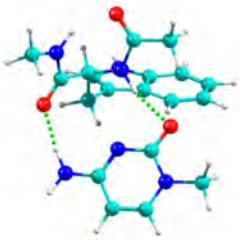


aWmC13
[19.85]

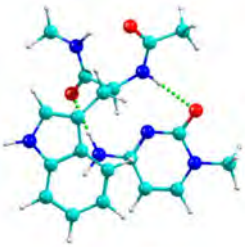
aWmC14
[21.16]



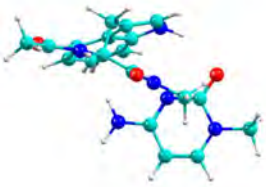
aWmC15
[21.57]



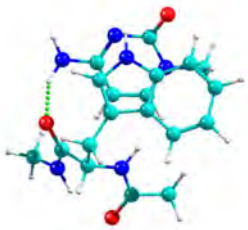
aWmC16
[21.78]



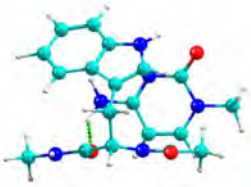
aWmC17
[21.94]



aWmC18
[22.03]

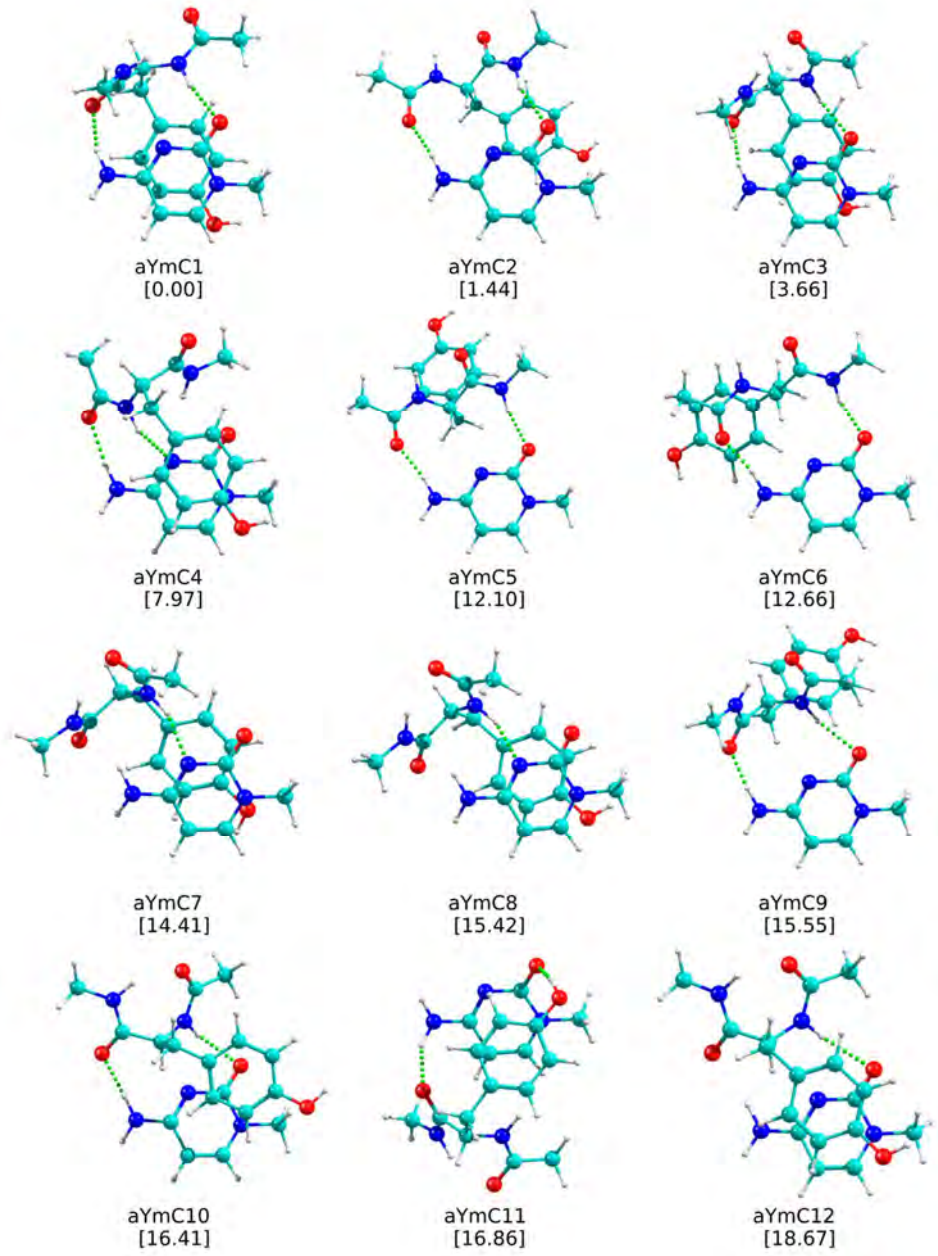


aWmC19
[22.35]



aWmC20
[23.02]

Figure 8.2.27 *The first 20 calculated structures for aYmC at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



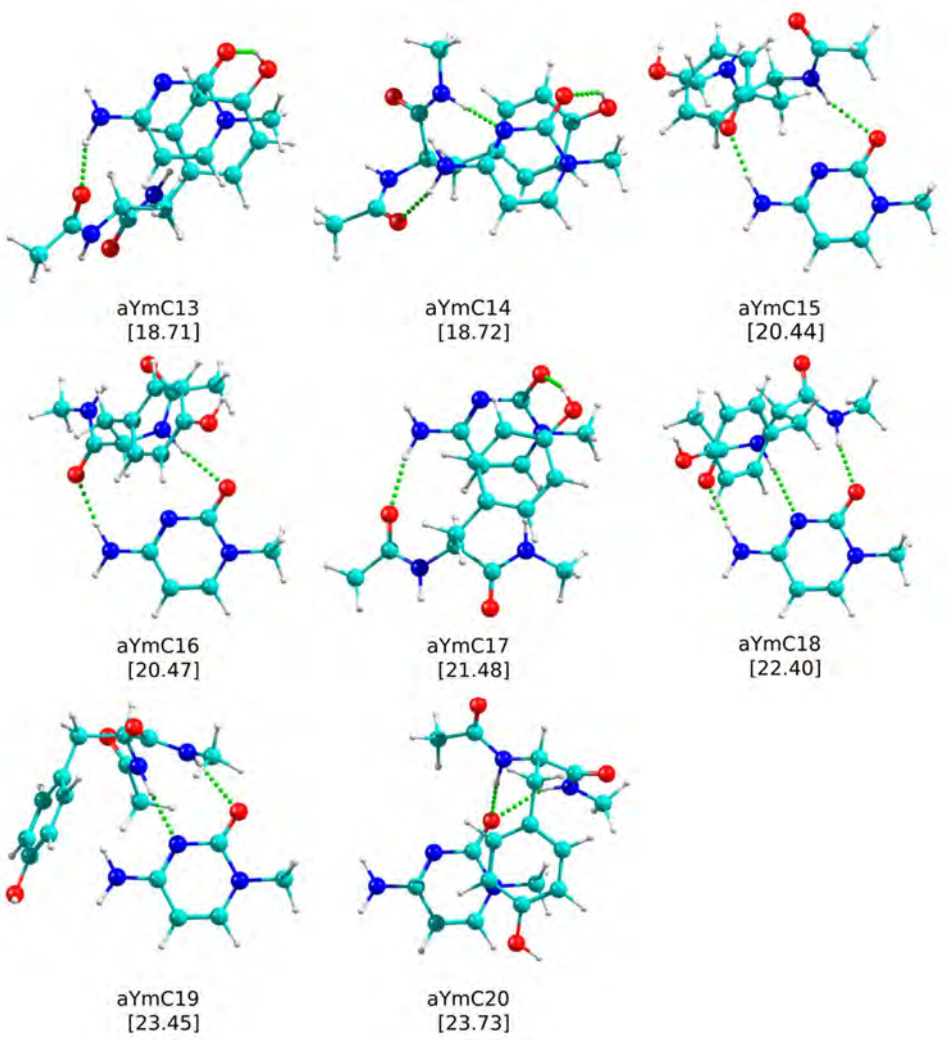
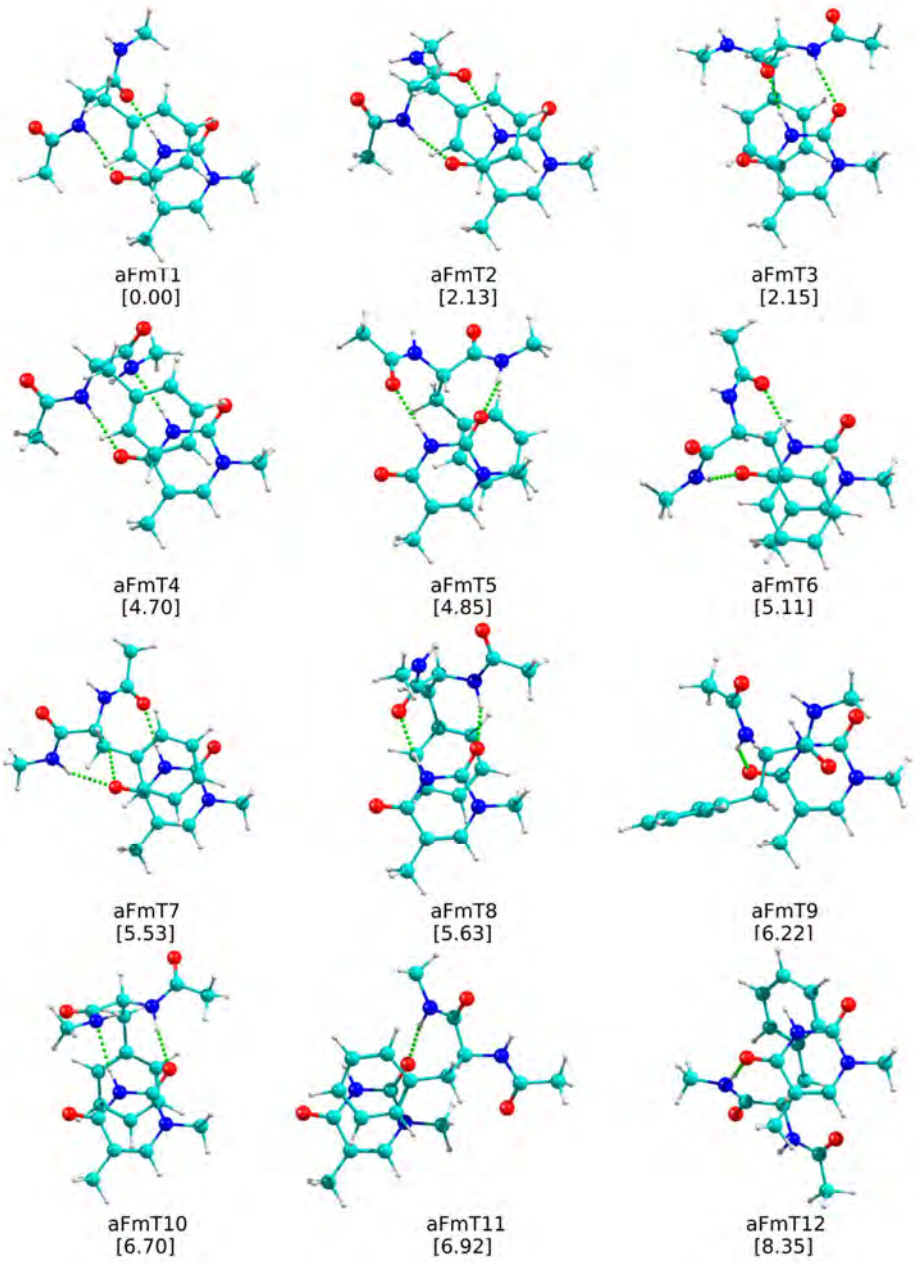

aYmC13
[18.71]
aYmC14
[18.72]
aYmC15
[20.44]
aYmC16
[20.47]
aYmC17
[21.48]
aYmC18
[22.40]
aYmC19
[23.45]
aYmC20
[23.73]

Figure 8.2.28 *The first 20 calculated structures for aFmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



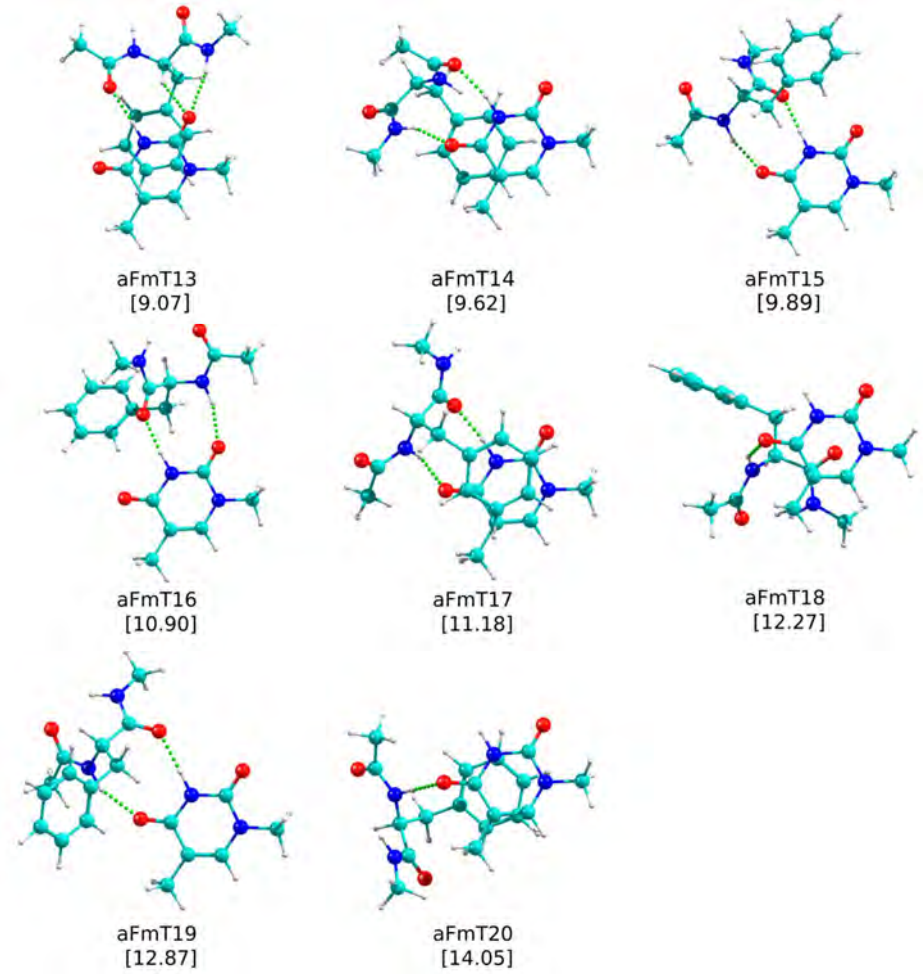
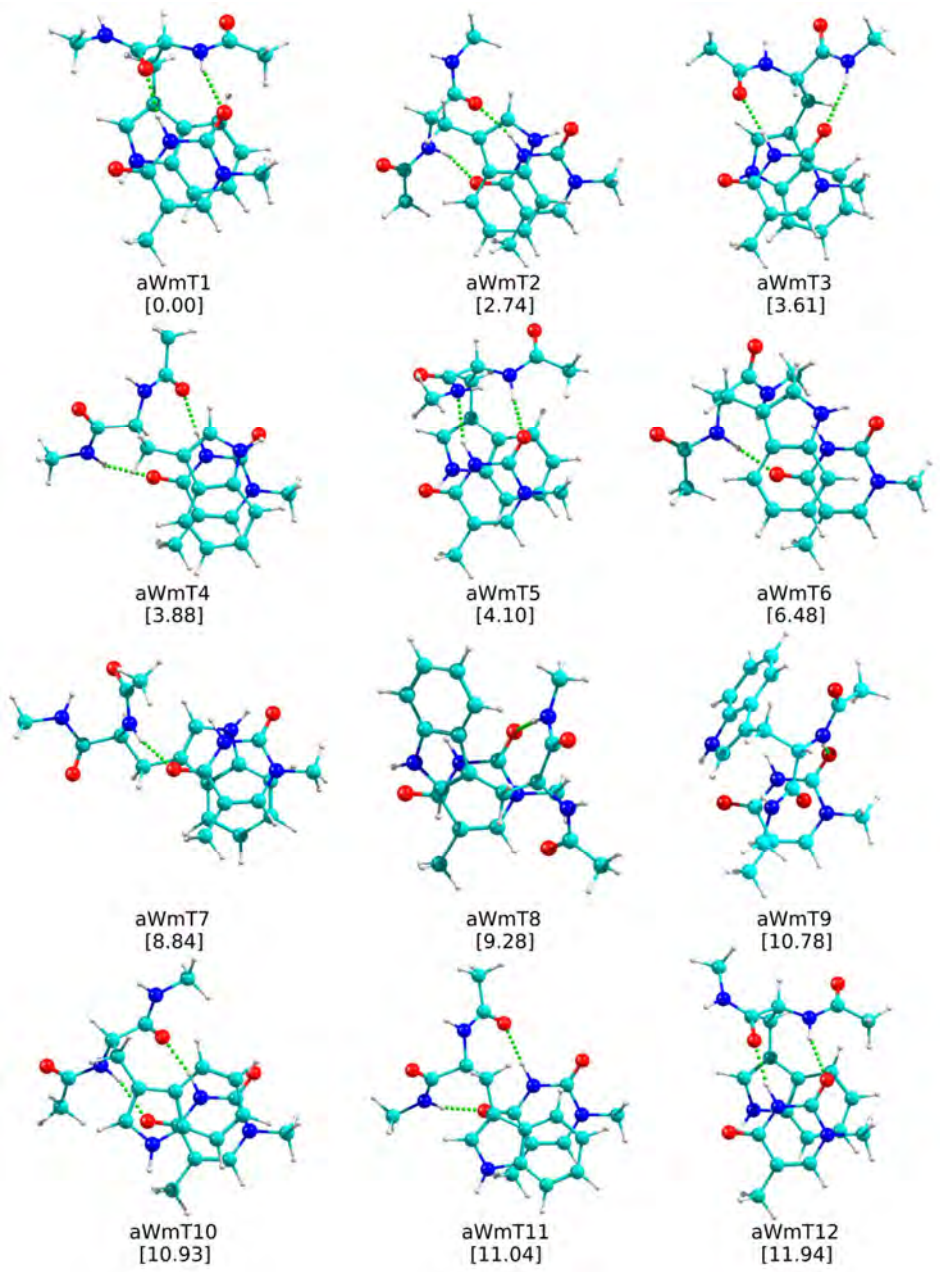
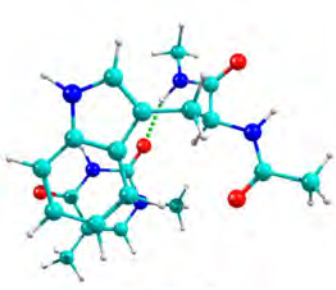
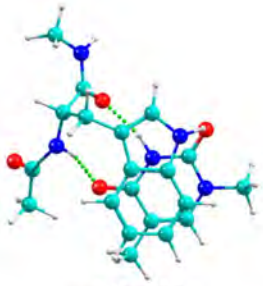

aFmT13
[9.07]
aFmT14
[9.62]
aFmT15
[9.89]
aFmT16
[10.90]
aFmT17
[11.18]
aFmT18
[12.27]
aFmT19
[12.87]
aFmT20
[14.05]

Figure 8.2.29 *The first 20 calculated structures for aWmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

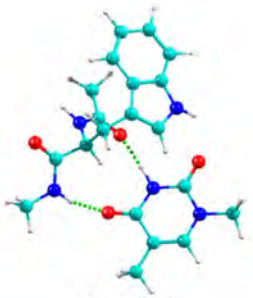




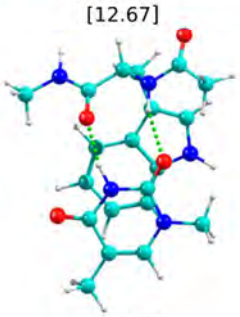
aWmT13
[12.67]



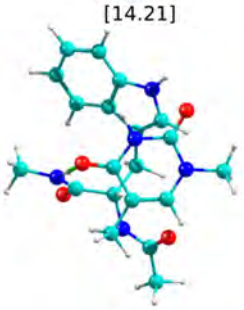
aWmT14
[14.21]



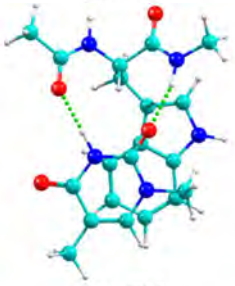
aWmT15
[14.33]



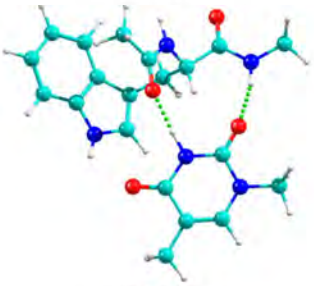
aWmT16
[14.41]



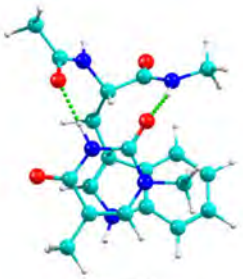
aWmT17
[15.25]



aWmT18
[15.81]

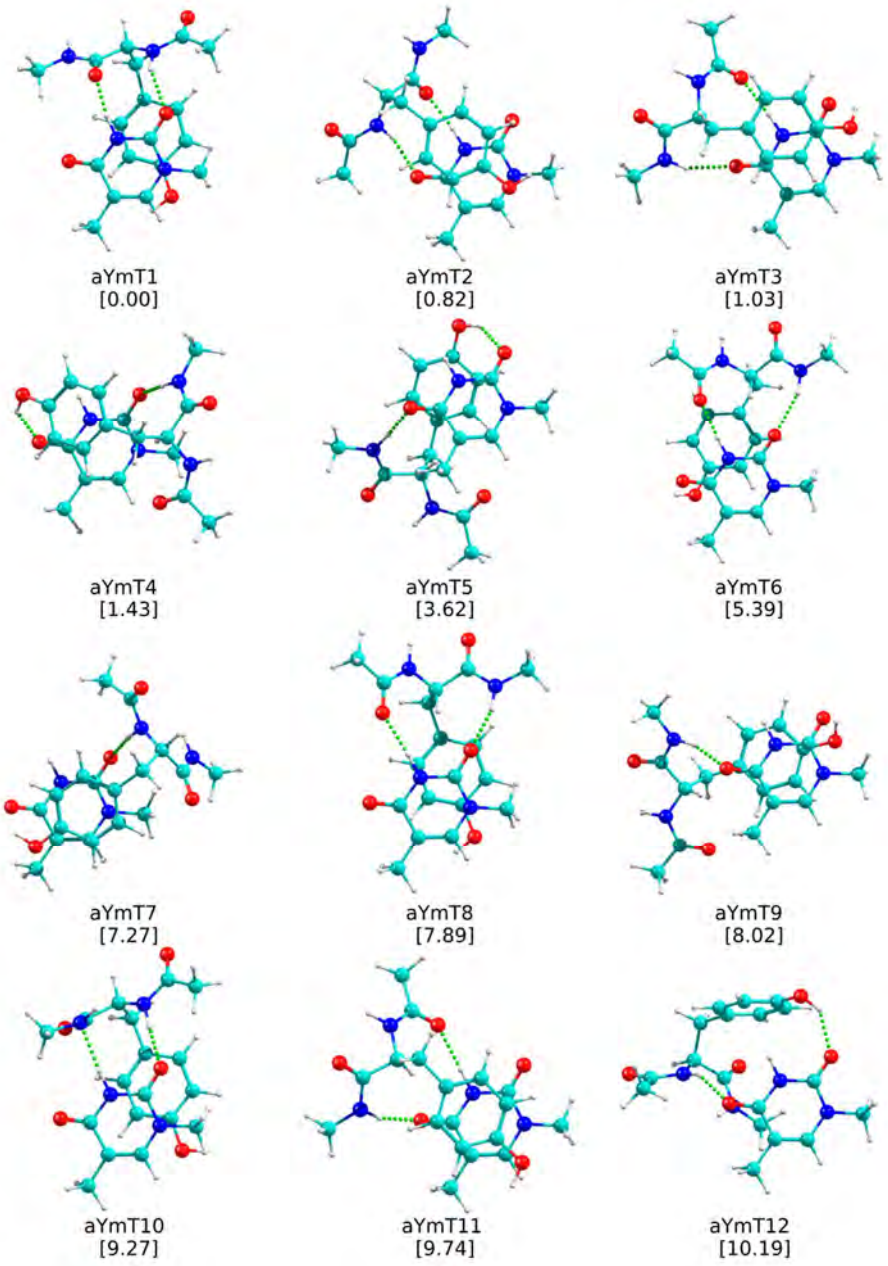


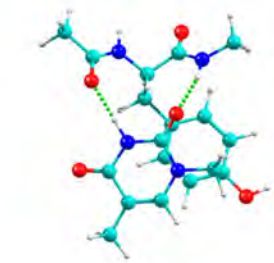
aWmT19
[16.12]



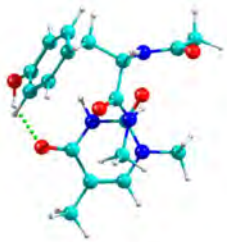
aWmT20
[16.15]

Figure 8.2.30 *The first 20 calculated structures for aYmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

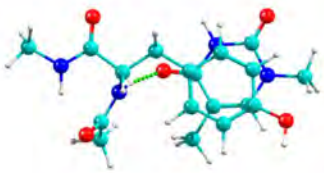




aYmT13
[10.60]



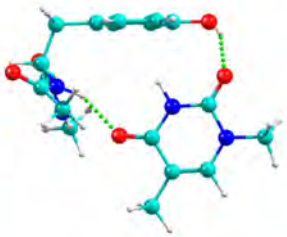
aYmT14
[10.64]



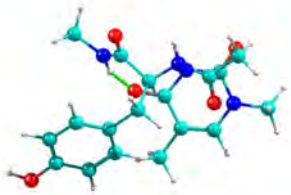
aYmT15
[12.53]

aYmT16
[13.86]

aYmT17
[14.25]



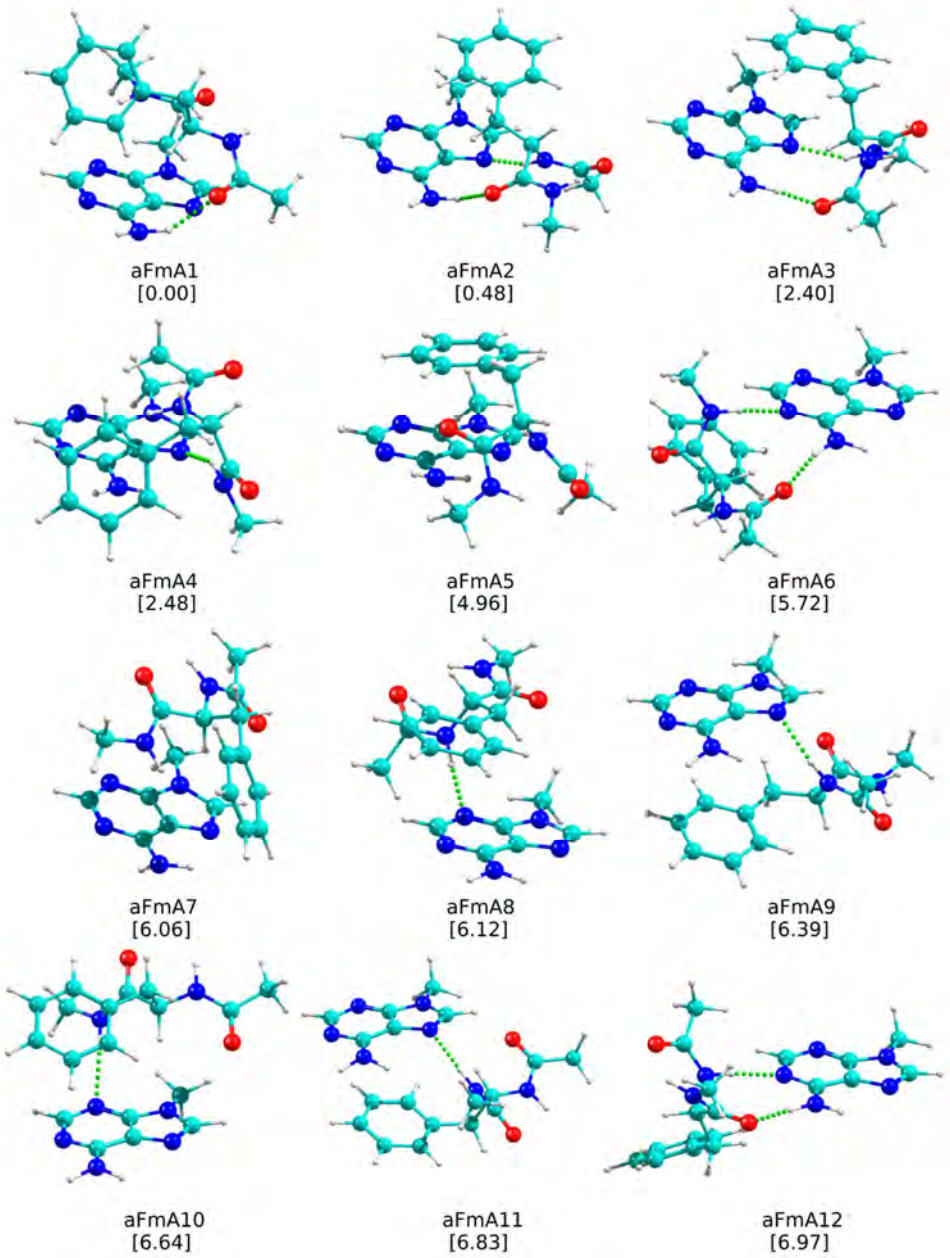
aYmT18
[14.50]



aYmT19
[14.89]

aYmT20
[15.06]

Figure 8.2.31 *The first 20 calculated structures for aFmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



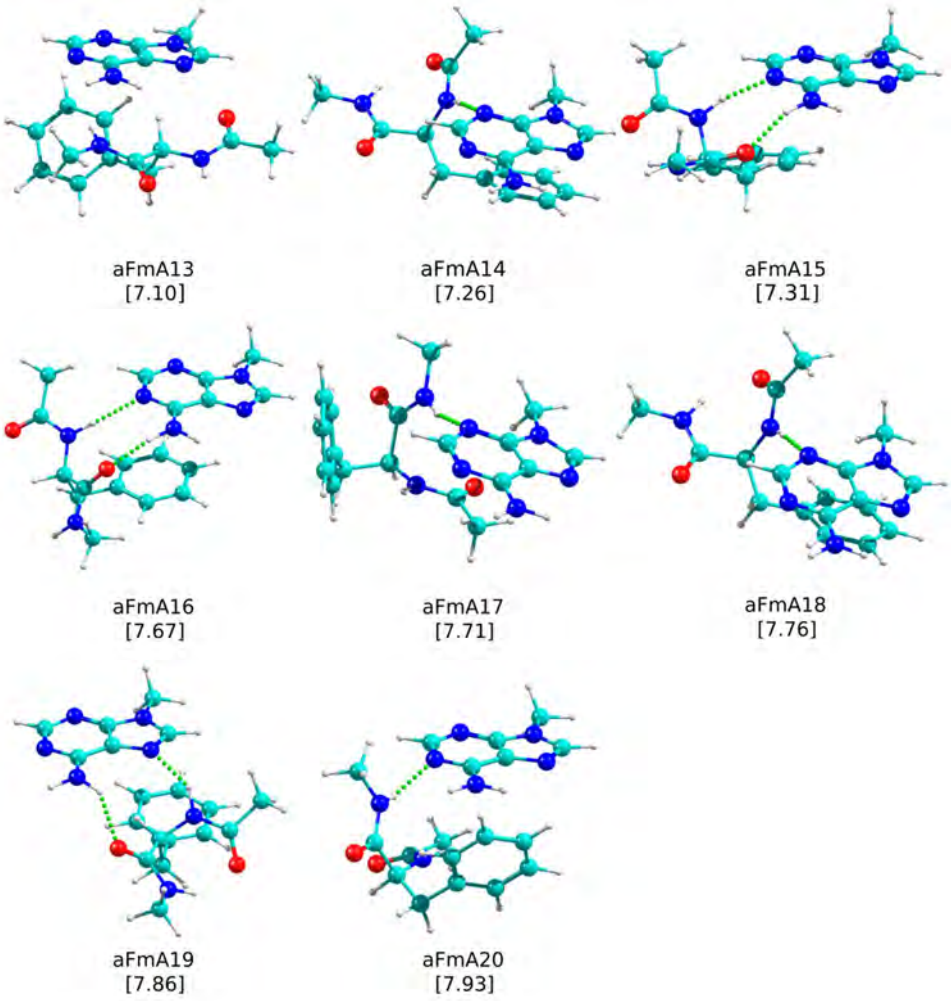
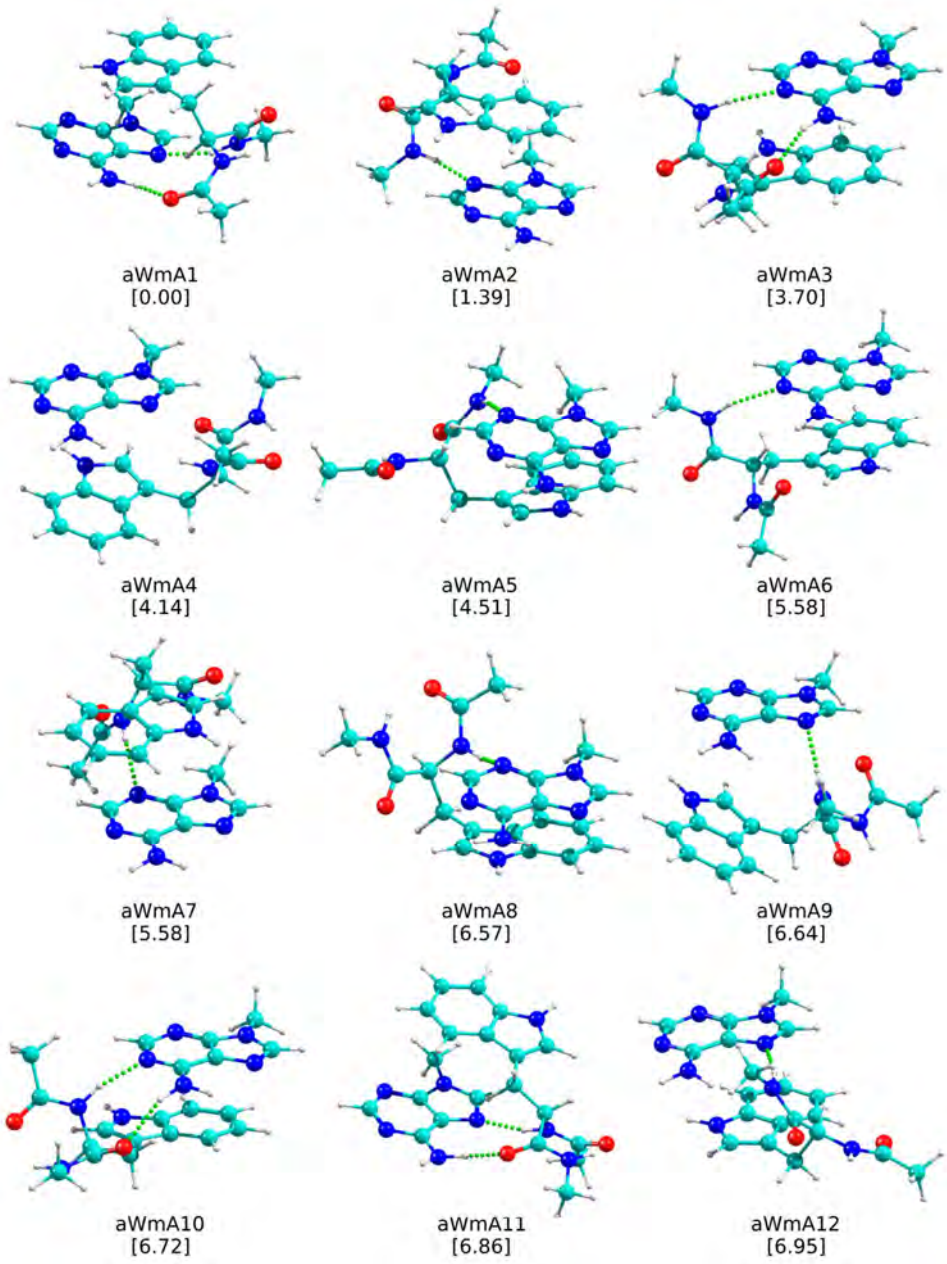
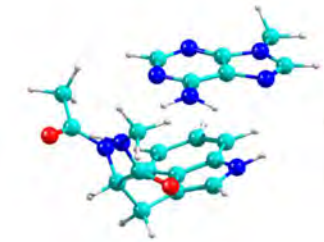
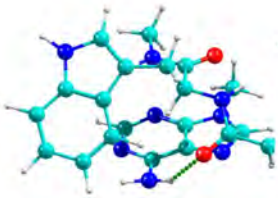

aFmA13
[7.10]
aFmA14
[7.26]
aFmA15
[7.31]
aFmA16
[7.67]
aFmA17
[7.71]
aFmA18
[7.76]
aFmA19
[7.86]
aFmA20
[7.93]

Figure 8.2.32 *The first 20 calculated structures for aWmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

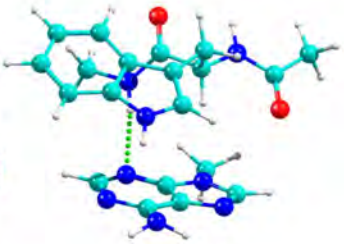




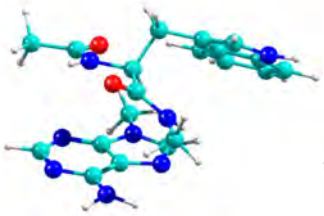
aWmA13
[8.12]



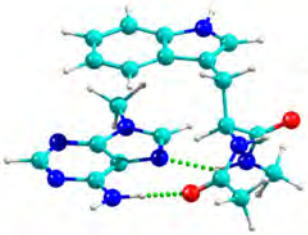
aWmA14
[8.70]



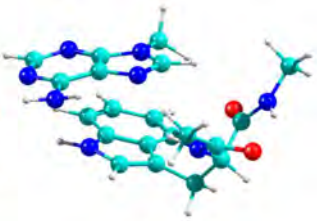
aWmA15
[9.46]



aWmA16
[9.89]

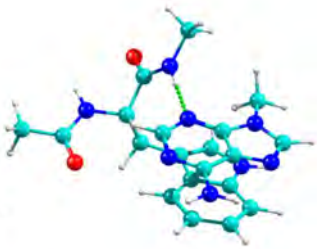


aWmA17
[9.94]



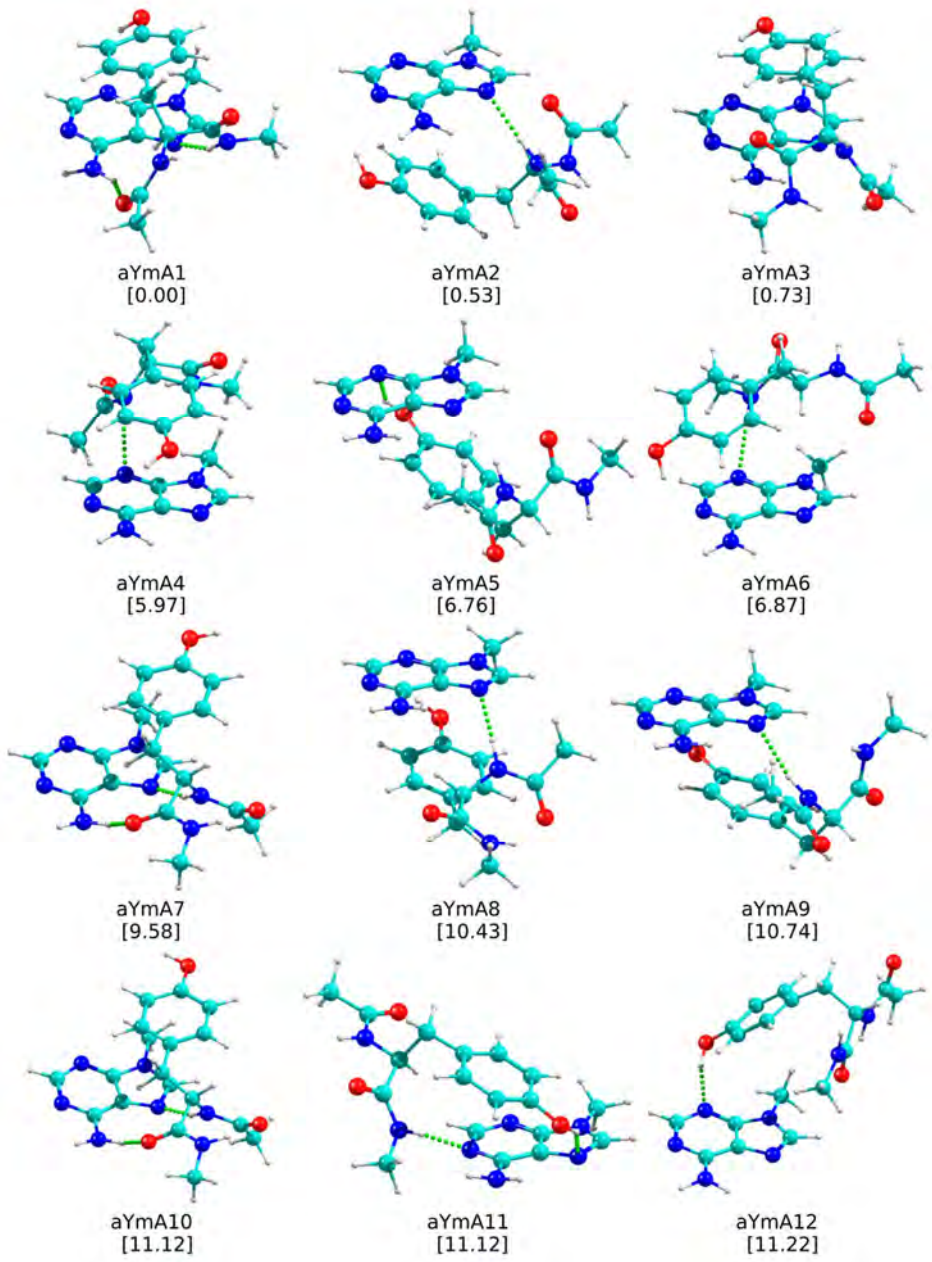
aWmA18
[11.90]

aWmA19
[12.67]



aWmA20
[13.07]

Figure 8.2.33 *The first 20 calculated structures for aYmA at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



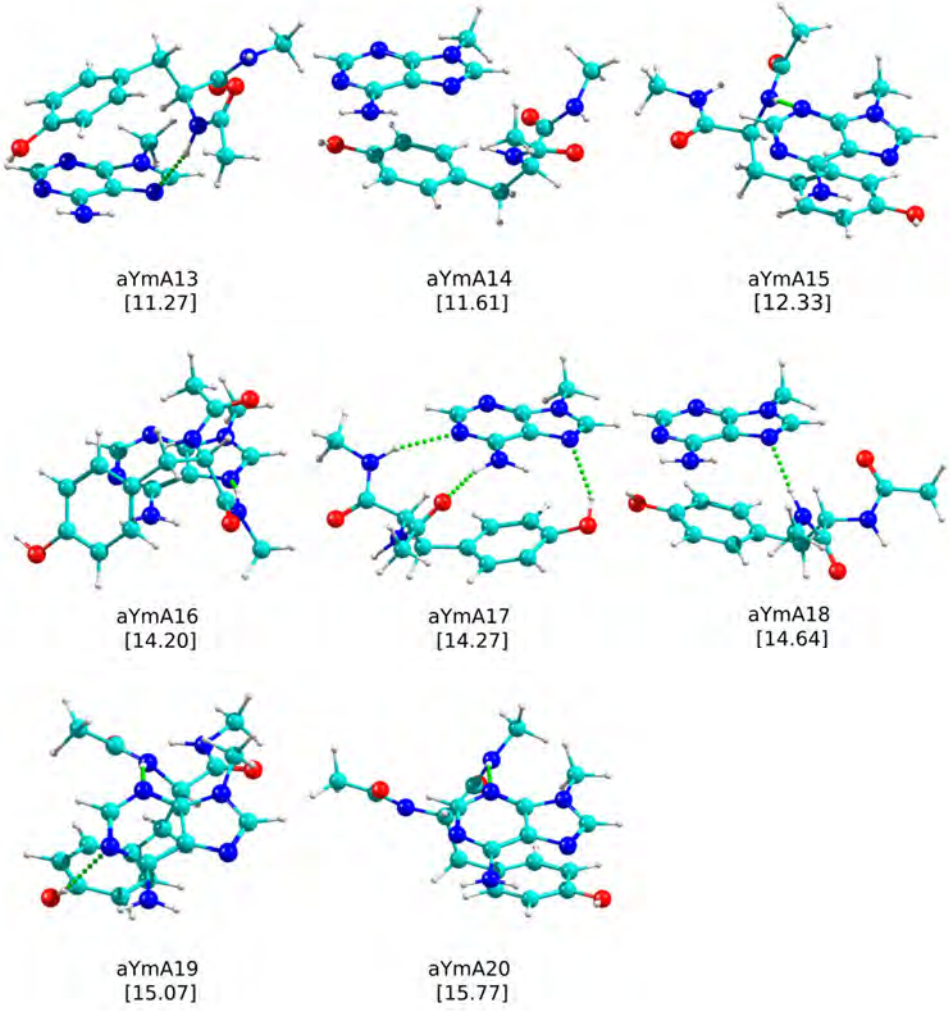
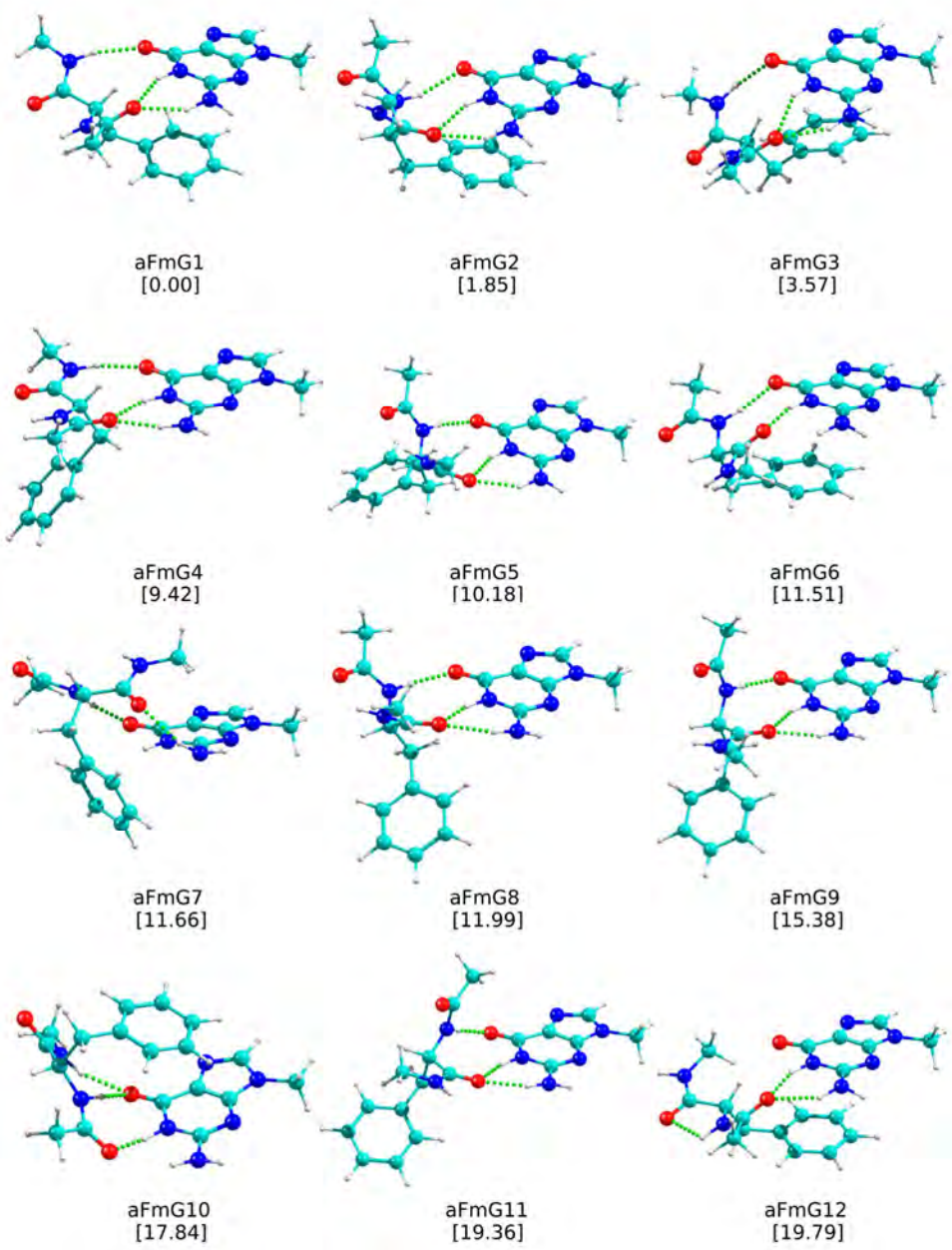

aYmA13
[11.27]
aYmA14
[11.61]
aYmA15
[12.33]
aYmA16
[14.20]
aYmA17
[14.27]
aYmA18
[14.64]
aYmA19
[15.07]
aYmA20
[15.77]

Figure 8.2.34 *The first 20 calculated structures for aFmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



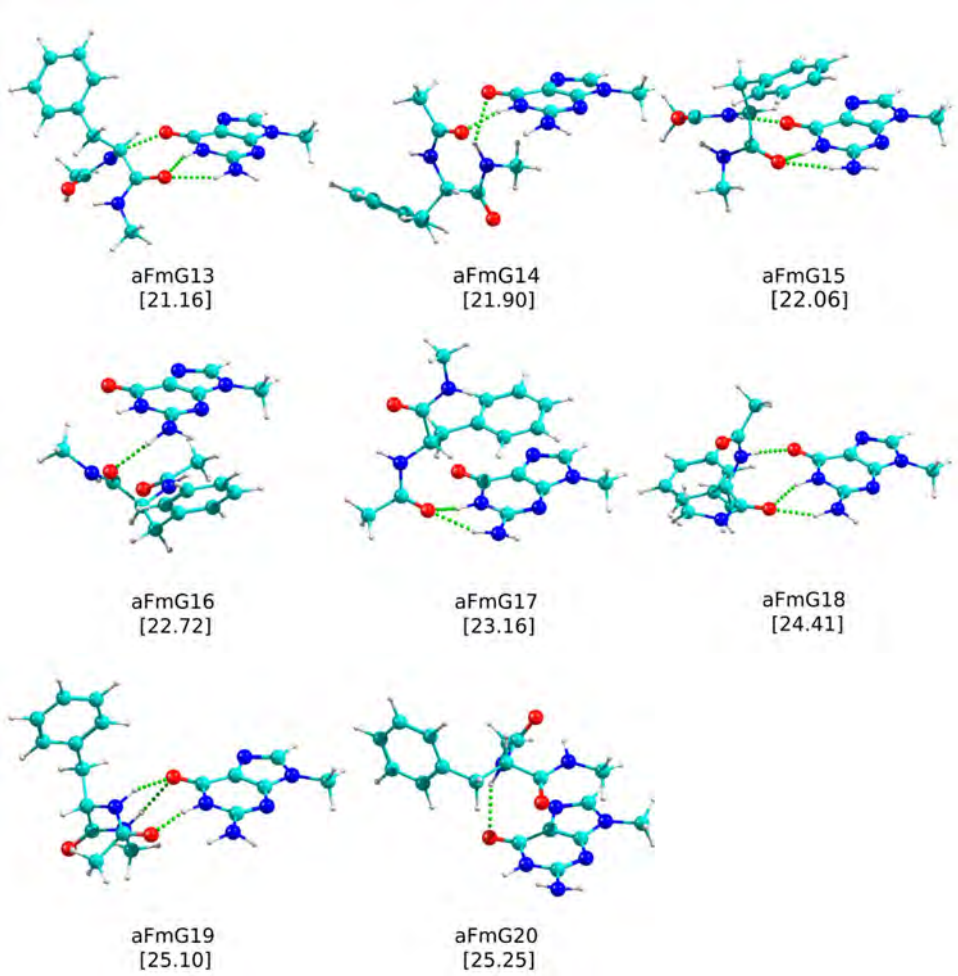
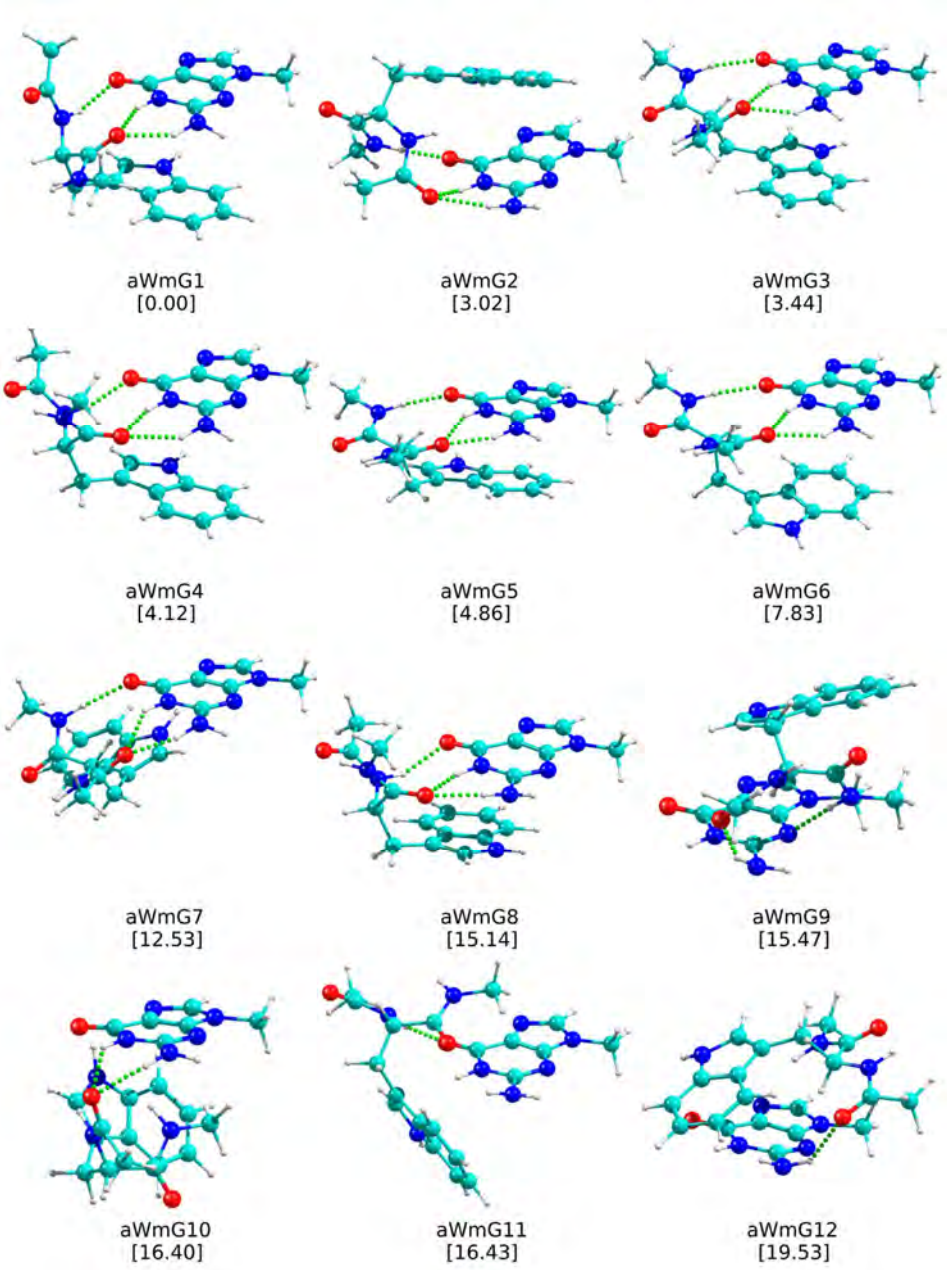
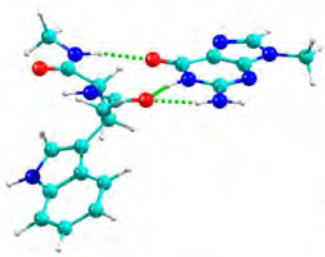
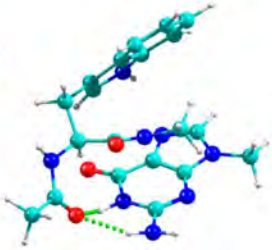

aFmG13
[21.16]
aFmG14
[21.90]
aFmG15
[22.06]
aFmG16
[22.72]
aFmG17
[23.16]
aFmG18
[24.41]
aFmG19
[25.10]
aFmG20
[25.25]

Figure 8.2.35 *The first 20 calculated structures for aWmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

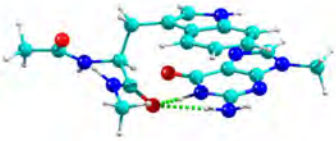




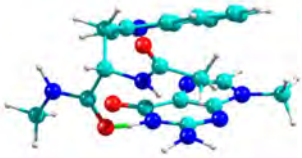
aWmG13
[19.58]



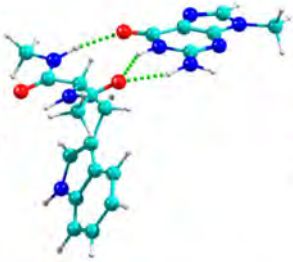
aWmG14
[19.61]



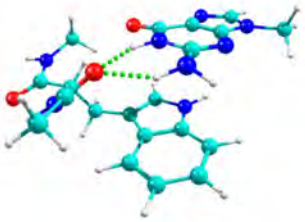
aWmG15
[20.35]



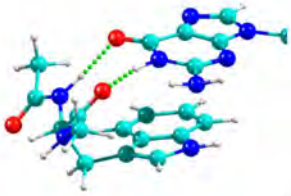
aWmG16
[21.07]



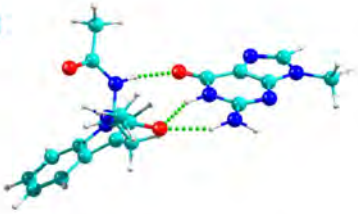
aWmG17
[21.36]



aWmG18
[21.37]

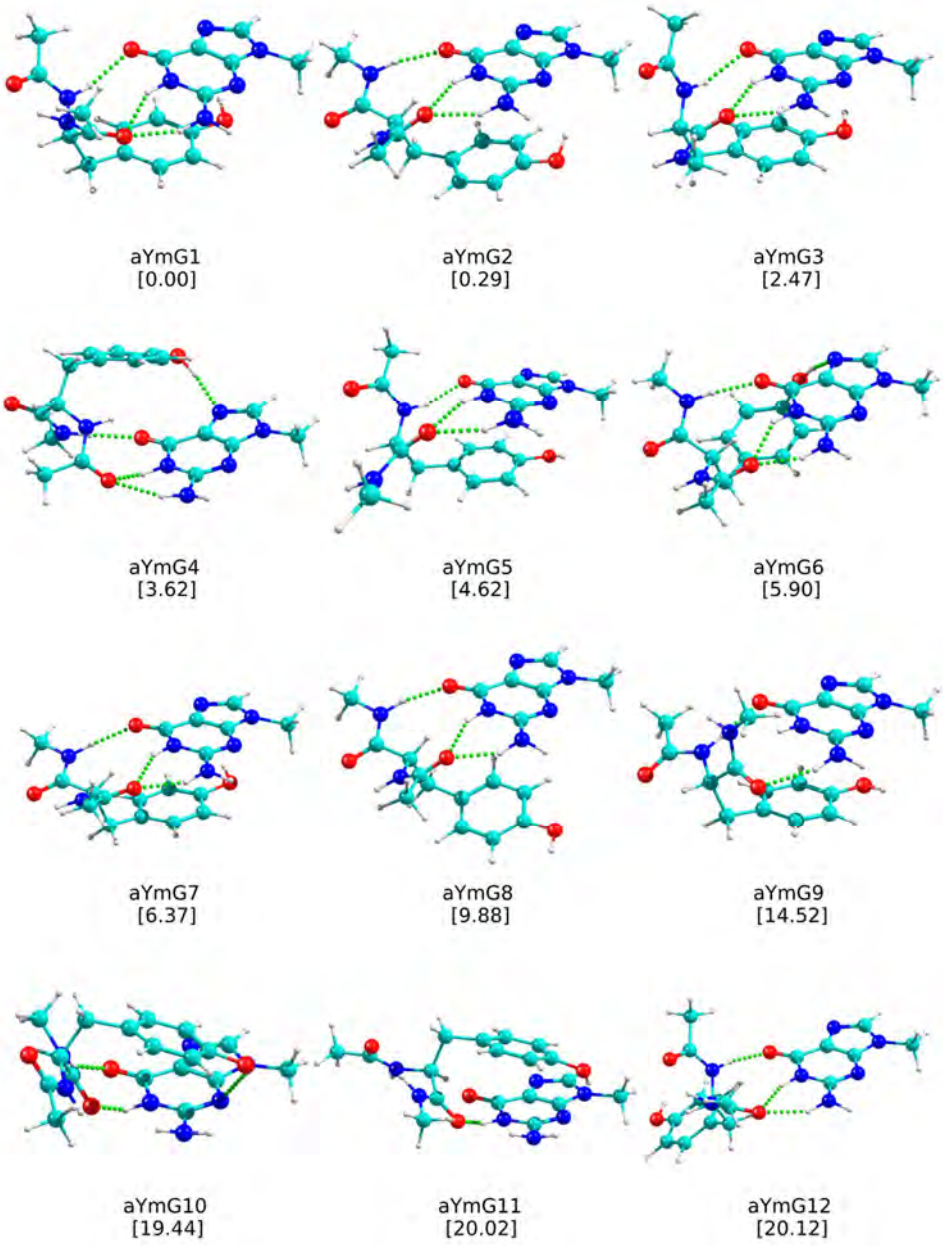


aWmG19
[22.06]



aWmG20
[23.38]

Figure 8.2.36 *The first 20 calculated structures for aYmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



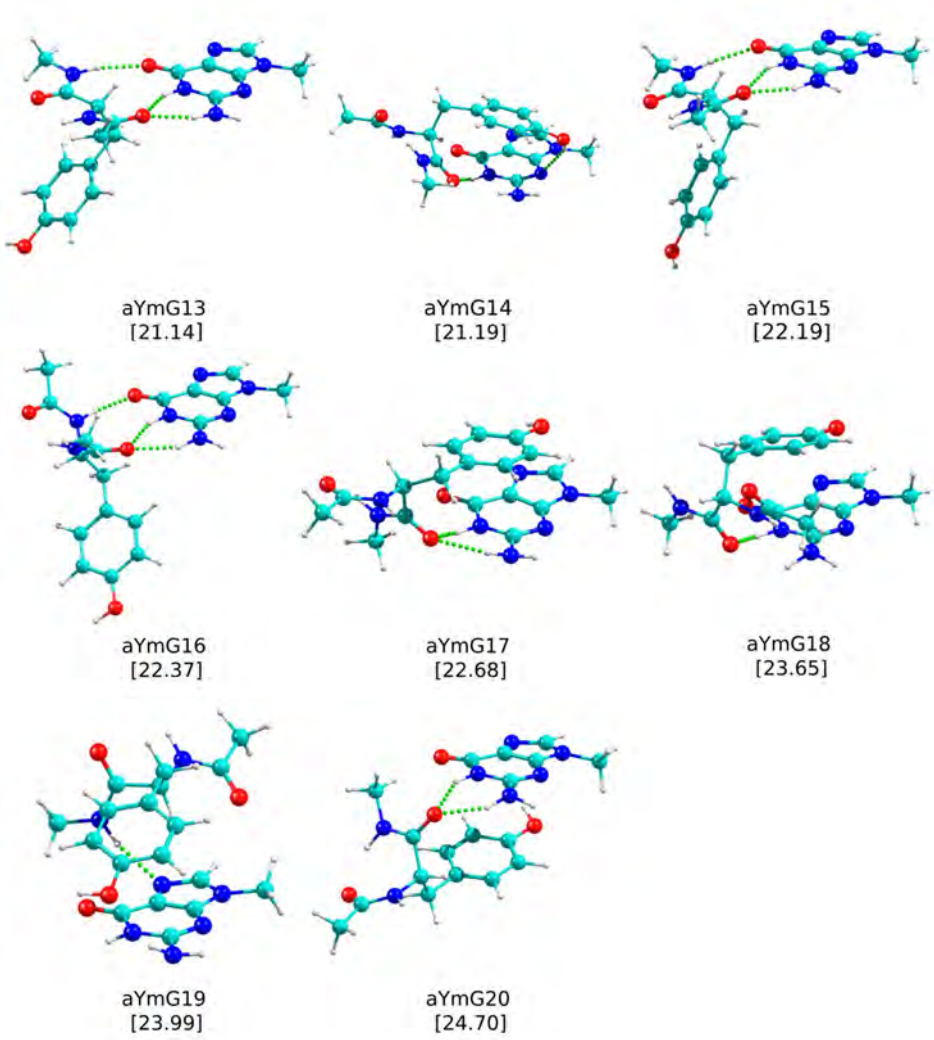

aYmG13
[21.14]
aYmG14
[21.19]
aYmG15
[22.19]
aYmG16
[22.37]
aYmG17
[22.68]
aYmG18
[23.65]
aYmG19
[23.99]
aYmG20
[24.70]

Figure 8.2.37 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aAmC, aVmC and aImC.*

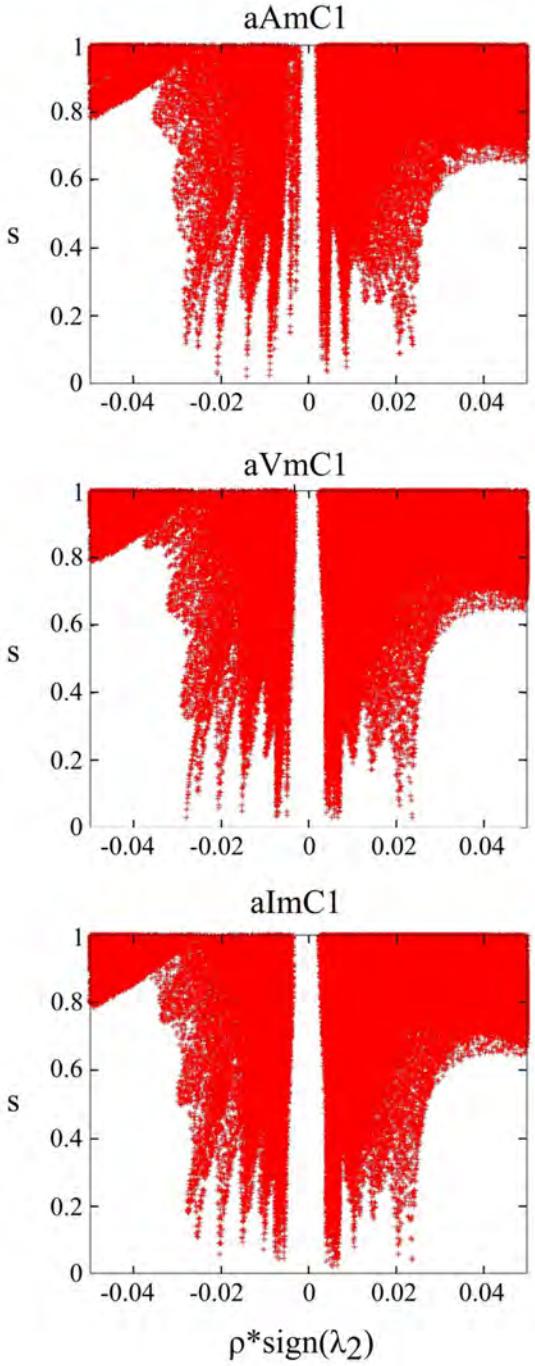


Figure 8.2.38 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aAmT, aVmT and aImT.*

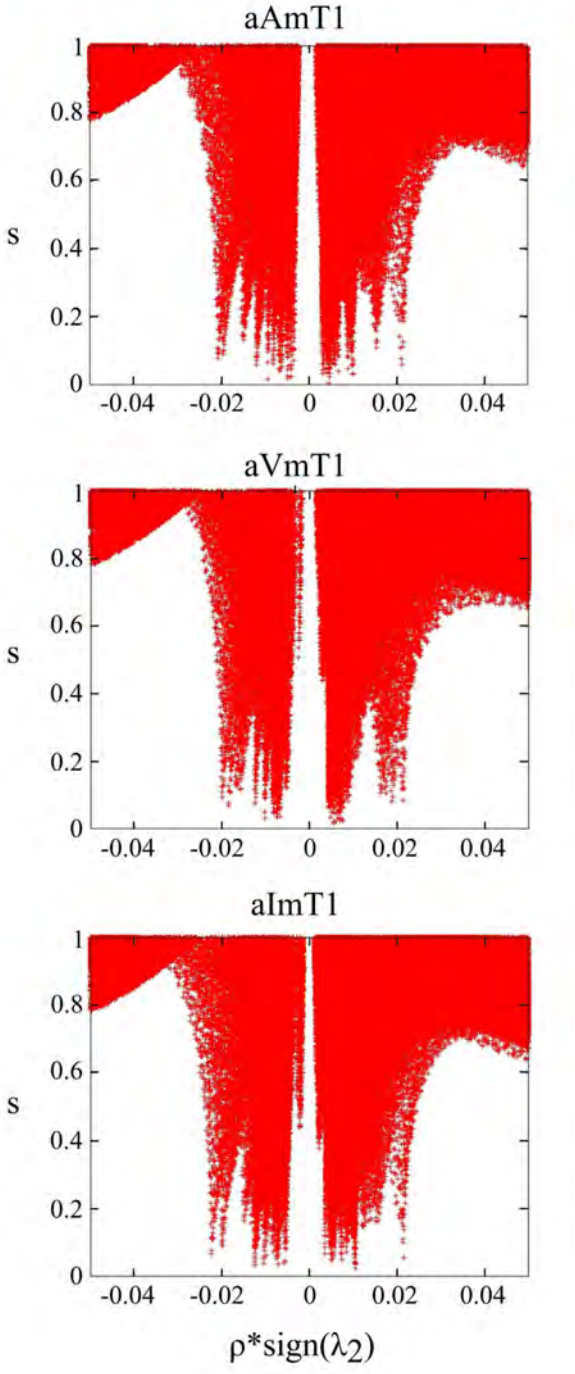


Figure 8.2.39 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aAmA, aVmA and aImA.*

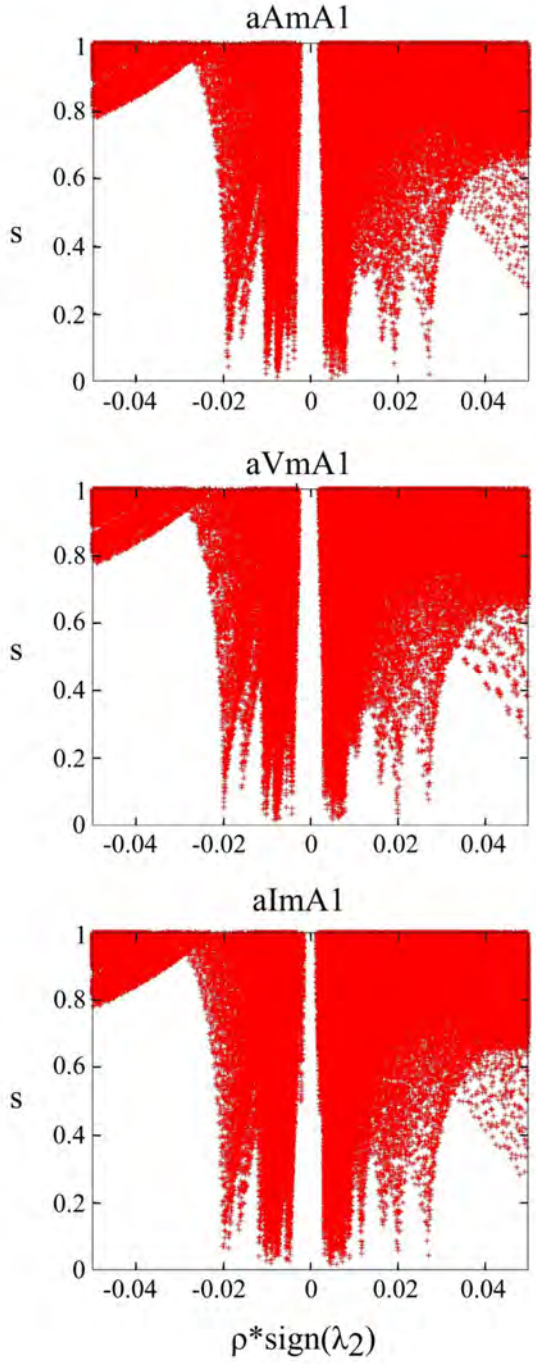


Figure 8.2.40 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aAmG, aVmG and aImG.*

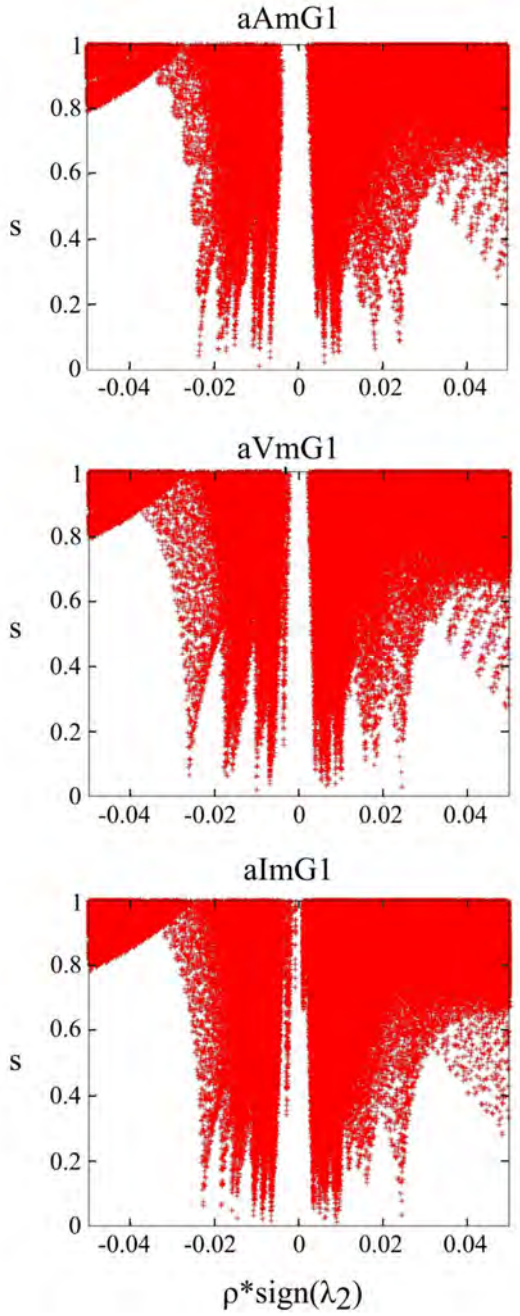


Figure 8.2.41 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aNmC, aQmC and aRmC.*

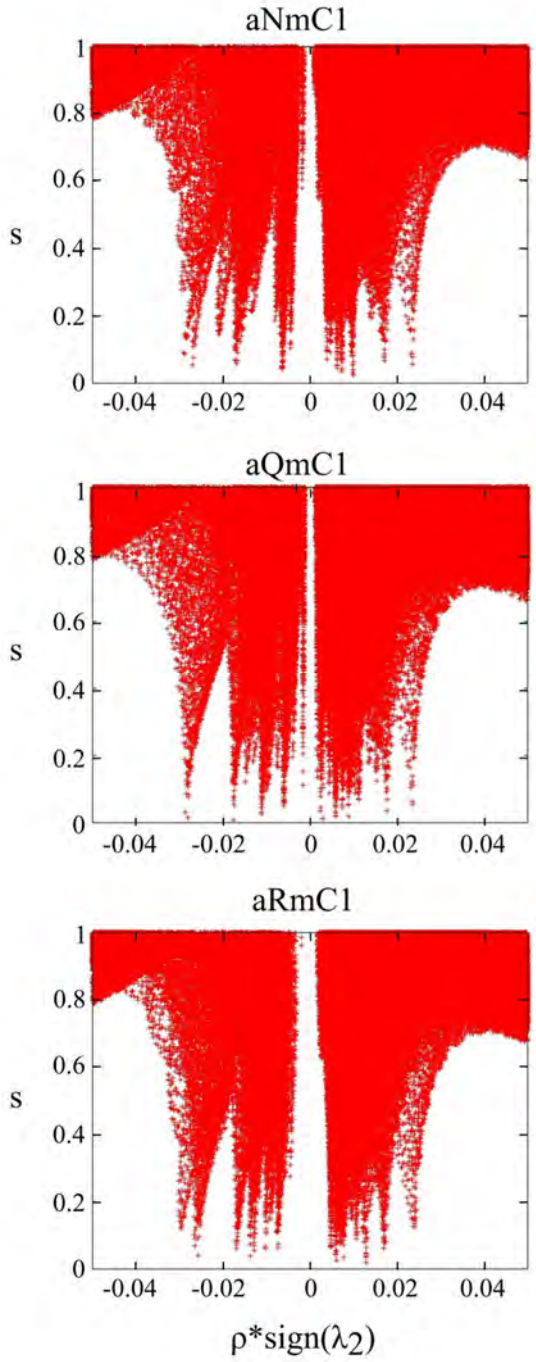


Figure 8.2.42 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aNmT, aQmT and aRmT.*

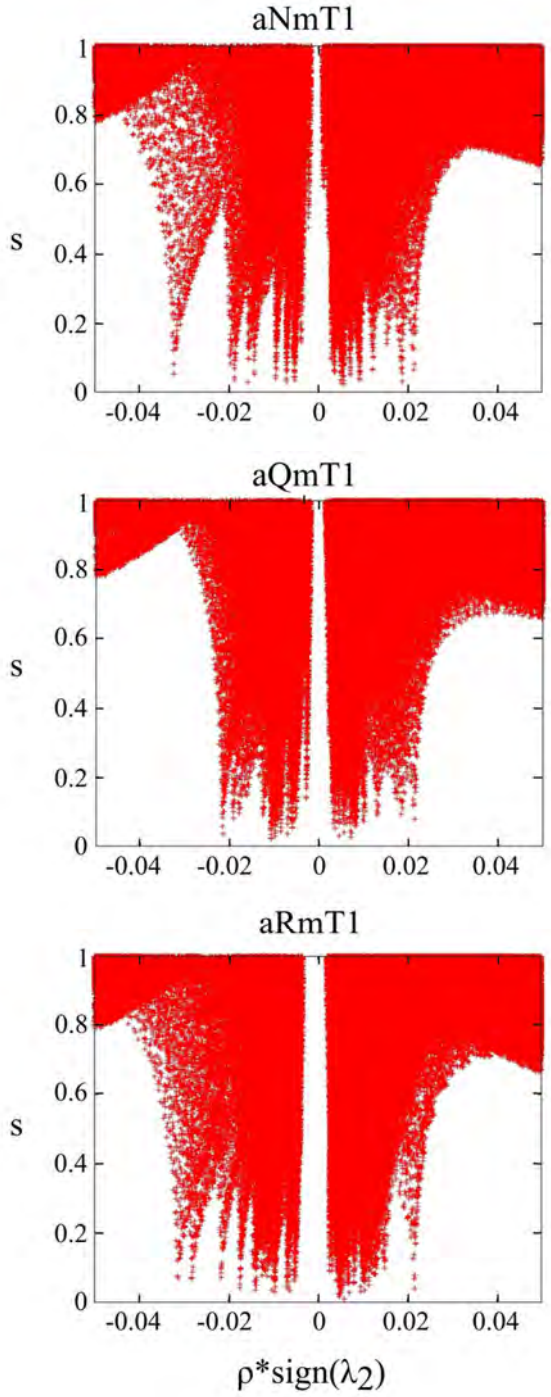


Figure 8.2.43 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aNmA, aQmA and aRmA.*

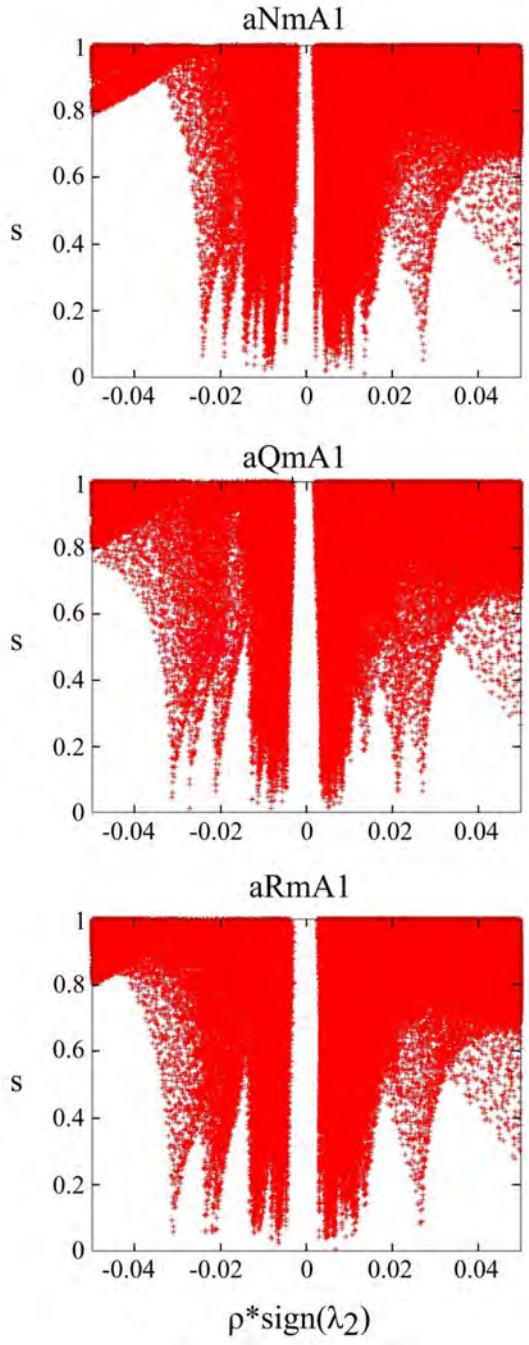


Figure 8.2.44 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aNmG, aQmG and aRmG.*

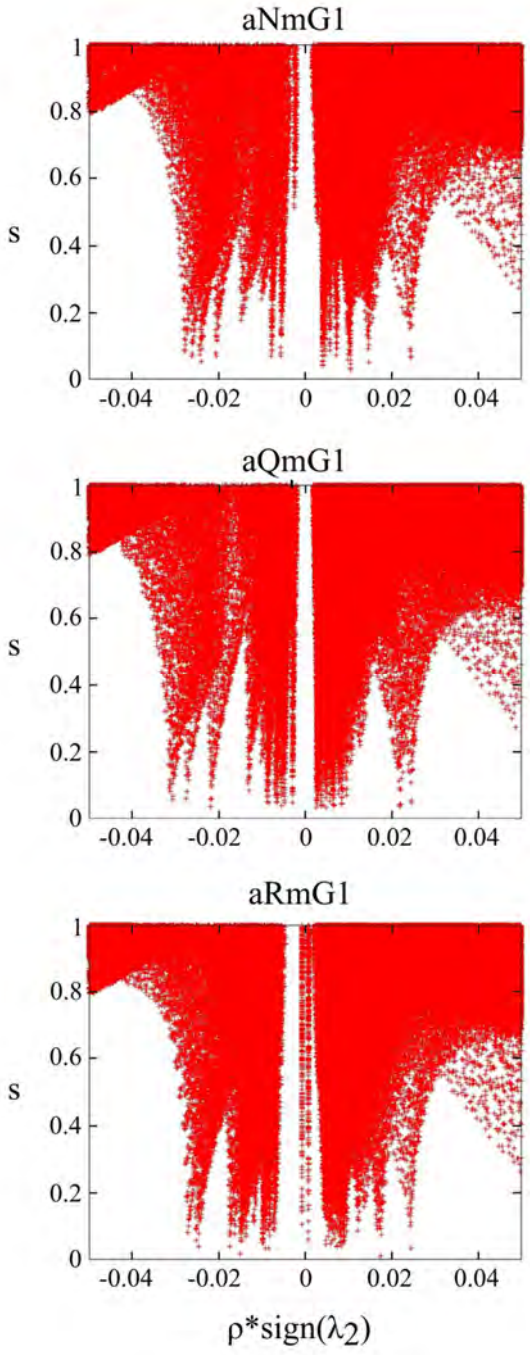


Figure 8.2.45 *2D-NCI plots (s vs- ρ*sign(λ_2)) for the most stable conformations of aFmC, aWmC and aYmC.*

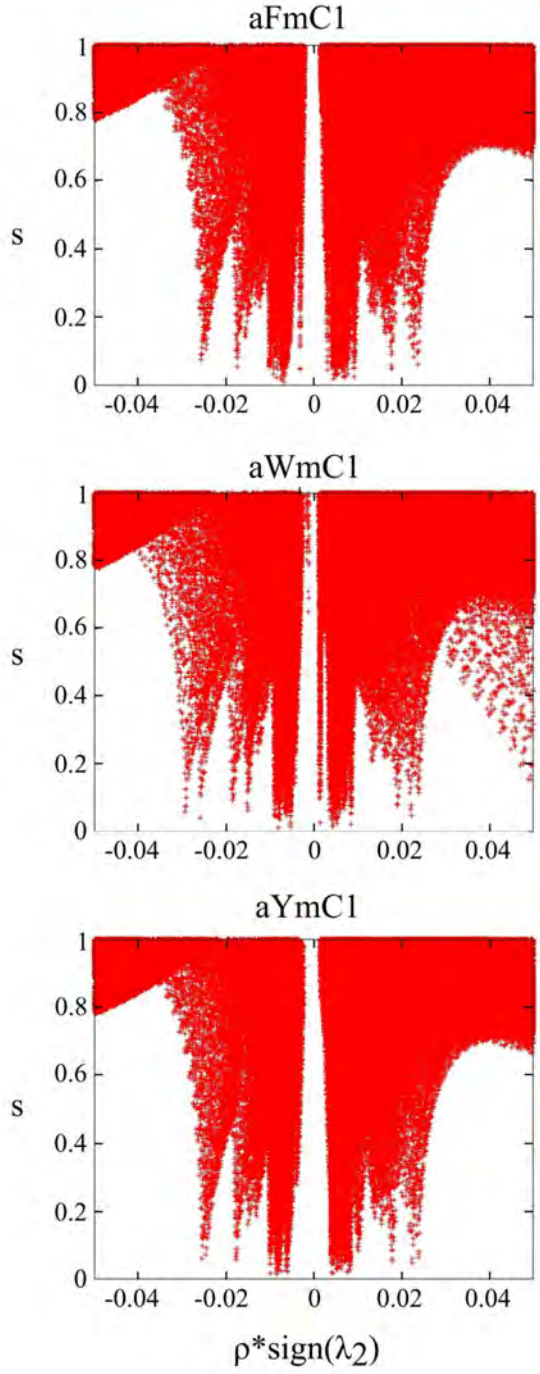


Figure 8.2.46 *2-DNCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aFmT, aWmT and aYmT.*

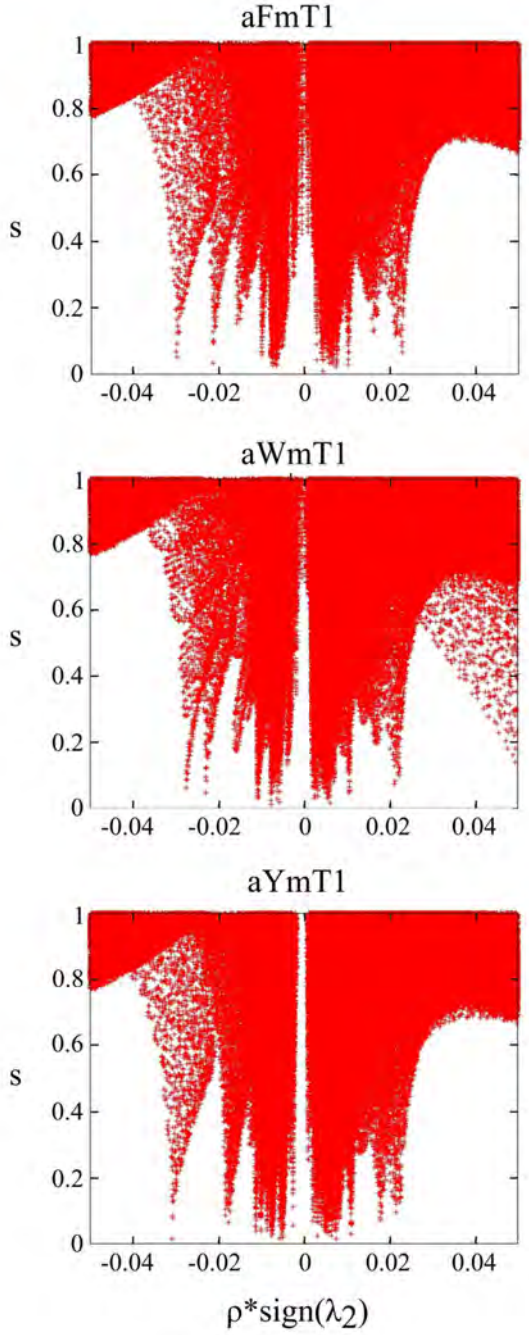


Figure 8.2.47 *2-DNCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aFmA, aWmA and aYmA.*

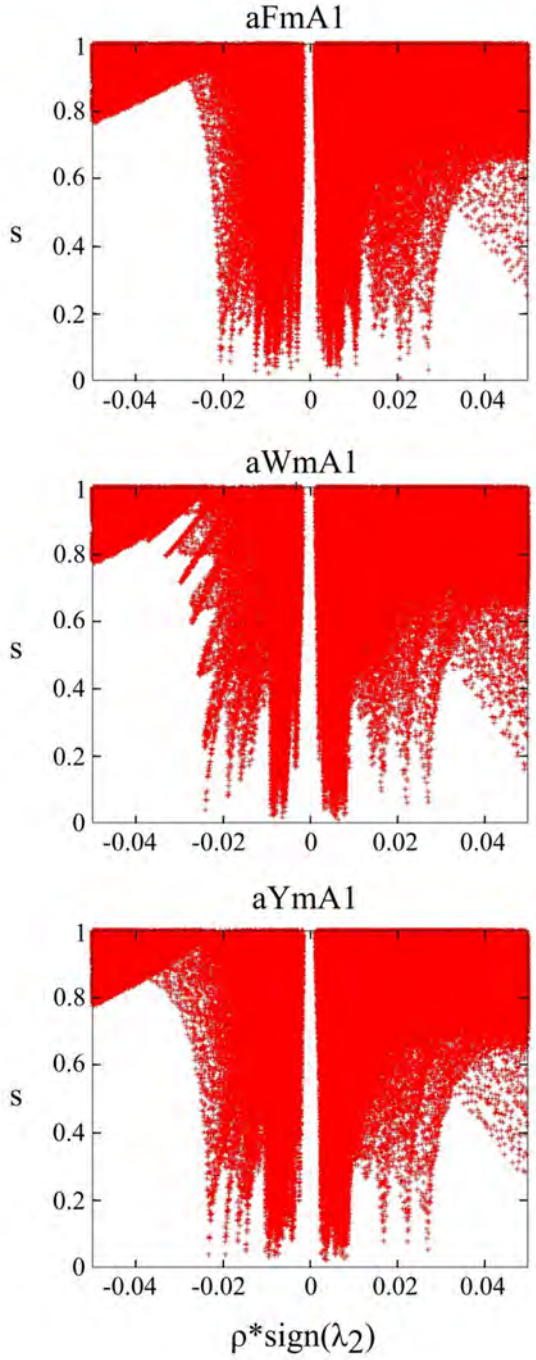


Figure 8.2.48 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aFmG, aWmG and aYmG.*

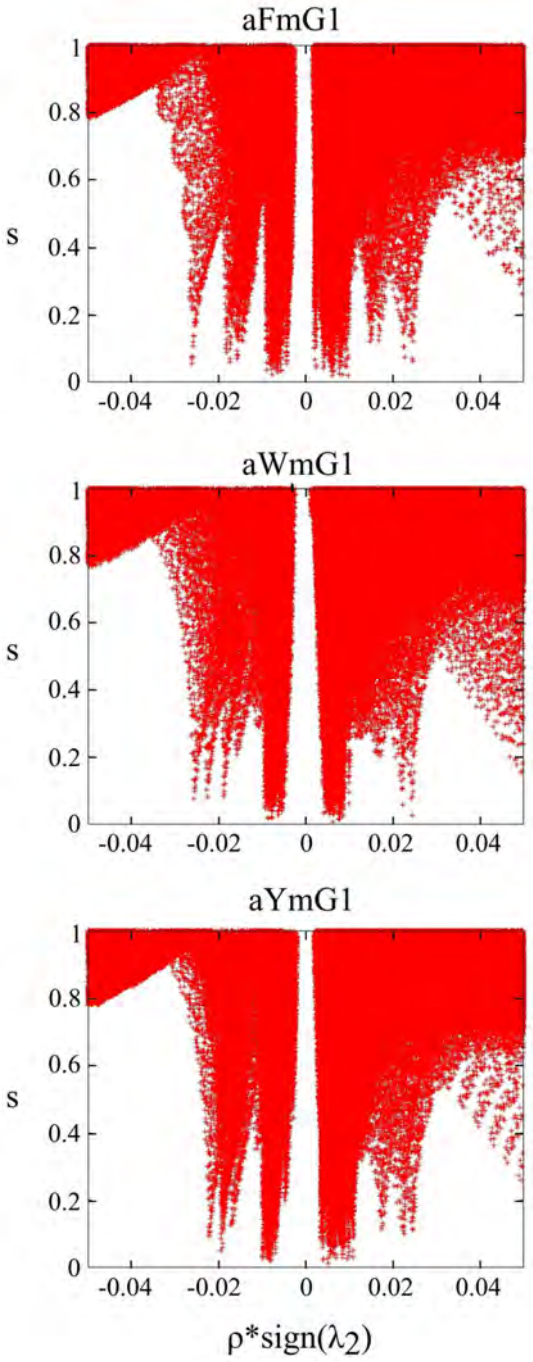
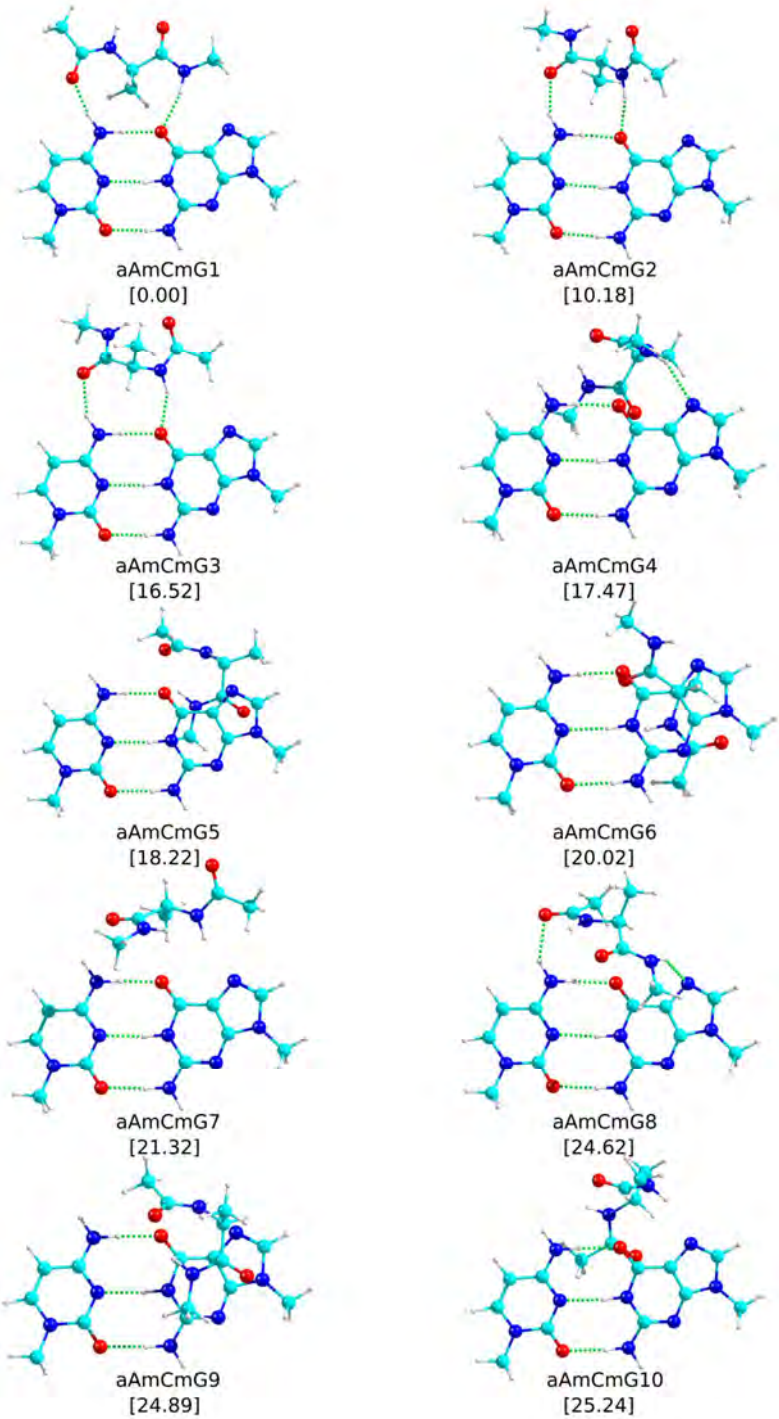
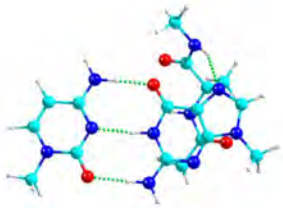
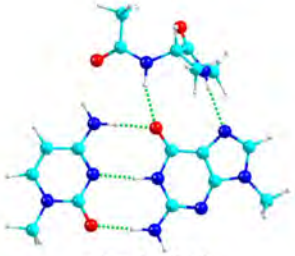


Figure 8.2.49 *The first 20 calculated structures for aAmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

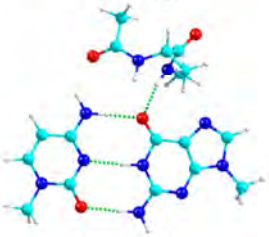




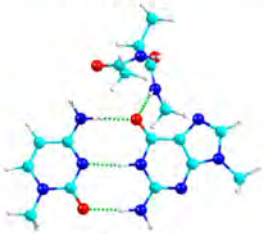
aAmCmG11
[25.32]



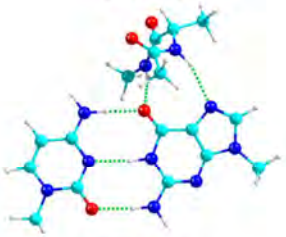
aAmCmG12
[26.29]



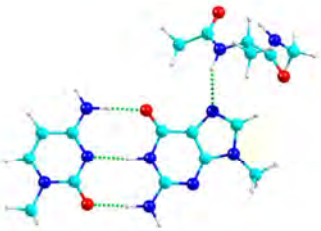
aAmCmG13
[26.31]



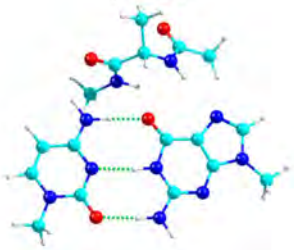
aAmCmG14
[28.34]



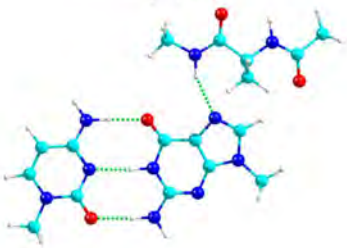
aAmCmG15
[29.03]



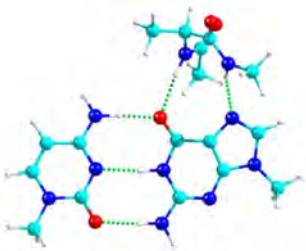
aAmCmG16
[29.25]



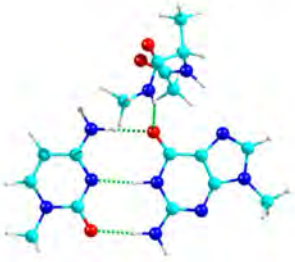
aAmCmG17
[29.67]



aAmCmG18
[30.74]

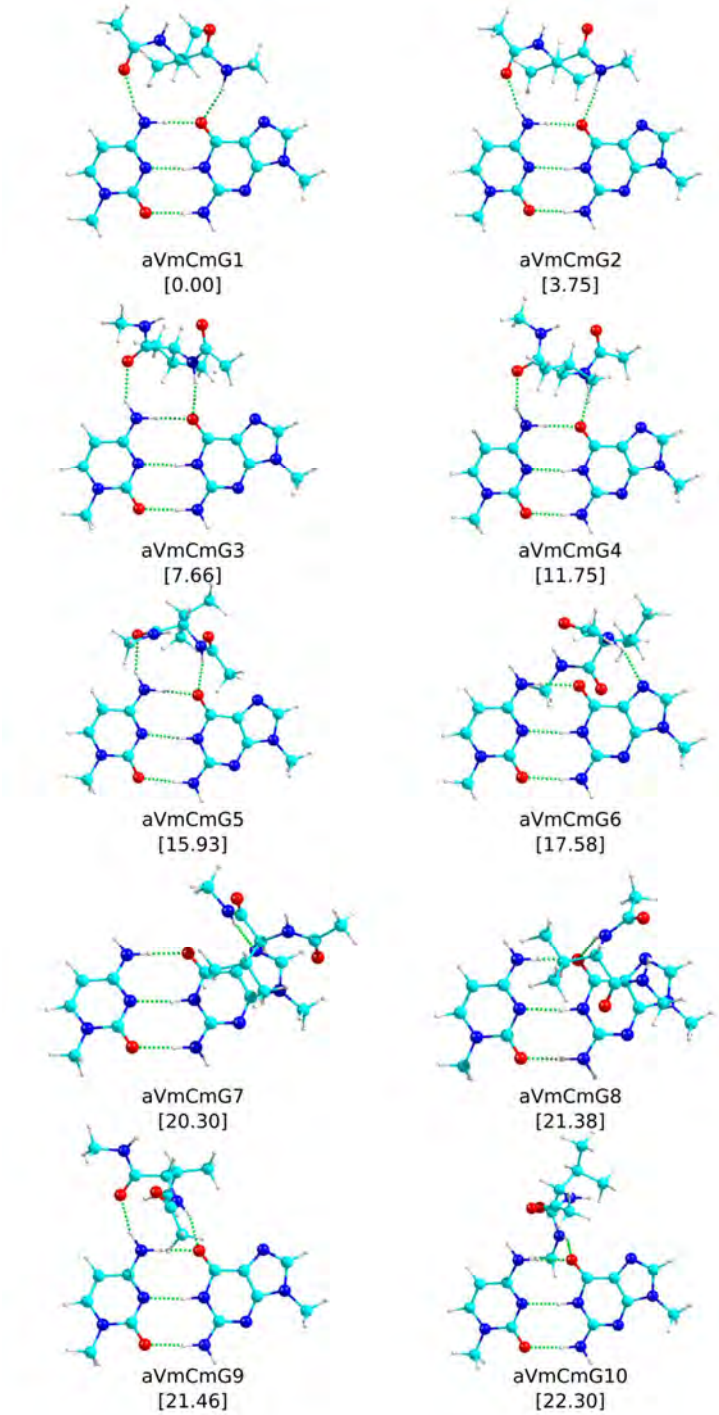


aAmCmG19
[34.46]



aAmCmG20
[35.04]

Figure 8.2.50 *The first 20 calculated structures for aVmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



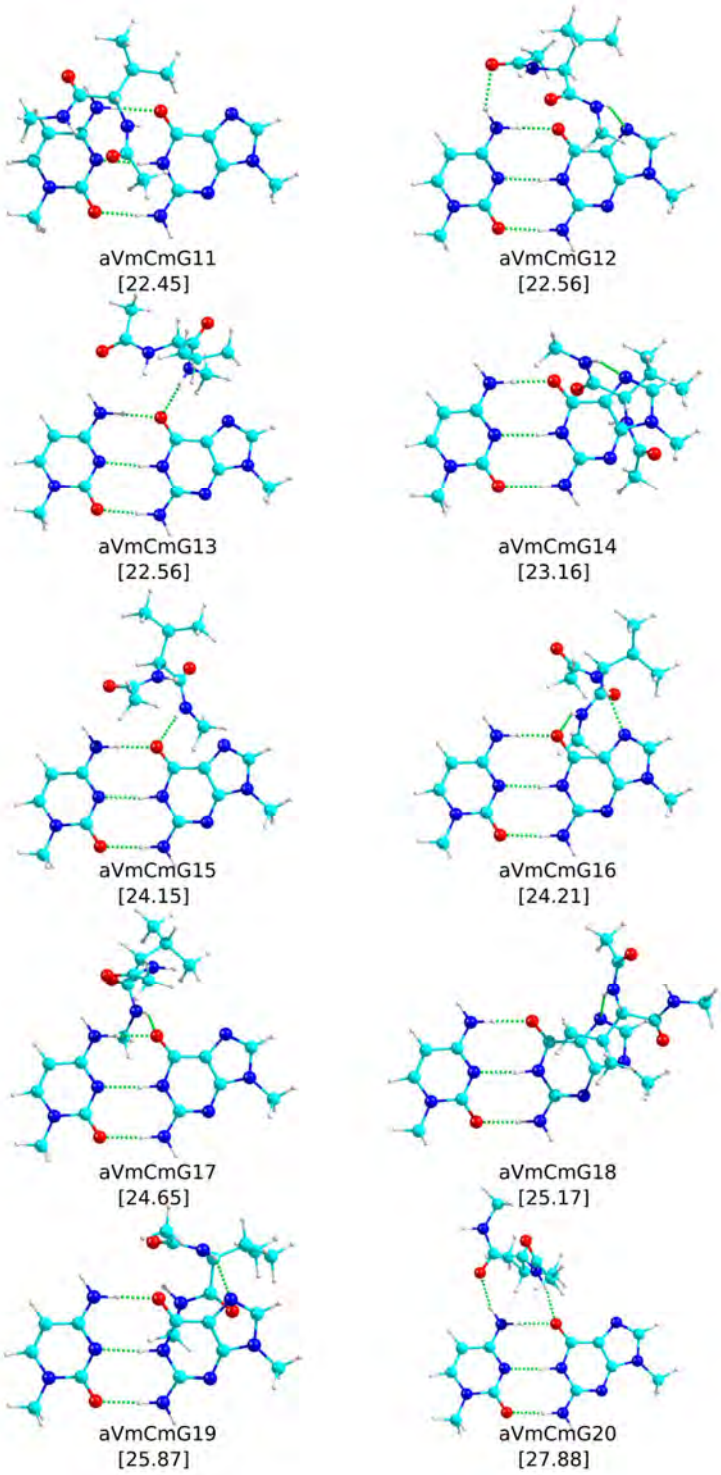
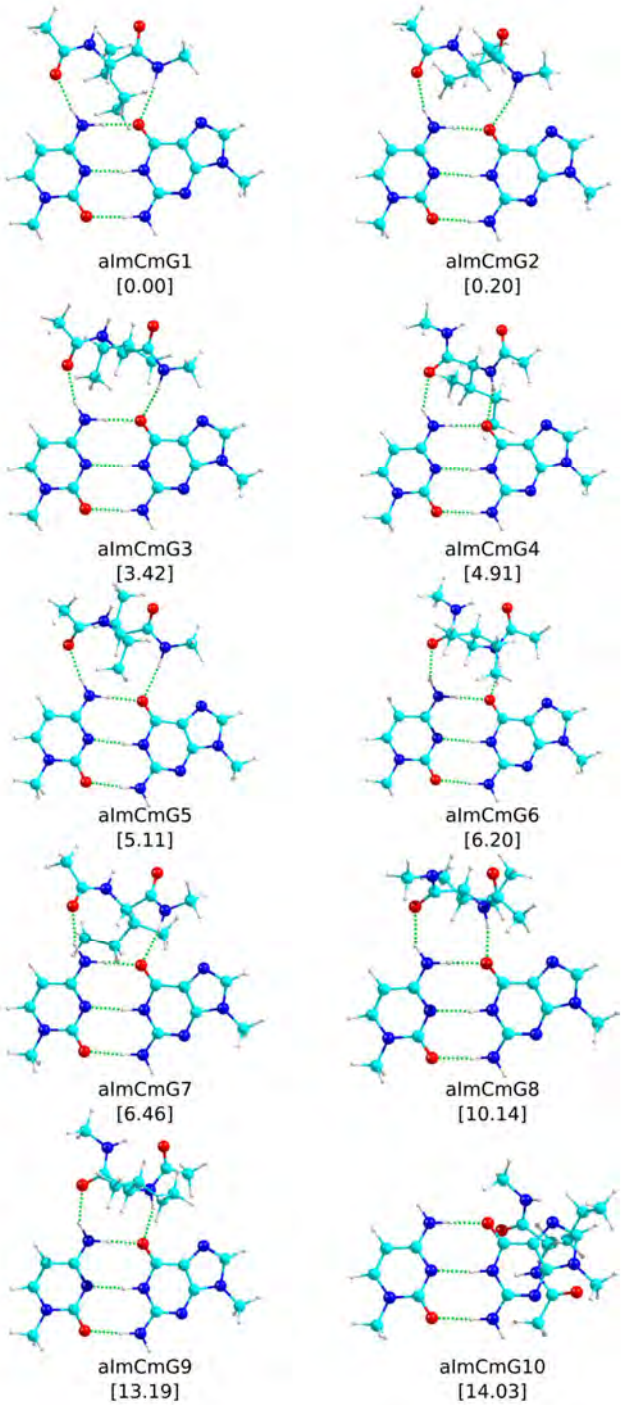

aVmCmG11
[22.45]
aVmCmG12
[22.56]
aVmCmG13
[22.56]
aVmCmG14
[23.16]
aVmCmG15
[24.15]
aVmCmG16
[24.21]
aVmCmG17
[24.65]
aVmCmG18
[25.17]
aVmCmG19
[25.87]
aVmCmG20
[27.88]

Figure 8.2.51 *The first 20 calculated structures for aImCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



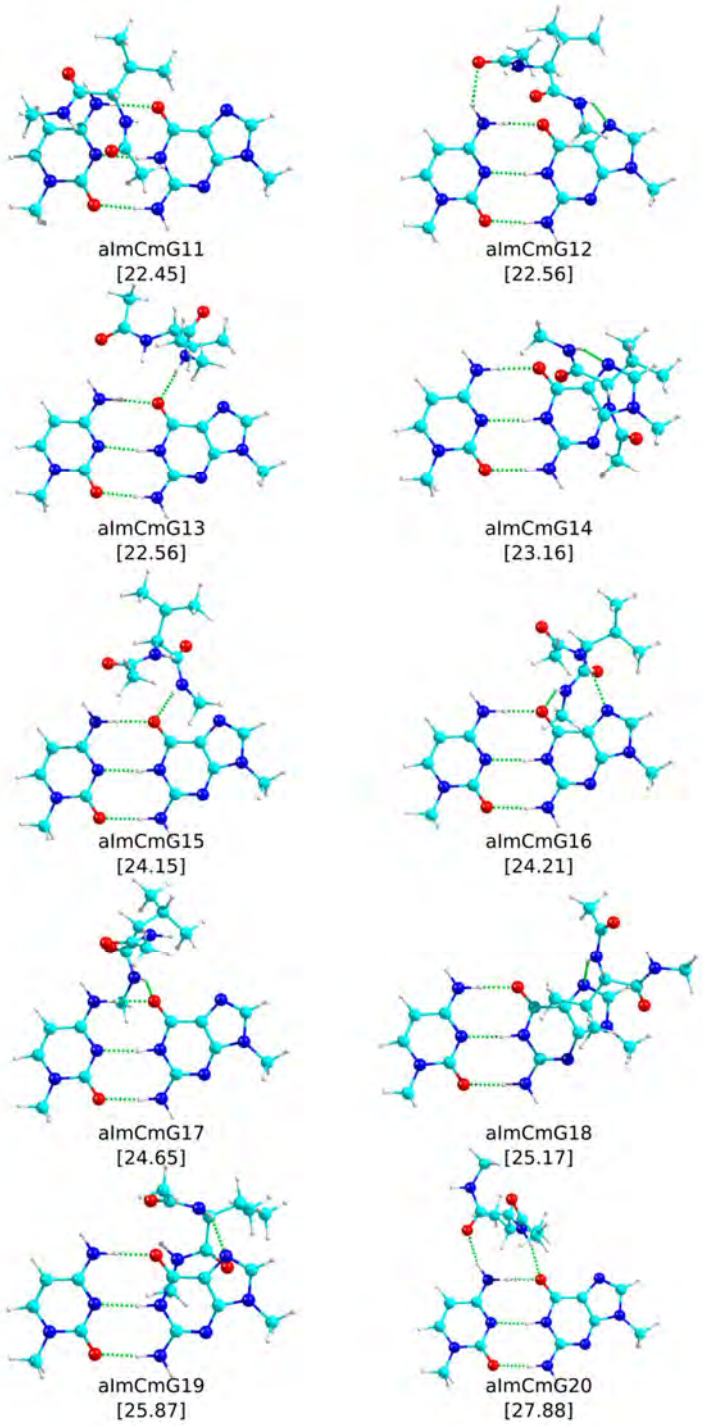
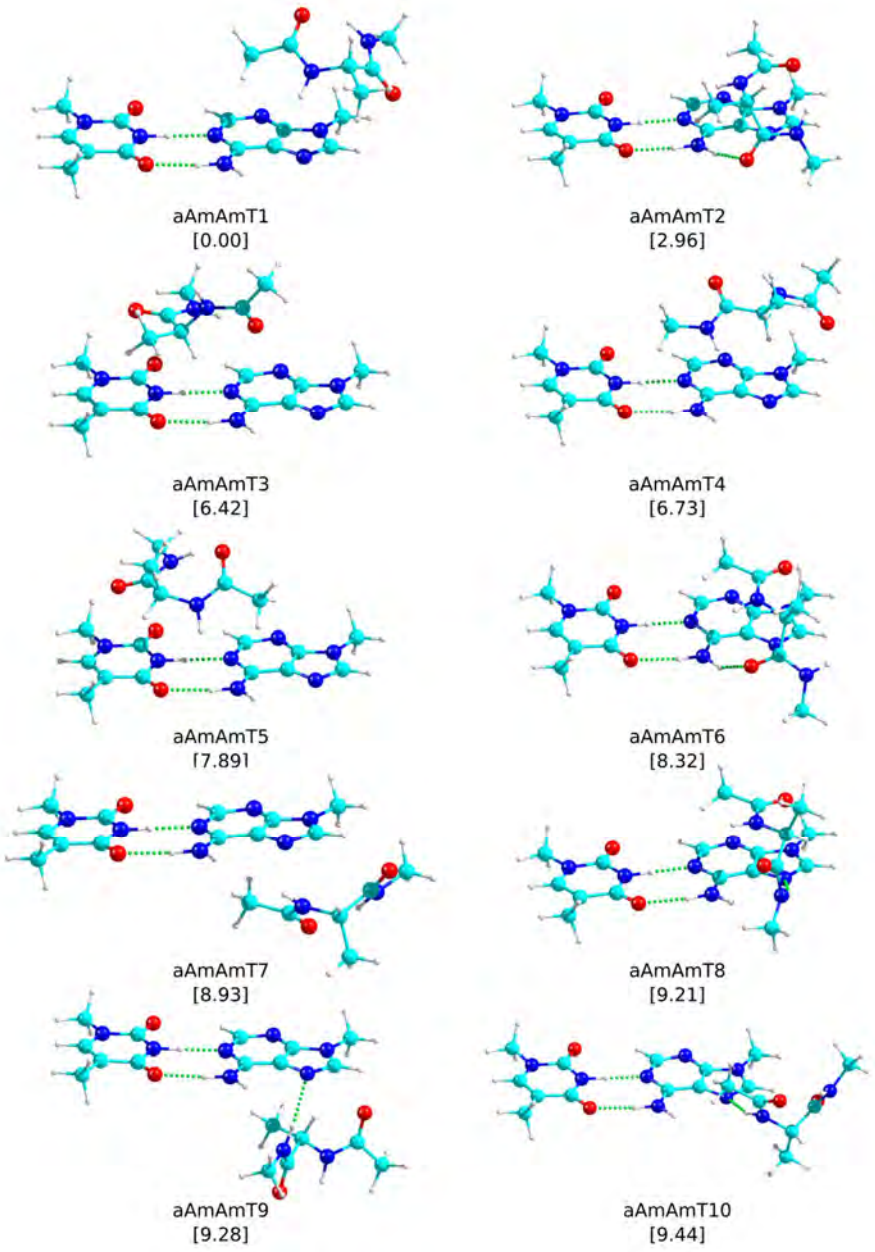
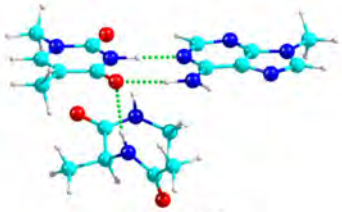
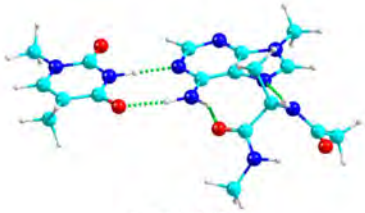

aImCmG11
[22.45]
aImCmG12
[22.56]
aImCmG13
[22.56]
aImCmG14
[23.16]
aImCmG15
[24.15]
aImCmG16
[24.21]
aImCmG17
[24.65]
aImCmG18
[25.17]
aImCmG19
[25.87]
aImCmG20
[27.88]

Figure 8.2.52 *The first 20 calculated structures for aAmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

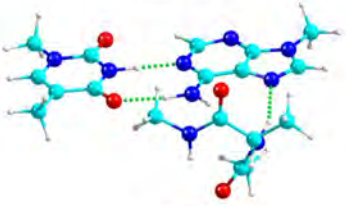




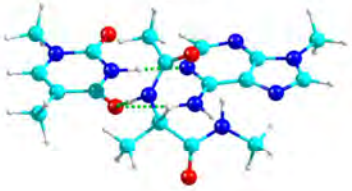
aAmAmT11
[11.77]



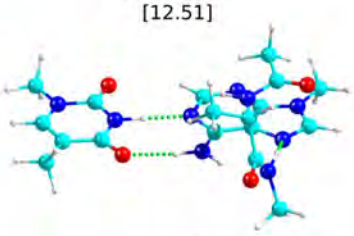
aAmAmT12
[12.30]



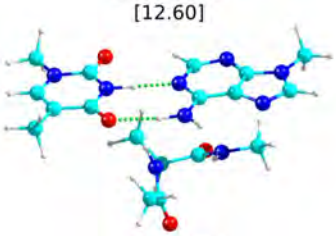
aAmAmT13
[12.51]



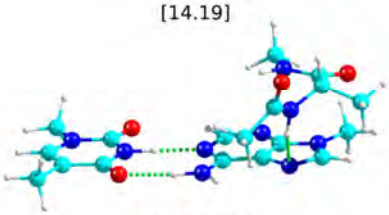
aAmAmT14
[12.60]



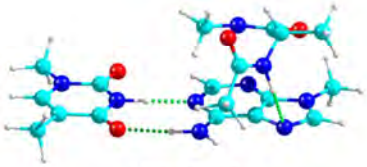
aAmAmT15
[14.19]



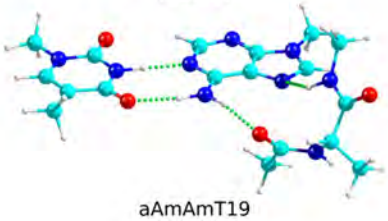
aAmAmT16
[15.16]



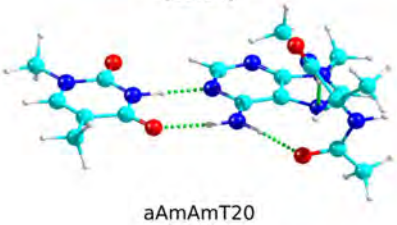
aAmAmT17
[15.23]



aAmAmT18
[15.84]

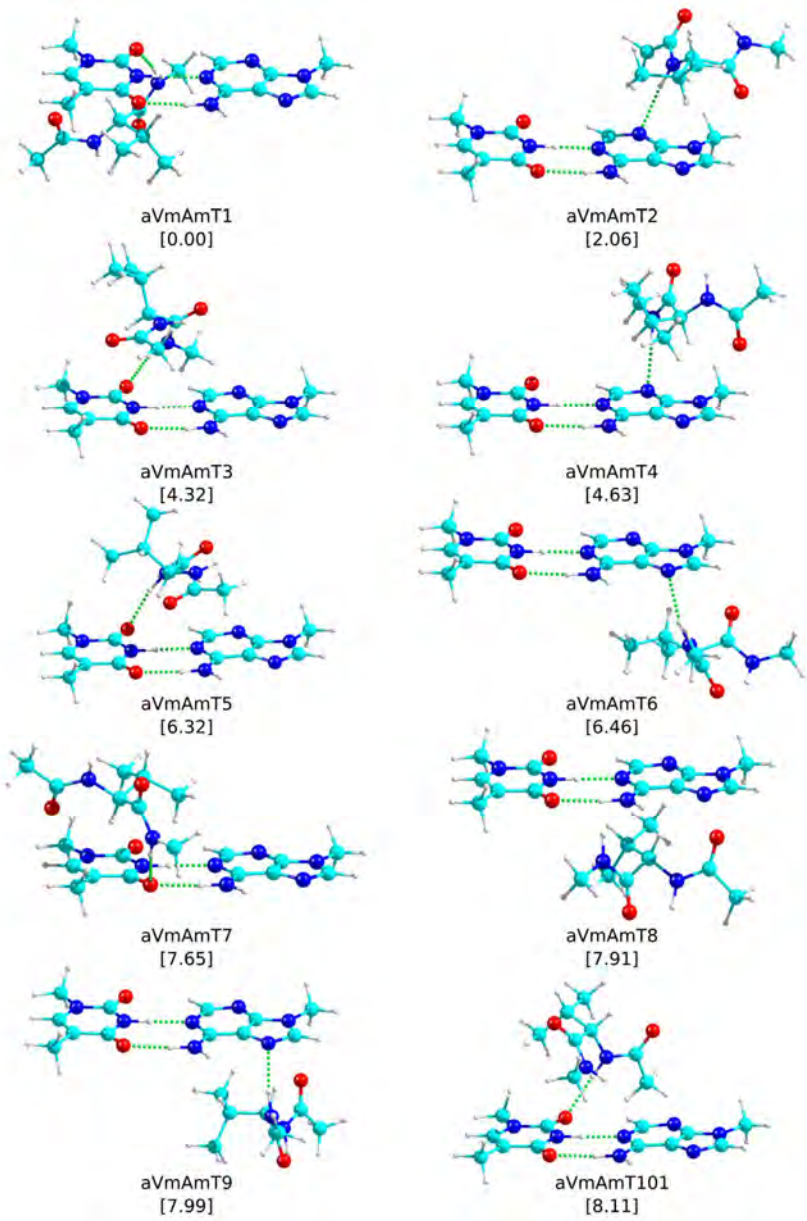


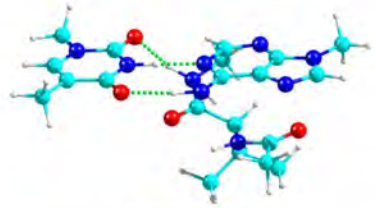
aAmAmT19
[15.98]



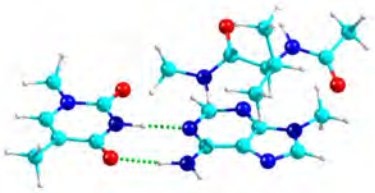
aAmAmT20
[15.98]

Figure 8.2.53 *The first 20 calculated structures for aVmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

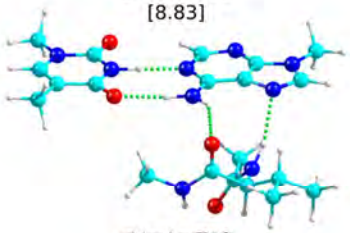




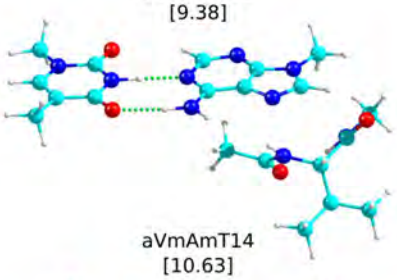
aVmAmT11
[8.83]



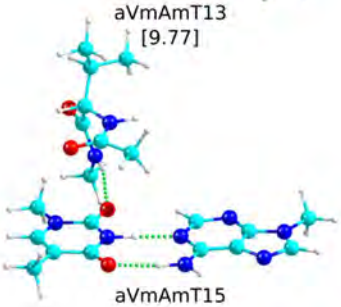
aVmAmT12
[9.38]



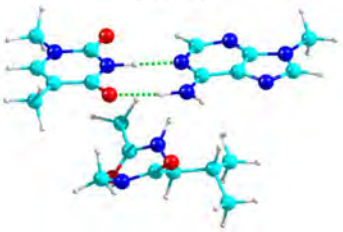
aVmAmT13
[9.77]



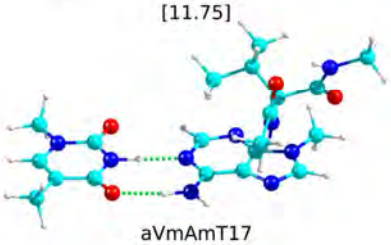
aVmAmT14
[10.63]



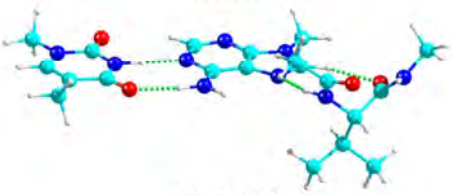
aVmAmT15
[11.75]



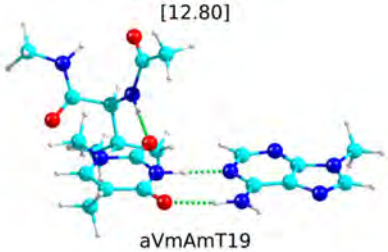
aVmAmT16
[12.01]



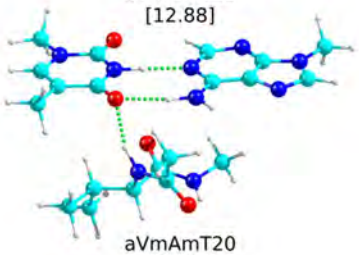
aVmAmT17
[12.80]



aVmAmT18
[12.88]

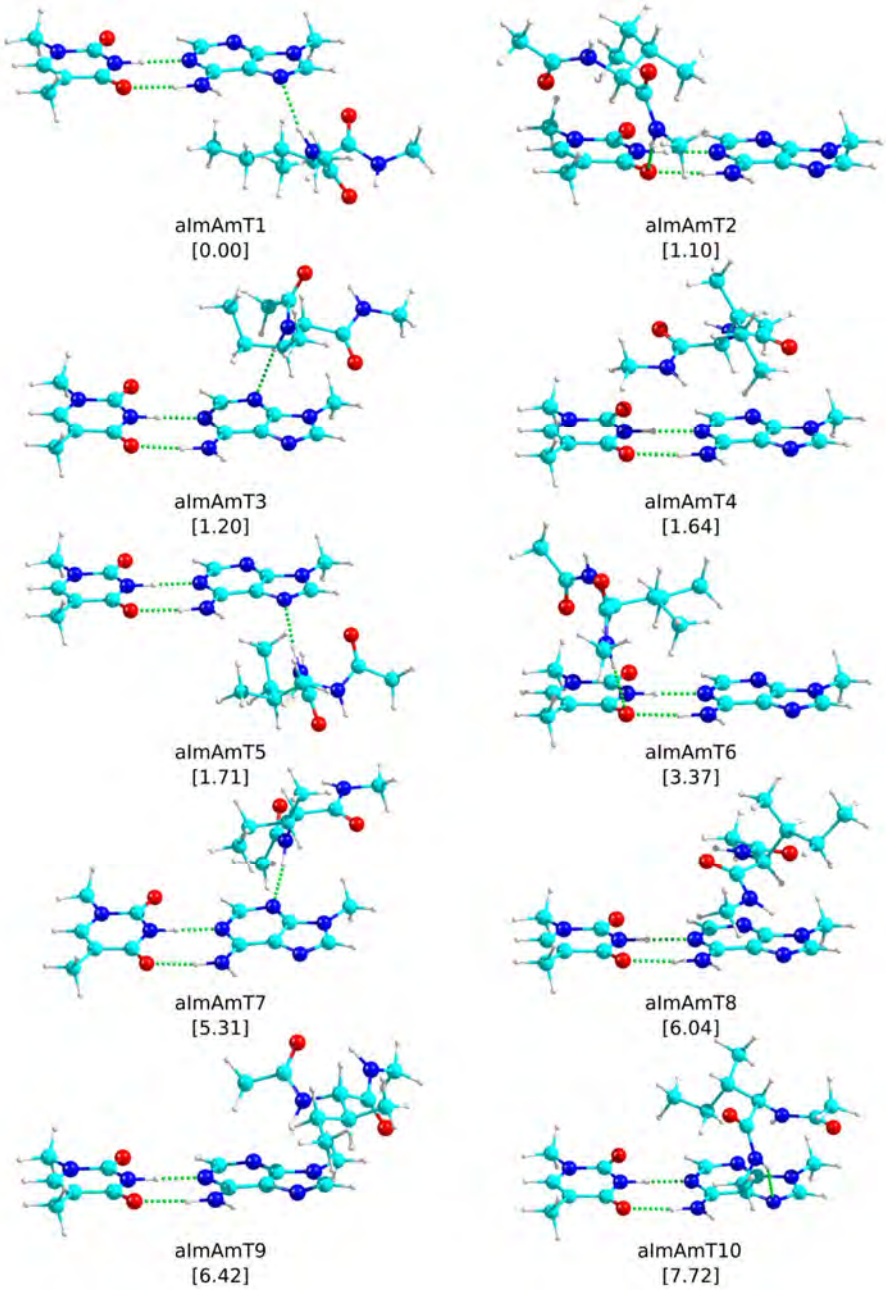


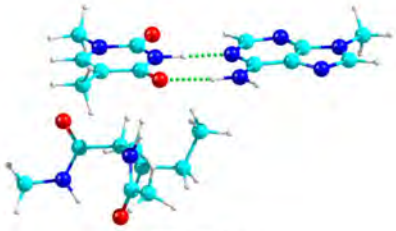
aVmAmT19
[12.95]



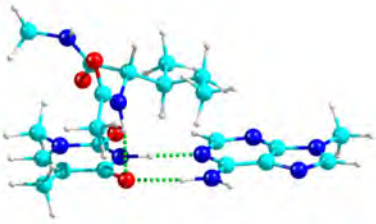
aVmAmT20
[13.40]

Figure 8.2.54 *The first 20 calculated structures for aImAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

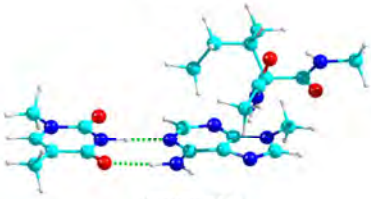




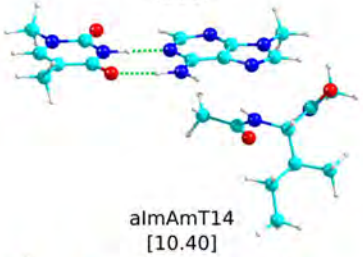
aImAmT11
[8.10]



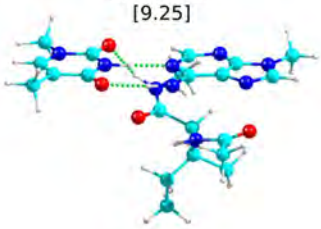
aImAmT12
[8.22]



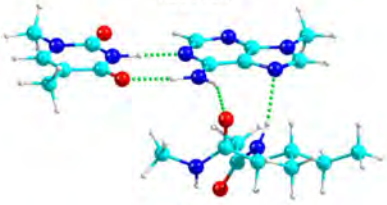
aImAmT13
[9.25]



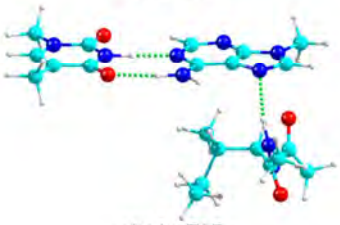
aImAmT14
[10.40]



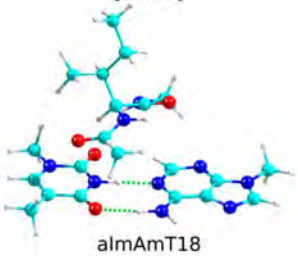
aImAmT15
[10.46]



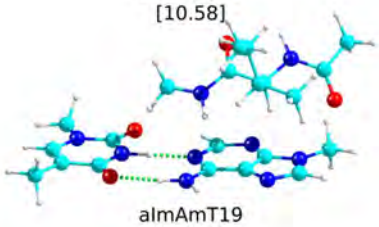
aImAmT16
[10.49]



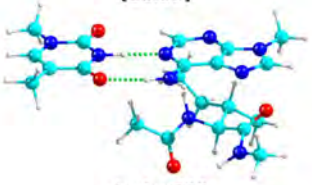
aImAmT17
[10.58]



aImAmT18
[10.68]

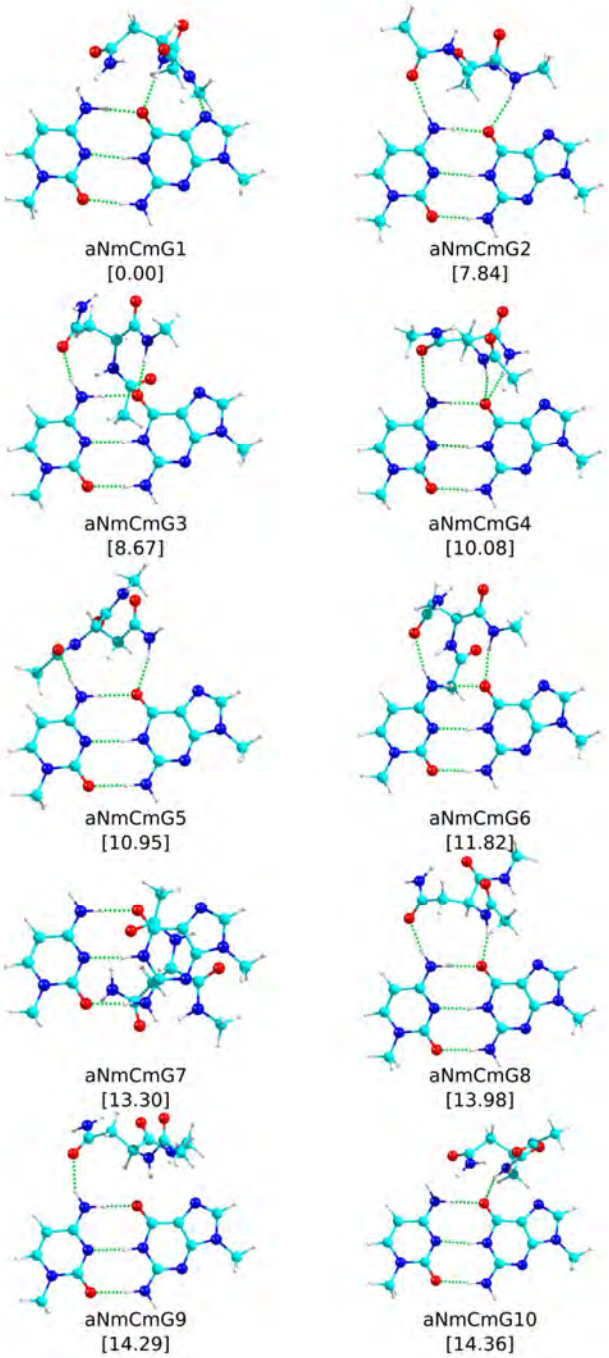


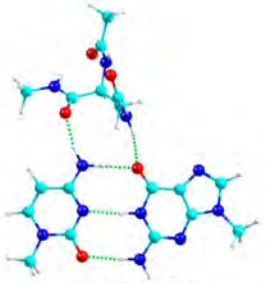
aImAmT19
[10.74]



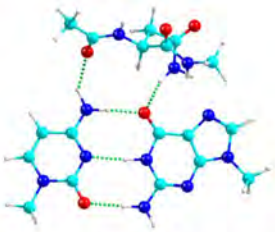
aImAmT20
[10.86]

Figure 8.2.55 *The first 20 calculated structures for aNmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

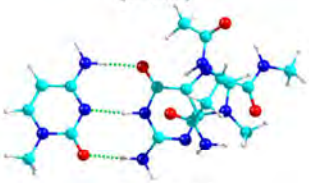




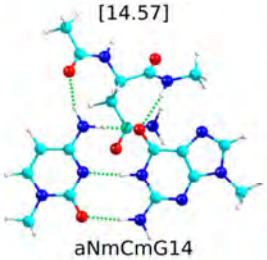
aNmCmG11
[14.51]



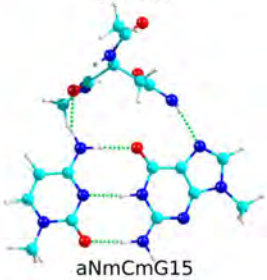
aNmCmG12
[14.57]



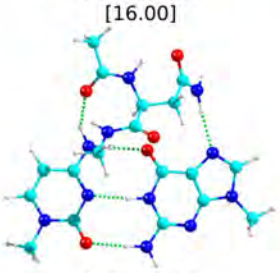
aNmCmG13
[15.96]



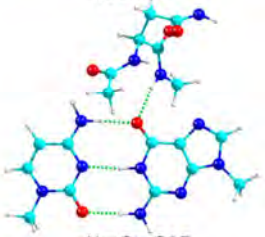
aNmCmG14
[16.00]



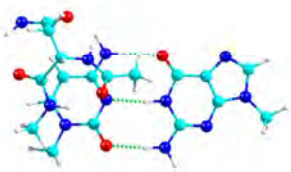
aNmCmG15
[16.29]



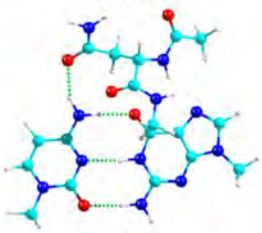
aNmCmG16
[16.40]



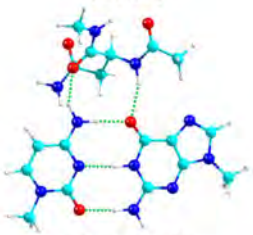
aNmCmG17
[16.62]



aNmCmG18
[16.85]

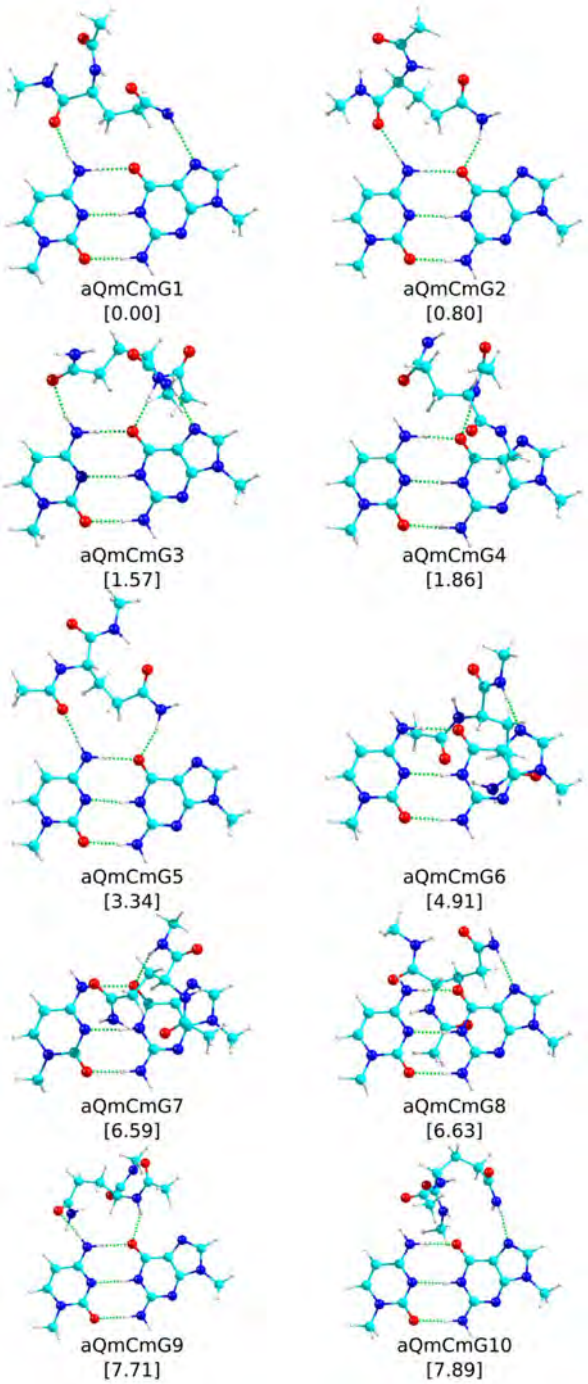


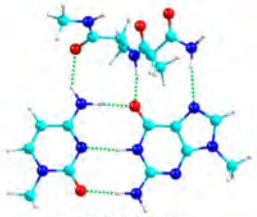
aNmCmG19
[17.30]



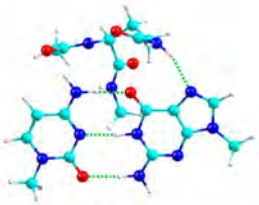
aNmCmG20
[17.52]

Figure 8.2.56 *The first 20 calculated structures for aQmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

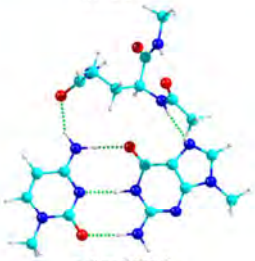




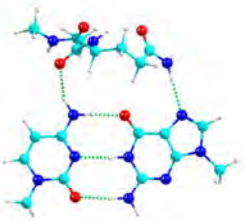
aQmCG11
[7.93]



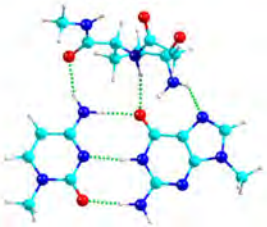
aQmCG12
[8.21]



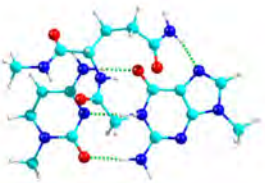
aQmCG13
[10.38]



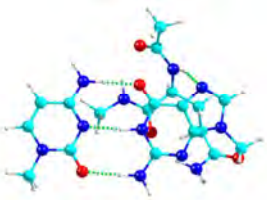
aQmCG14
[11.00]



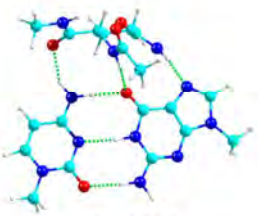
aQmCG15
[11.54]



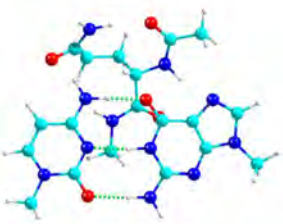
aQmCG16
[12.06]



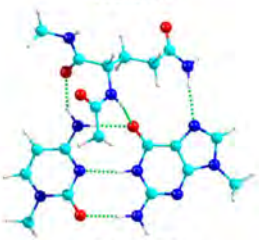
aQmCG17
[13.27]



aQmCG18
[13.57]

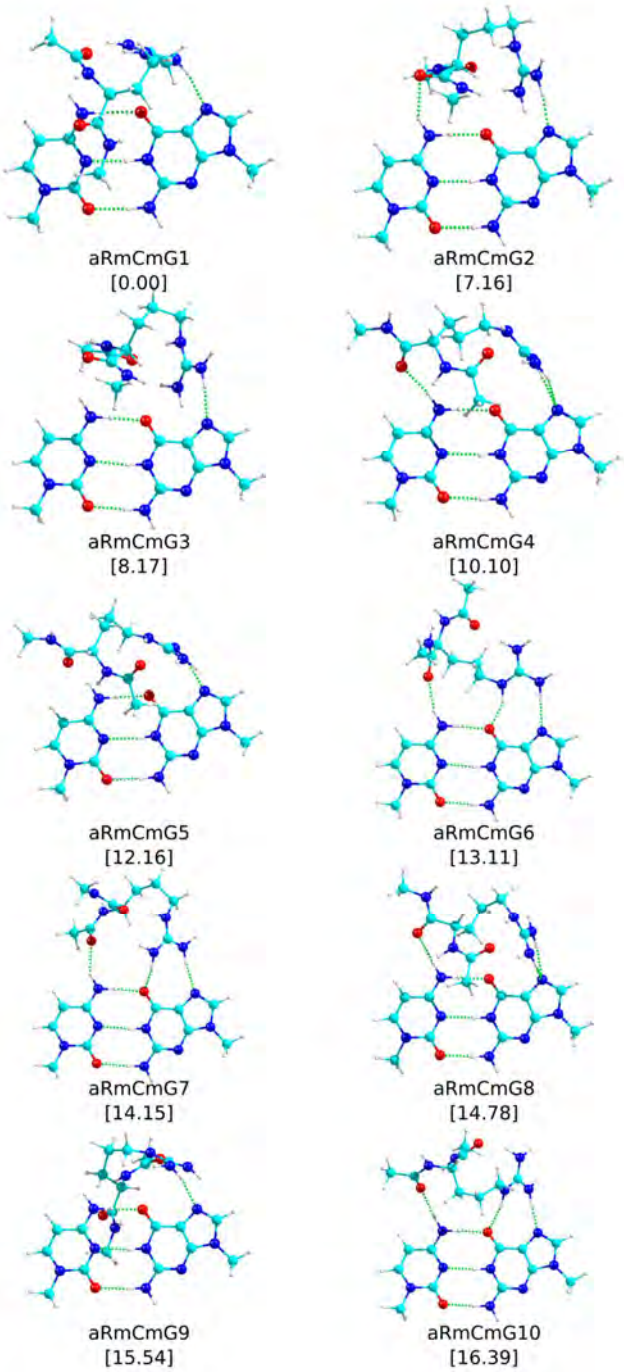


aQmCG19
[13.78]



aQmCG20
[14.85]

Figure 8.2.57 *The first 20 calculated structures for aRmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



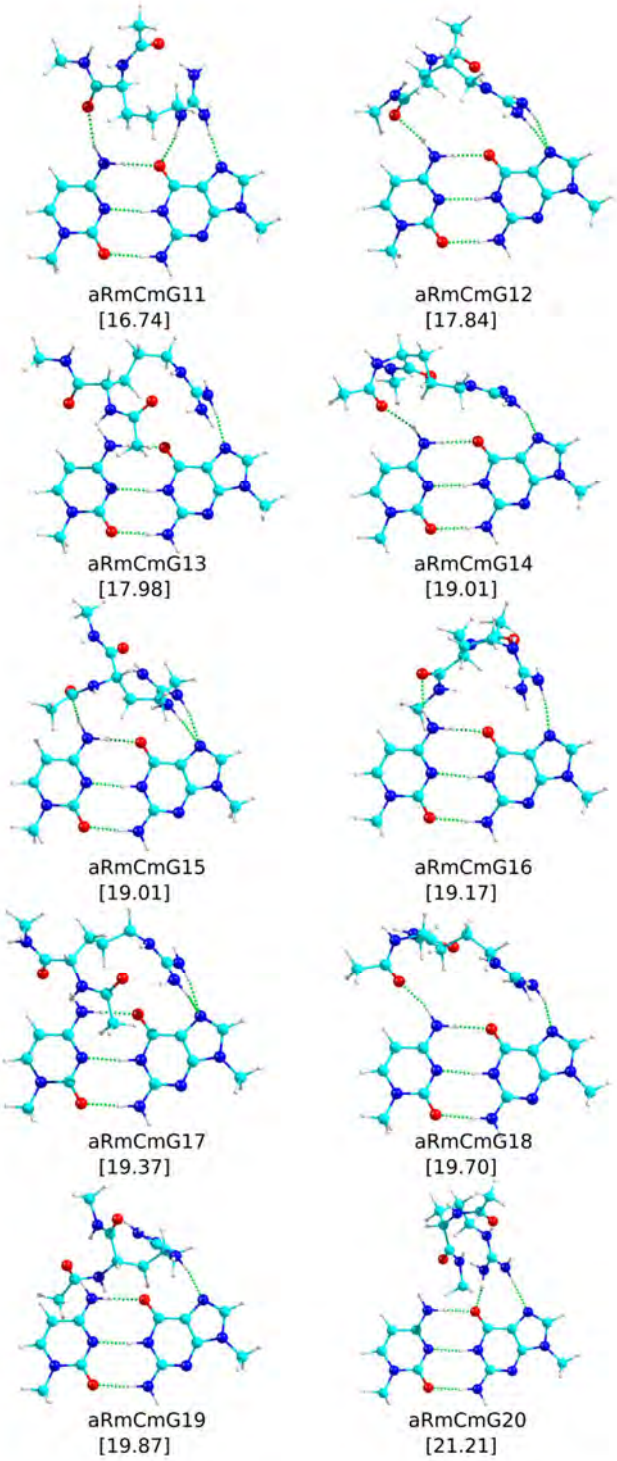
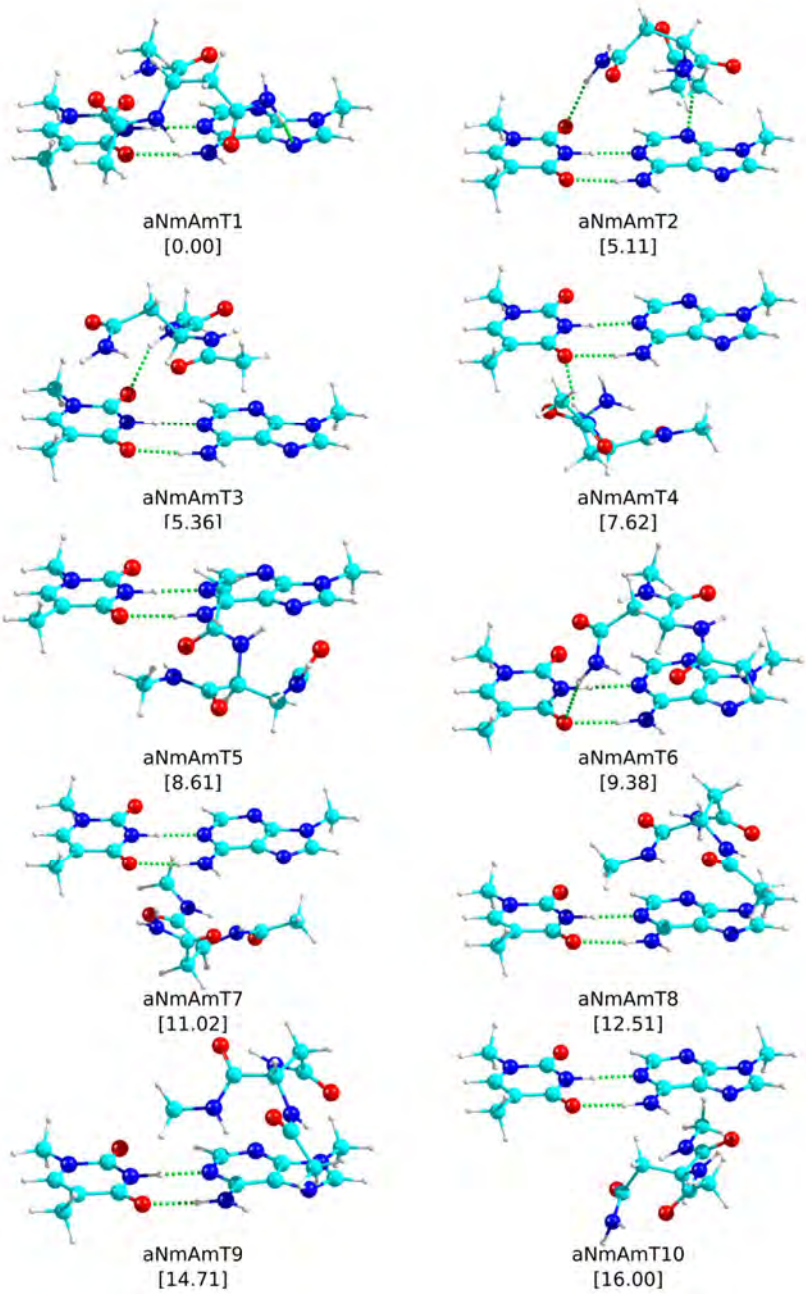
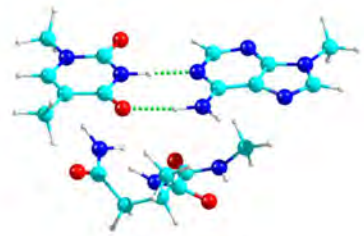
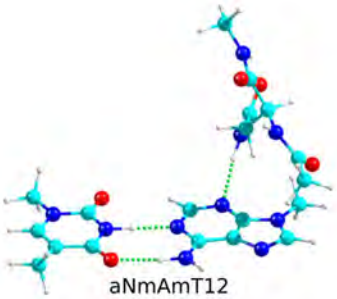

aRmCmG11
[16.74]
aRmCmG12
[17.84]
aRmCmG13
[17.98]
aRmCmG14
[19.01]
aRmCmG15
[19.01]
aRmCmG16
[19.17]
aRmCmG17
[19.37]
aRmCmG18
[19.70]
aRmCmG19
[19.87]
aRmCmG20
[21.21]

Figure 8.2.58 *The first 20 calculated structures for aNmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

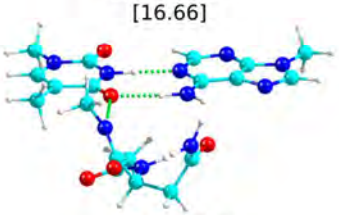




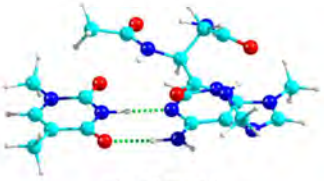
aNmAmT11
[16.66]



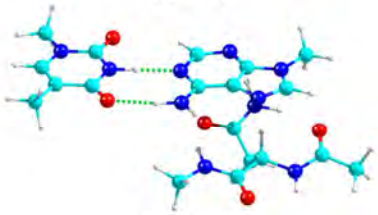
aNmAmT12
[16.69]



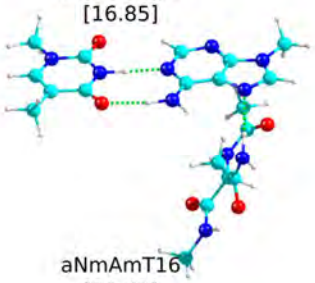
aNmAmT13
[16.77]



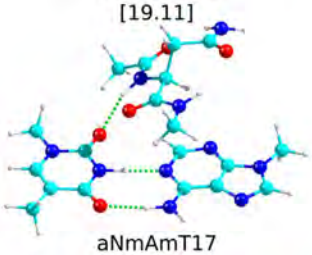
aNmAmT14
[16.85]



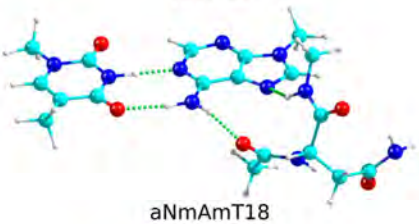
aNmAmT15
[19.11]



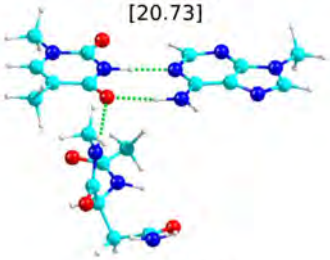
aNmAmT16
[20.41]



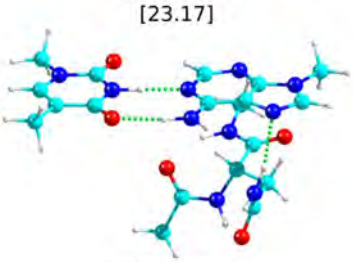
aNmAmT17
[20.73]



aNmAmT18
[23.17]

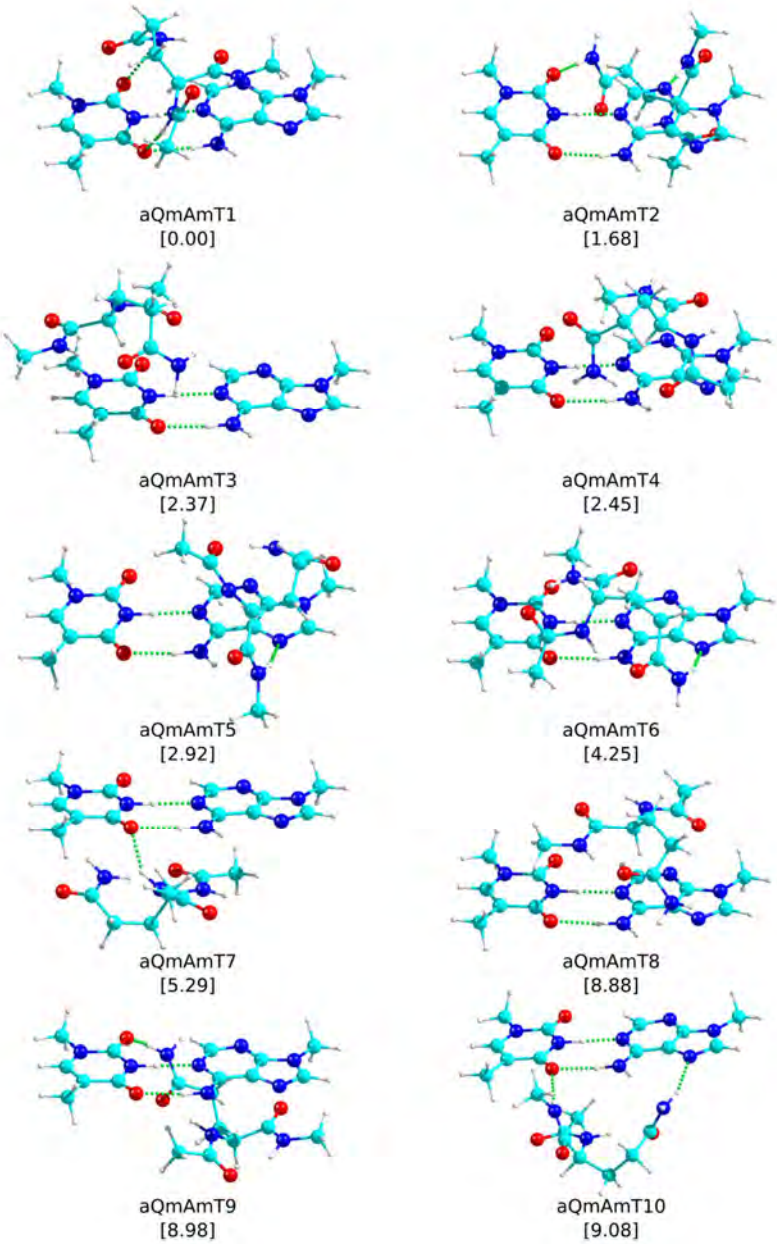


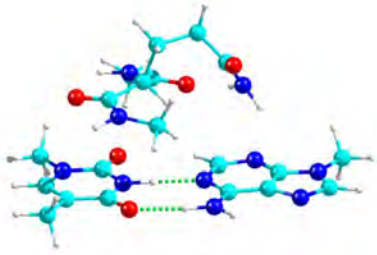
aNmAmT19
[23.29]



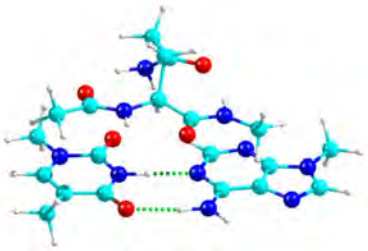
aNmAmT20
[24.76]

Figure 8.2.59 *The first 20 calculated structures for aQmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

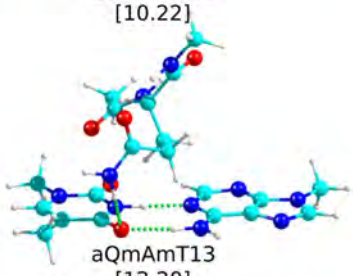




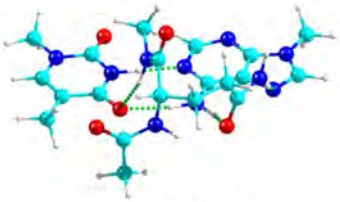
aQmAmT11
[10.22]



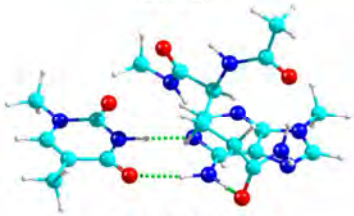
aQmAmT12
[11.76]



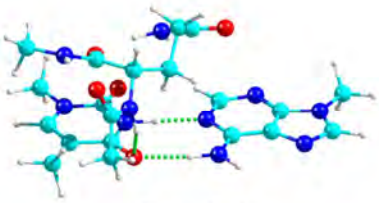
aQmAmT13
[12.20]



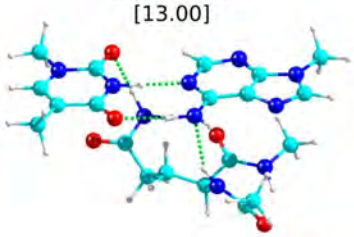
aQmAmT14
[12.59]



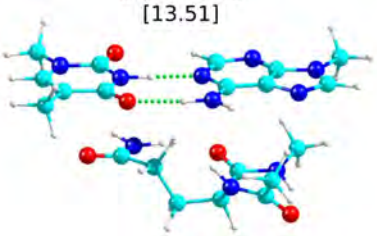
aQmAmT15
[13.00]



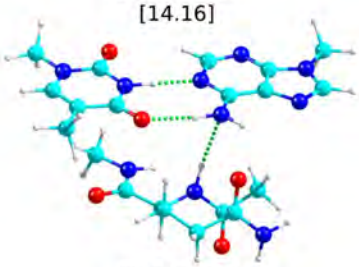
aQmAmT16
[13.51]



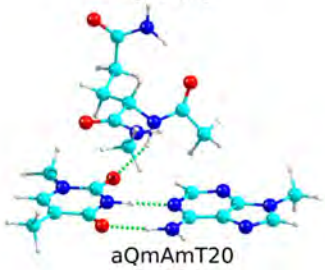
aQmAmT17
[14.16]



aQmAmT18
[14.53]

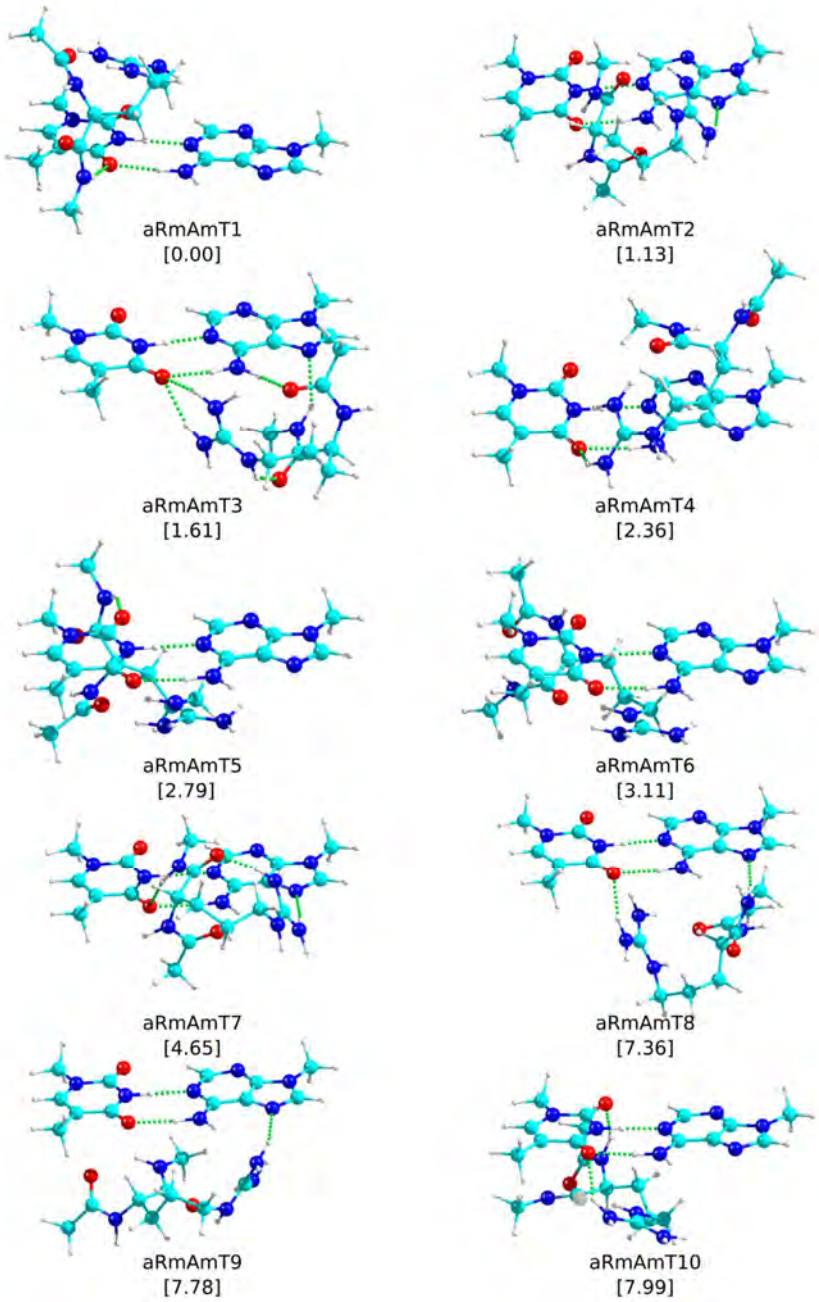


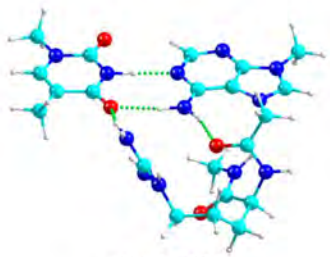
aQmAmT19
[16.10]



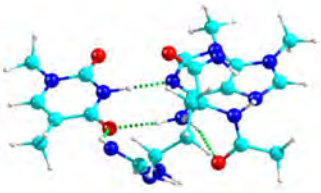
aQmAmT20
[16.19]

Figure 8.2.60 *The first 20 calculated structures for aRmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

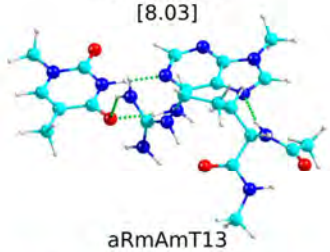




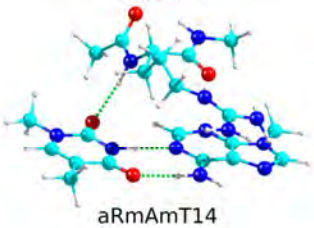
aRmAmT11
[8.03]



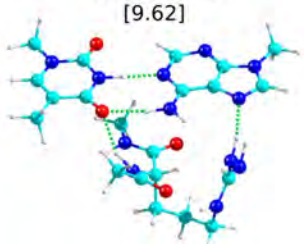
aRmAmT12
[8.96]



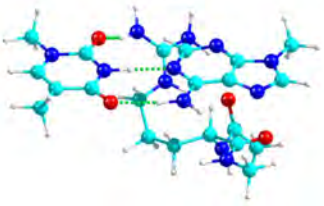
aRmAmT13
[9.62]



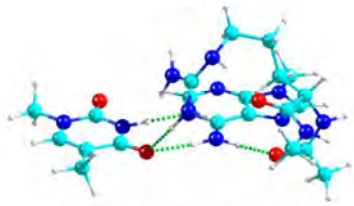
aRmAmT14
[9.89]



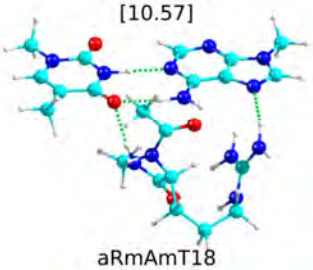
aRmAmT15
[9.91]



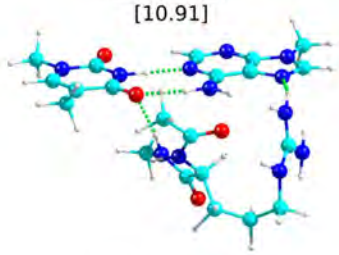
aRmAmT16
[10.57]



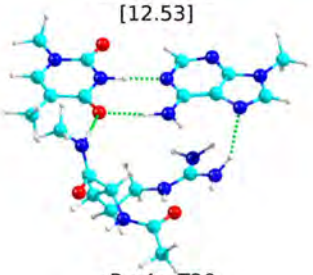
aRmAmT17
[10.91]



aRmAmT18
[12.53]

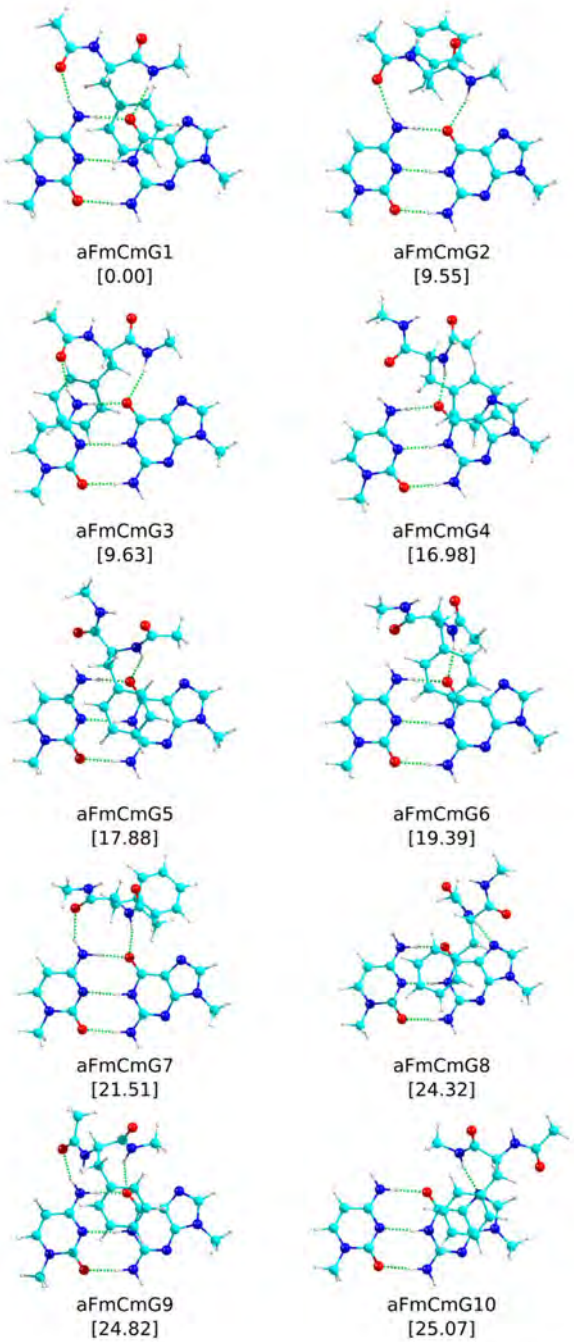


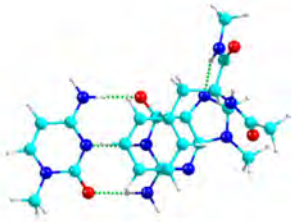
aRmAmT19
[12.59]



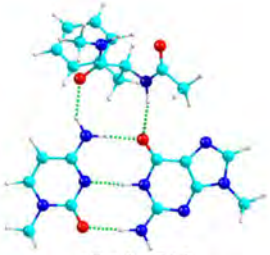
aRmAmT20
[12.96]

Figure 8.2.61 *The first 20 calculated structures for aFmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

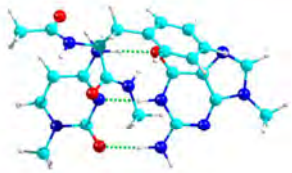




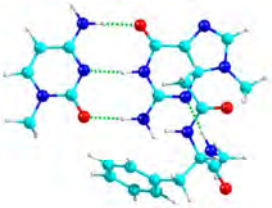
aFmCmG11
[25.63]



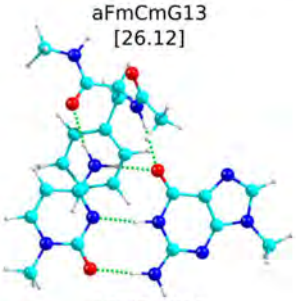
aFmCmG12
[25.90]



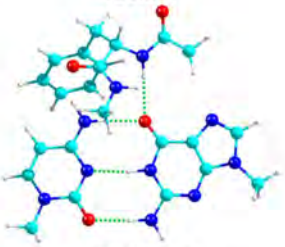
aFmCmG13
[26.12]



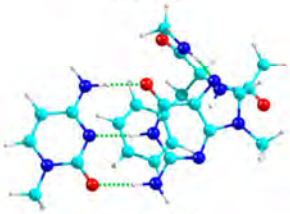
aFmCmG14
[26.27]



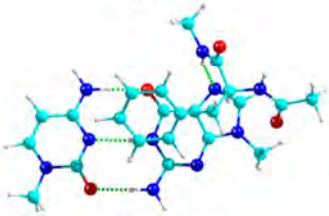
aFmCmG15
[27.39]



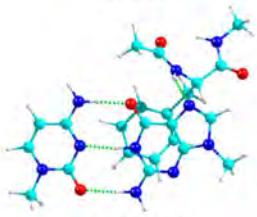
aFmCmG16
[27.84]



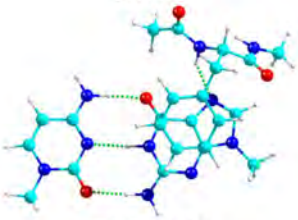
aFmCmG17
[27.85]



aFmCmG18
[27.93]

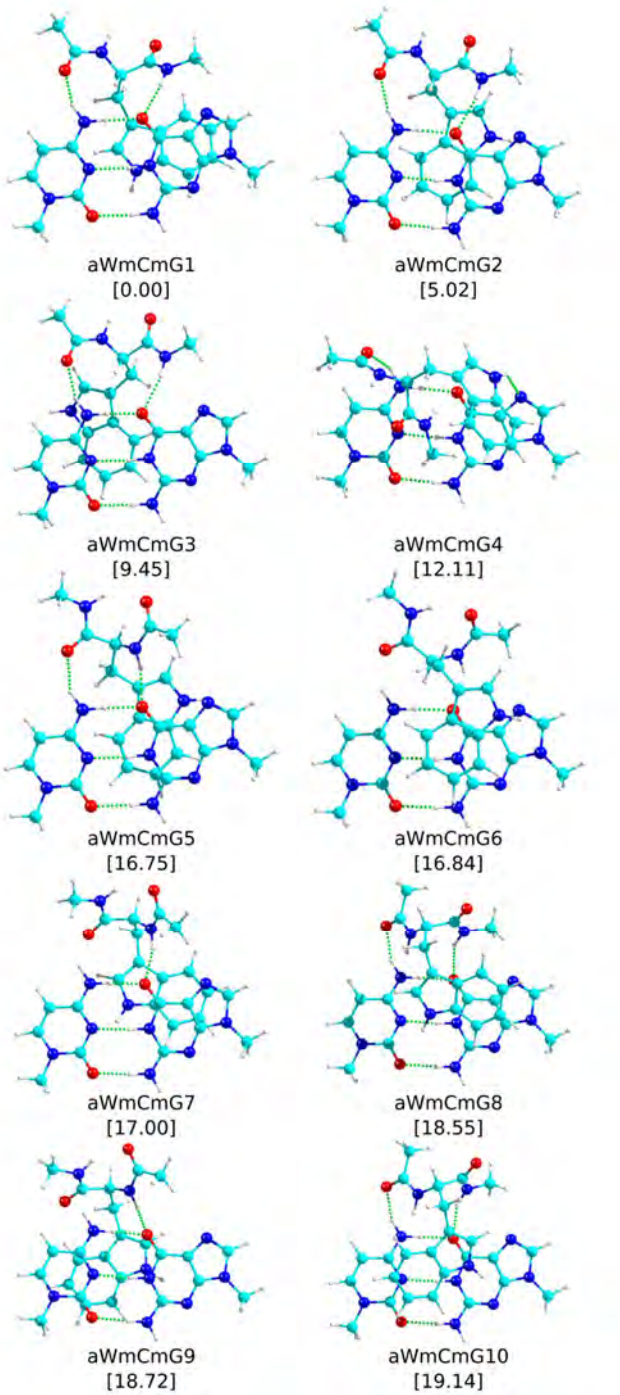


aFmCmG19
[28.49]



aFmCmG20
[28.71]

Figure 8.2.62 *The first 20 calculated structures for aWmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



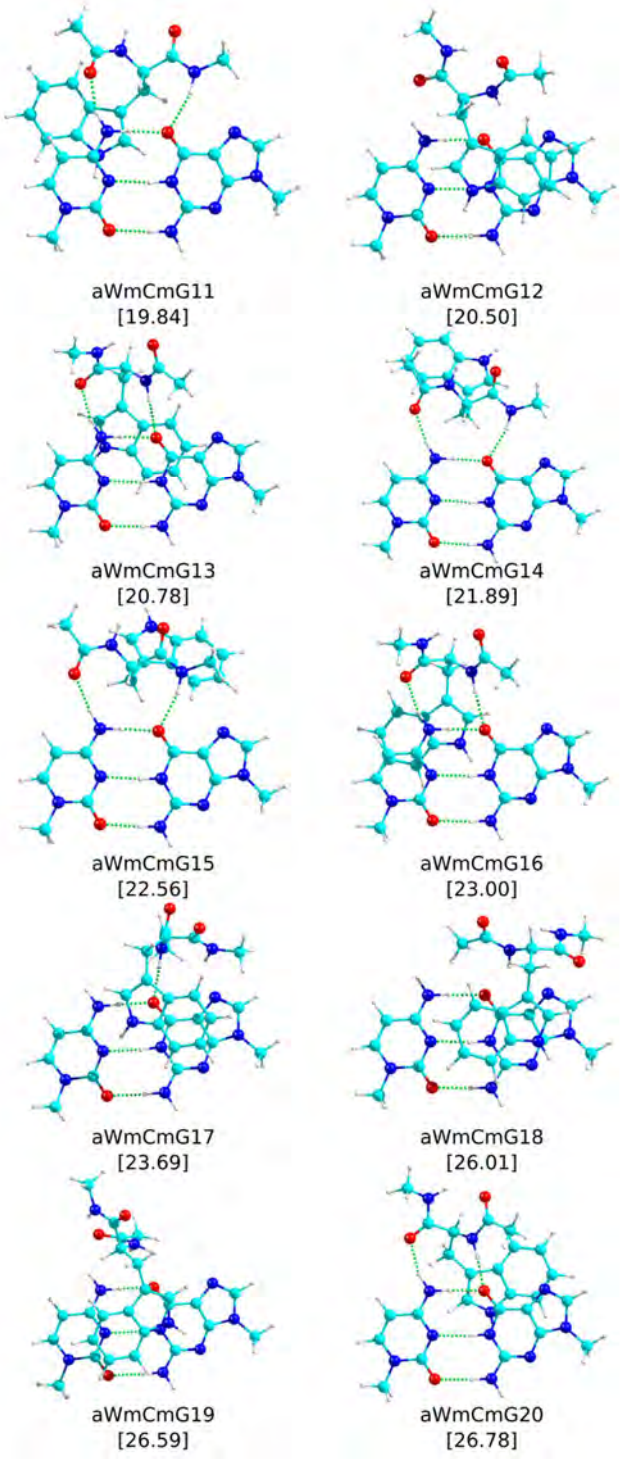
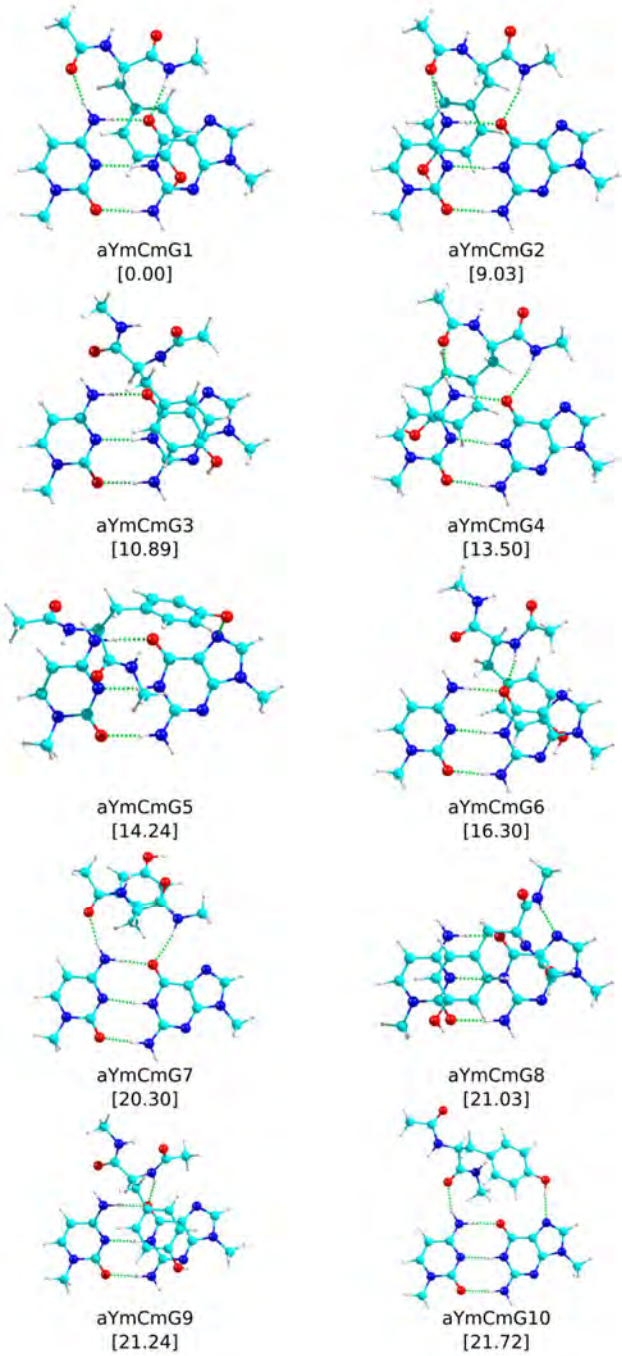
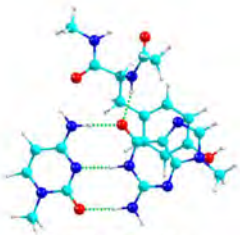
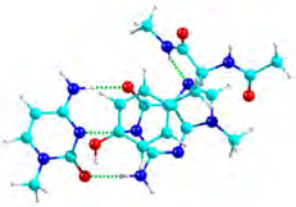

aWmCmG11
[19.84]
aWmCmG12
[20.50]
aWmCmG13
[20.78]
aWmCmG14
[21.89]
aWmCmG15
[22.56]
aWmCmG16
[23.00]
aWmCmG17
[23.69]
aWmCmG18
[26.01]
aWmCmG19
[26.59]
aWmCmG20
[26.78]

Figure 8.2.63 *The first 20 calculated structures for aYmCmG at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

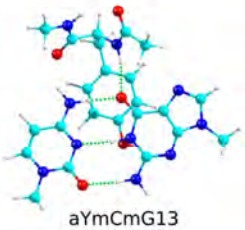




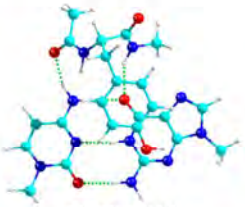
aYmCmG11
[22.20]



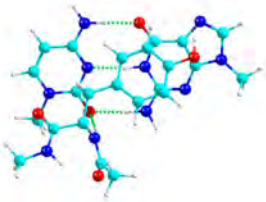
aYmCmG12
[22.65]



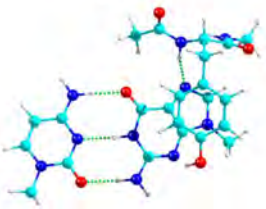
aYmCmG13
[23.33]



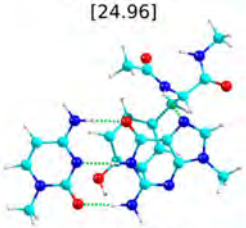
aYmCmG14
[24.80]



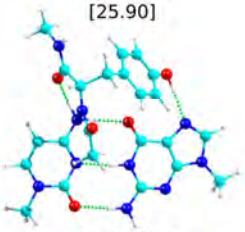
aYmCmG15
[24.96]



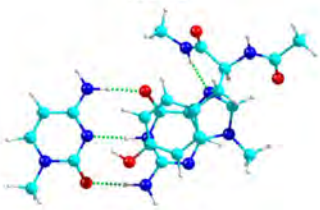
aYmCmG16
[25.90]



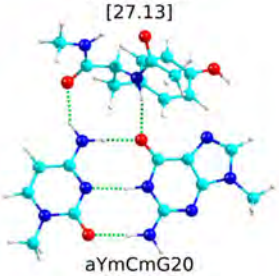
aYmCmG17
[25.91]



aYmCmG18
[27.13]

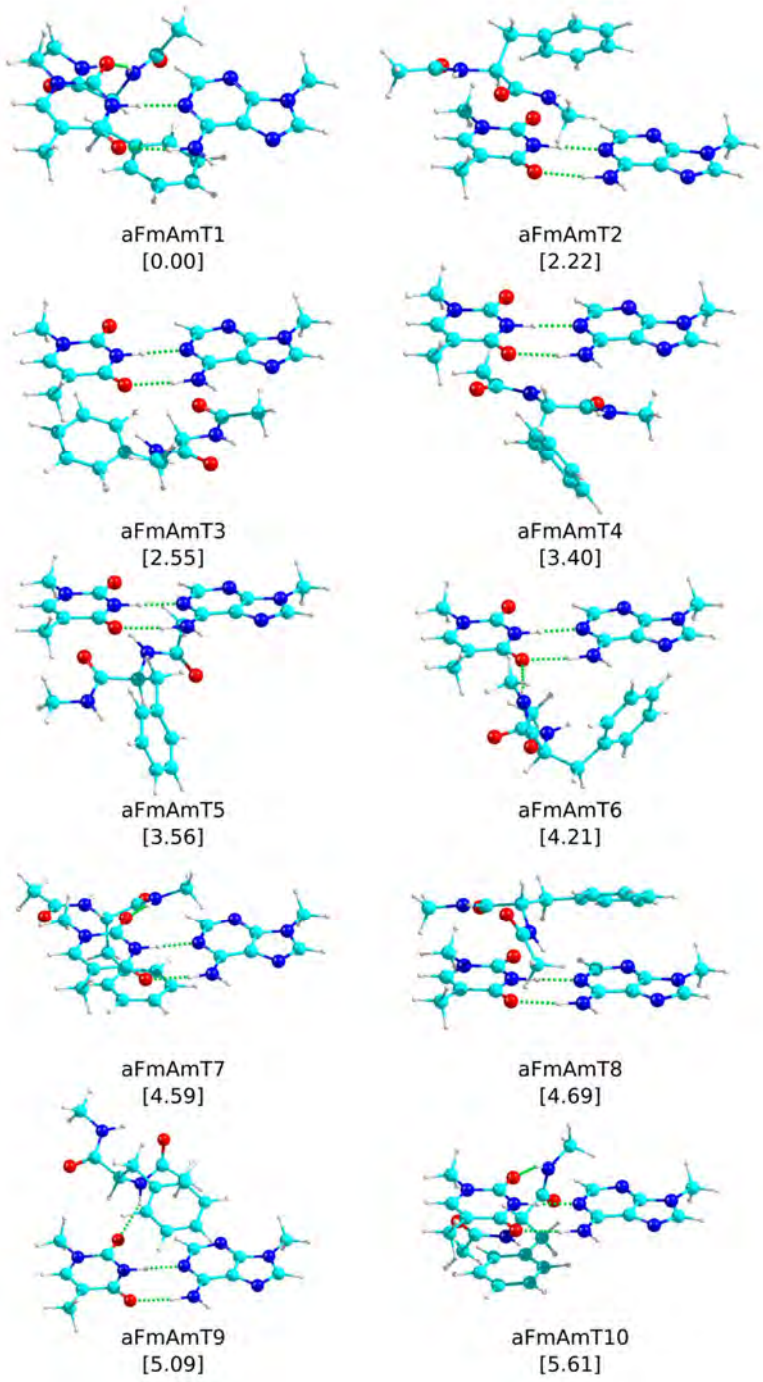


aYmCmG19
[27.65]



aYmCmG20
[28.71]

Figure 8.2.64 *The first 20 calculated structures for aFmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



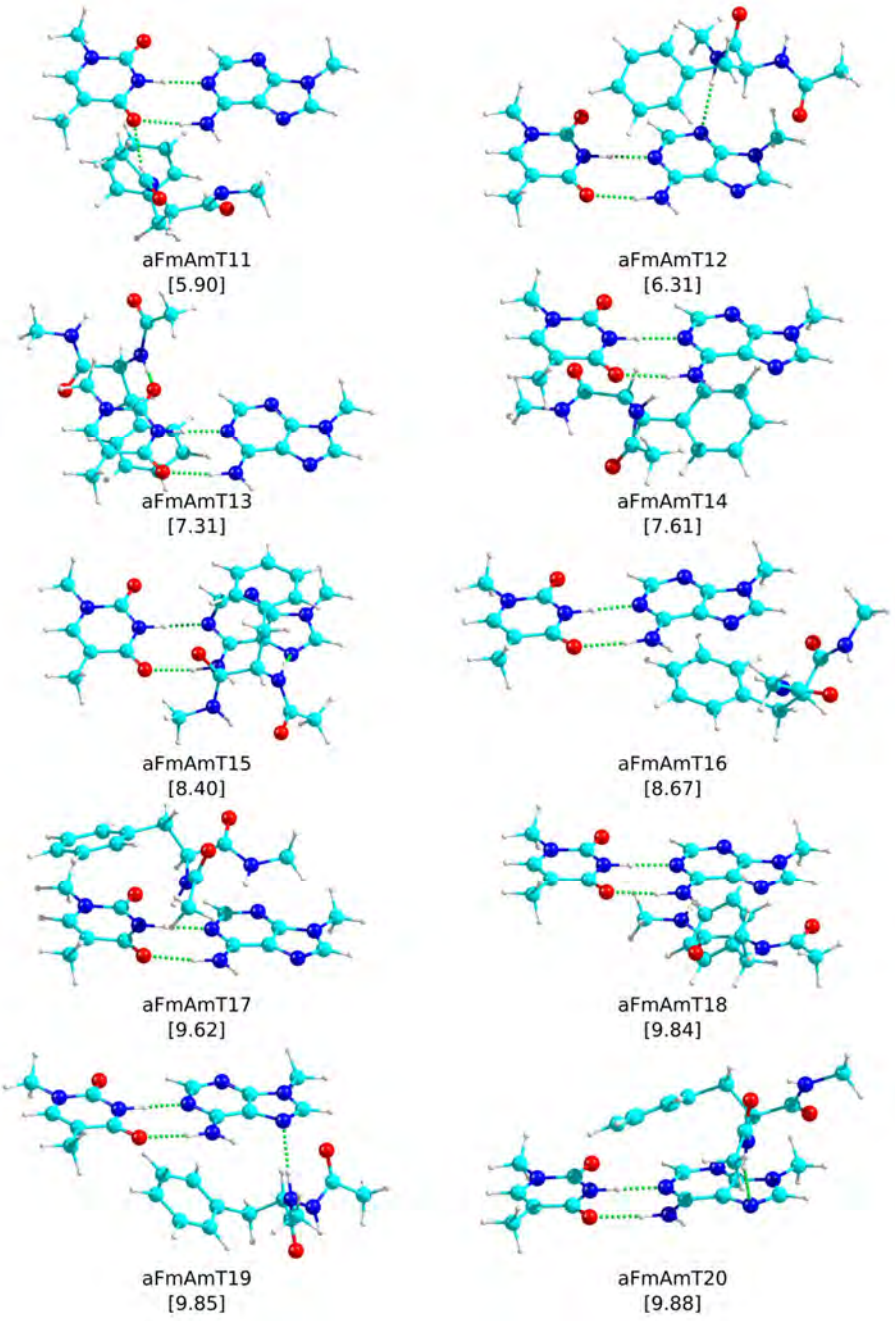
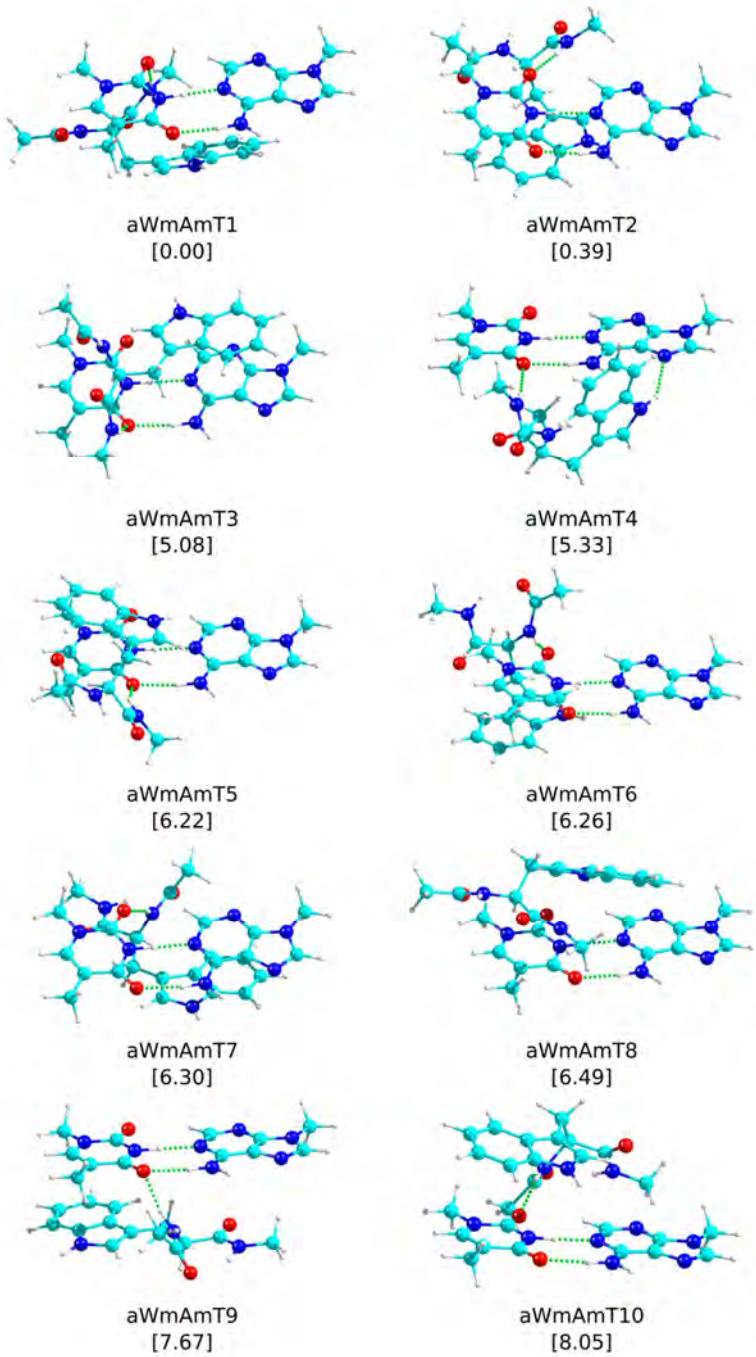
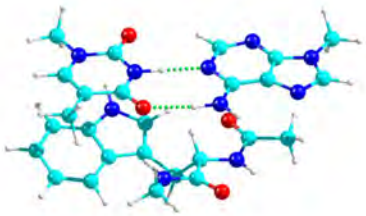
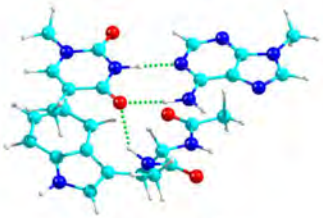

aFmAmT11
[5.90]
aFmAmT12
[6.31]
aFmAmT13
[7.31]
aFmAmT14
[7.61]
aFmAmT15
[8.40]
aFmAmT16
[8.67]
aFmAmT17
[9.62]
aFmAmT18
[9.84]
aFmAmT19
[9.85]
aFmAmT20
[9.88]

Figure 8.2.65 *The first 20 calculated structures for aWmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

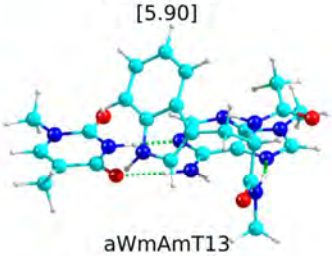




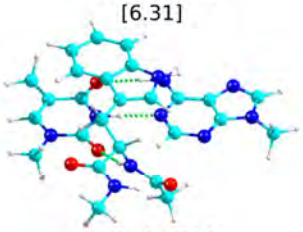
aWmAmT11
[5.90]



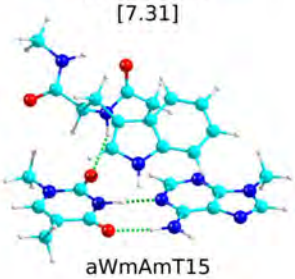
aWmAmT12
[6.31]



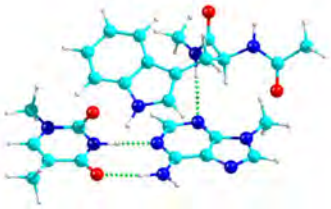
aWmAmT13
[7.31]



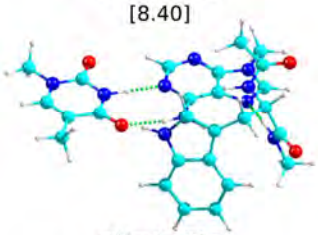
aWmAmT14
[7.61]



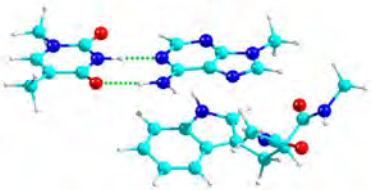
aWmAmT15
[8.40]



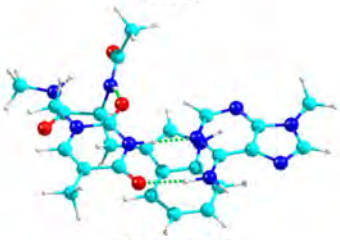
aWmAmT16
[8.67]



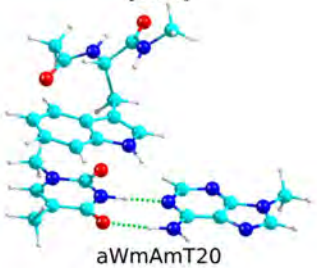
aWmAmT17
[9.62]



aWmAmT18
[9.84]

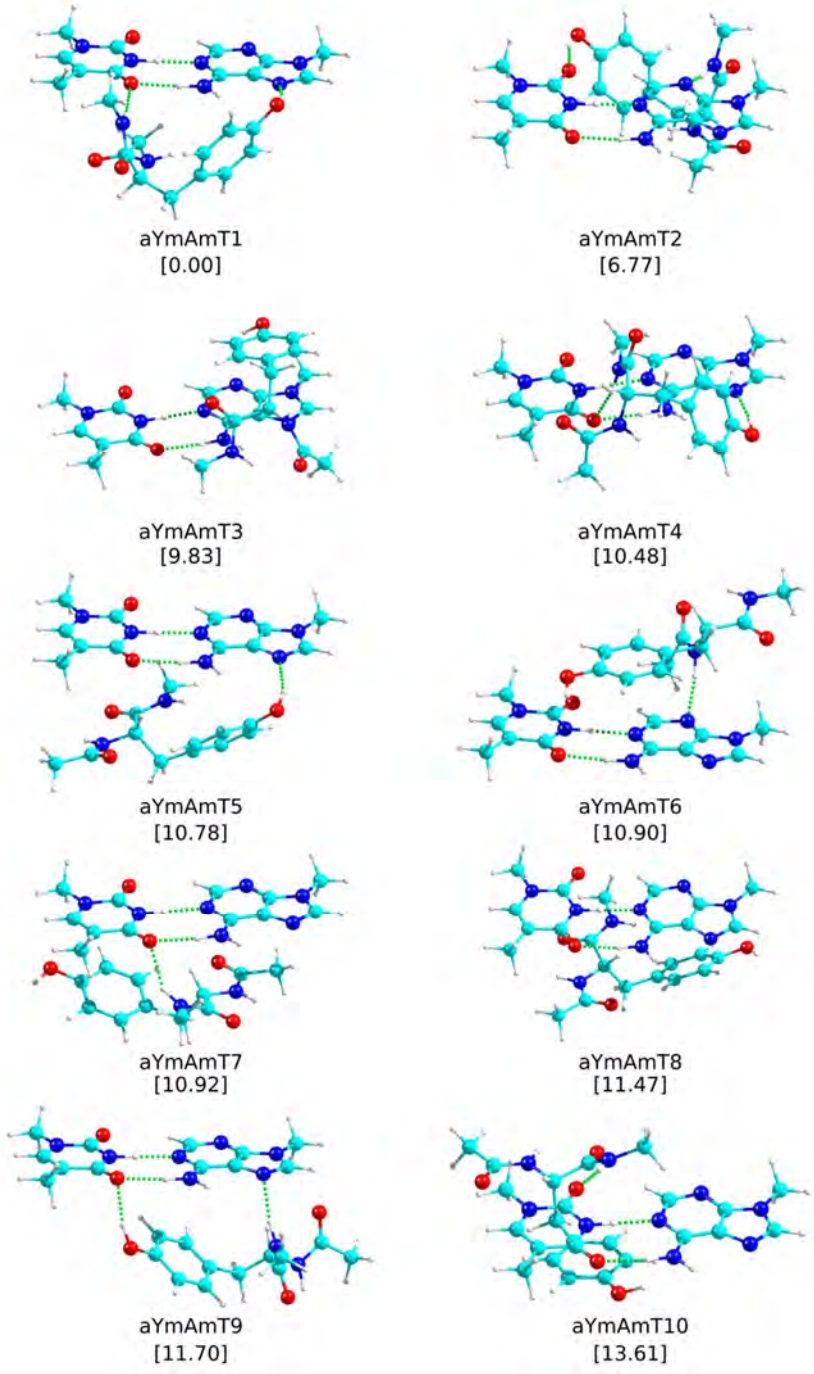


aWmAmT19
[9.85]



aWmAmT20
[9.88]

Figure 8.2.66 *The first 20 calculated structures for aYmAmT at M06-2X/6-311++G(d,p) level together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*



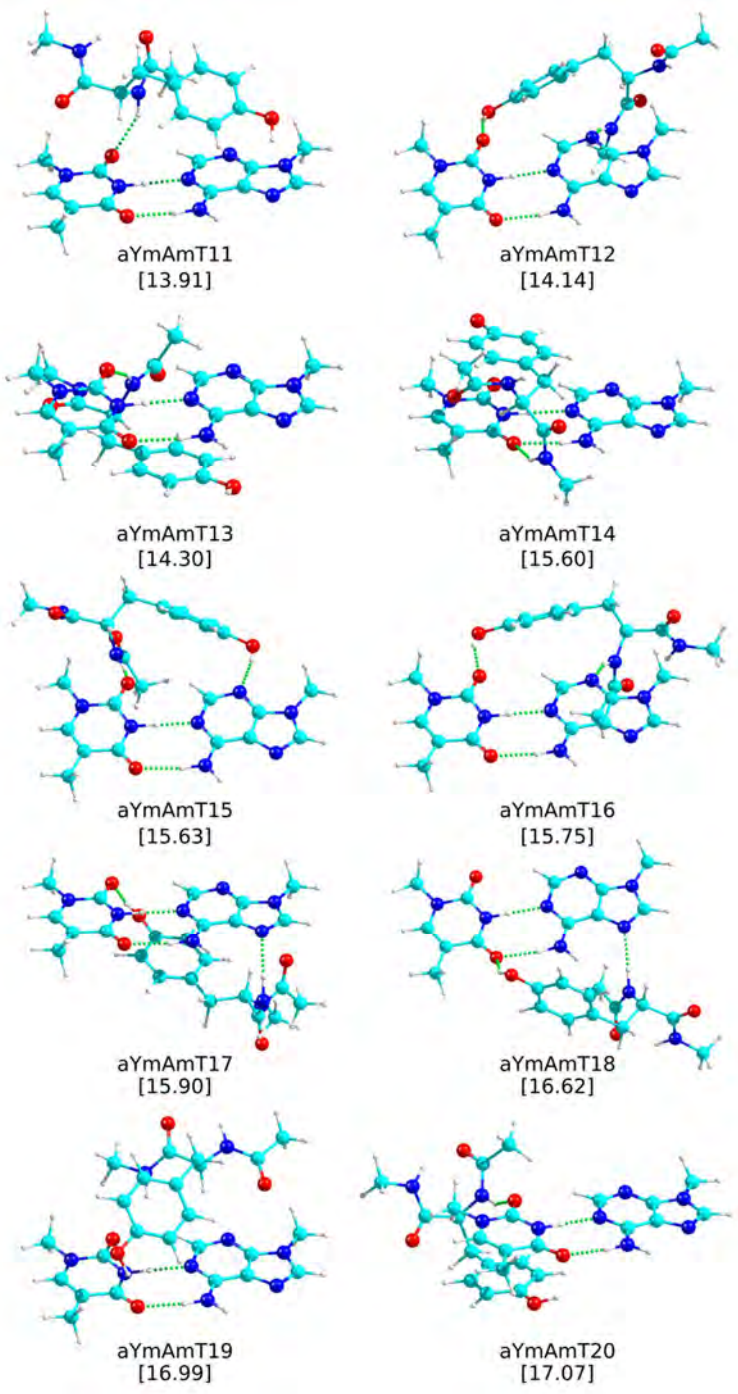

aYmAmT11
[13.91]
aYmAmT12
[14.14]
aYmAmT13
[14.30]
aYmAmT14
[15.60]
aYmAmT15
[15.63]
aYmAmT16
[15.75]
aYmAmT17
[15.90]
aYmAmT18
[16.62]
aYmAmT19
[16.99]
aYmAmT20
[17.07]

Figure 8.2.67 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aAmCmG, aImCmG and aVmCmG.*

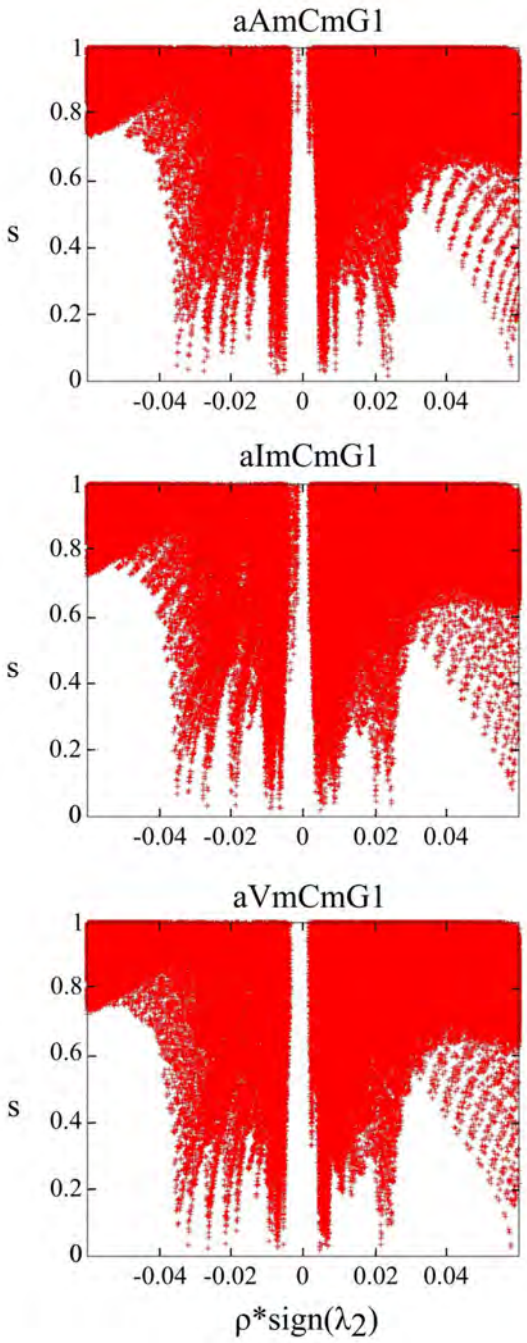


Figure 8.2.68 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aAmAmT, aImAmT and aVmAmT.*

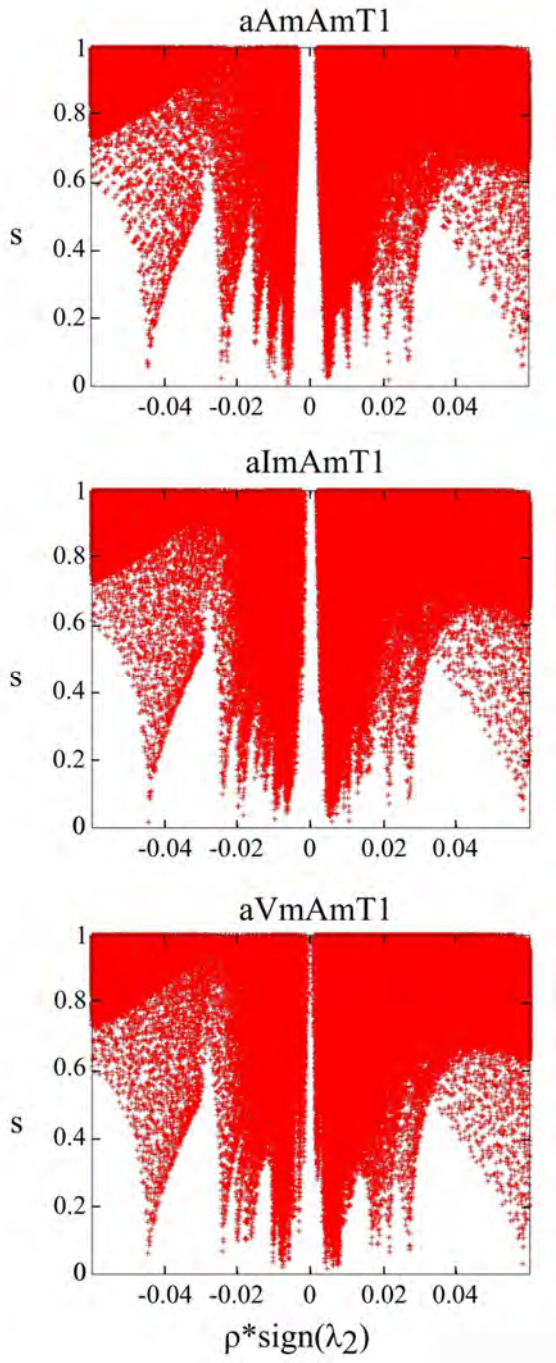


Figure 8.2.69 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aNmCmG, aQmCmG and aRmCmG.*

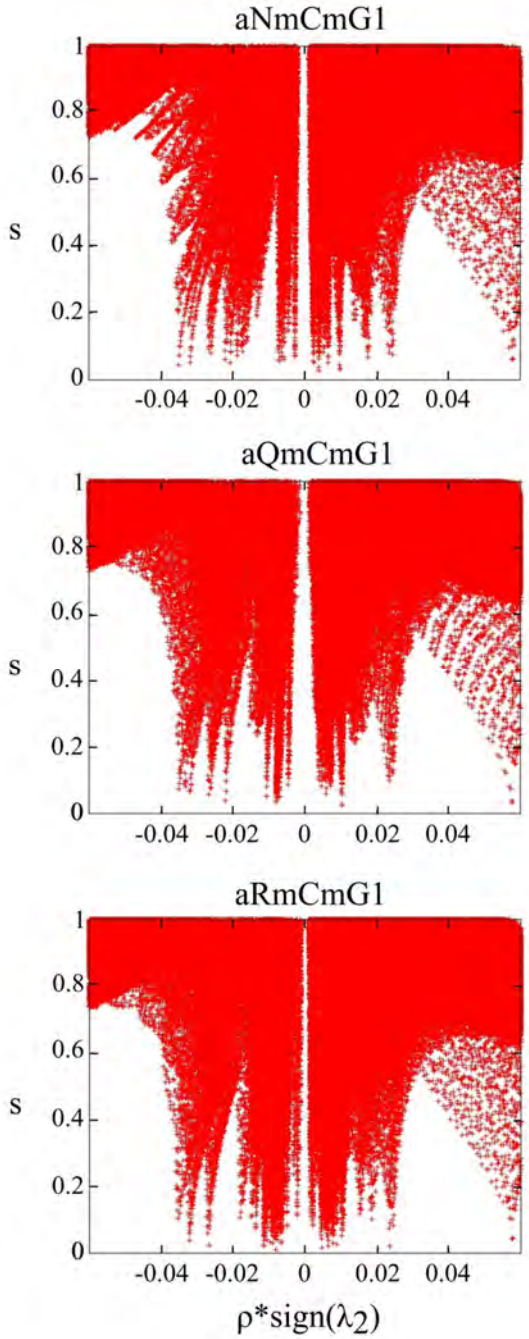


Figure 8.2.70 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aNmAmT, aQmAmT and aRmAmT.*

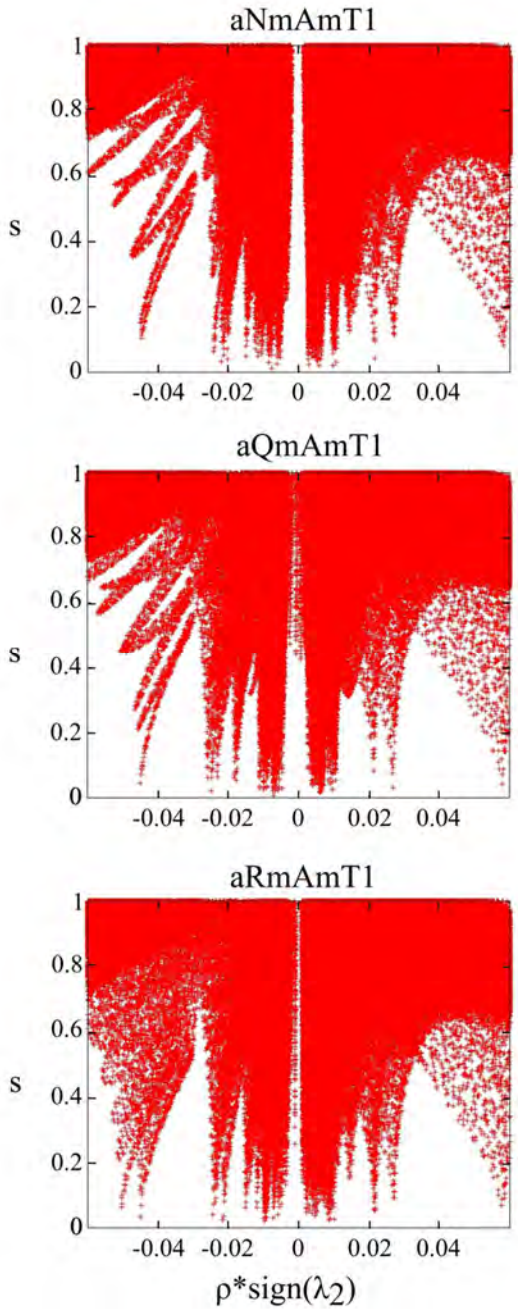


Figure 8.2.71 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aFmCmG, aWmCmG and aYmCmG.*

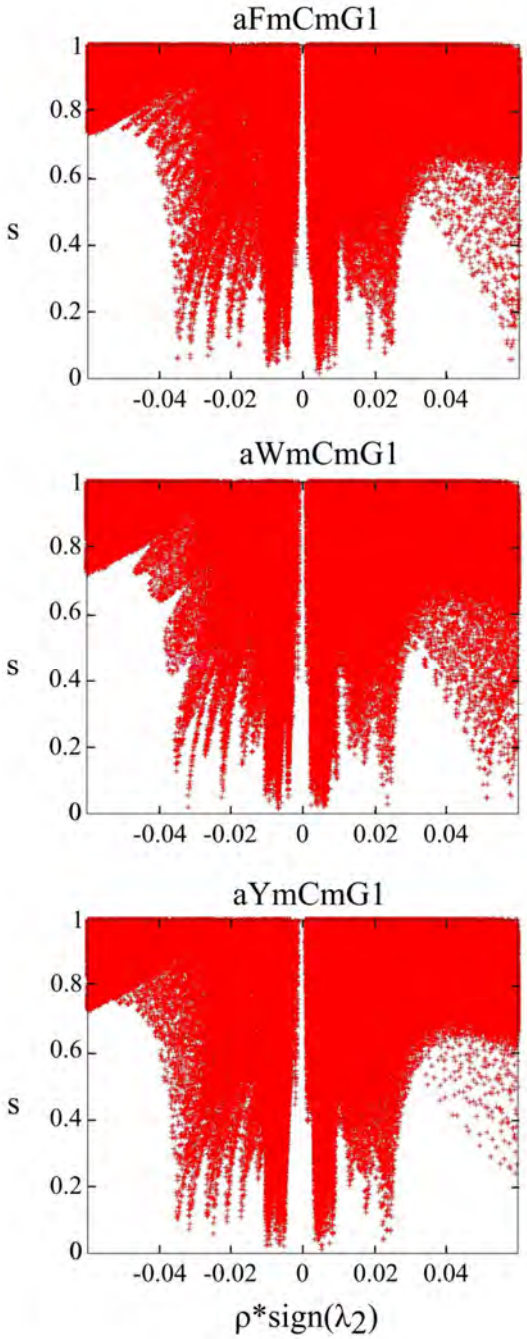
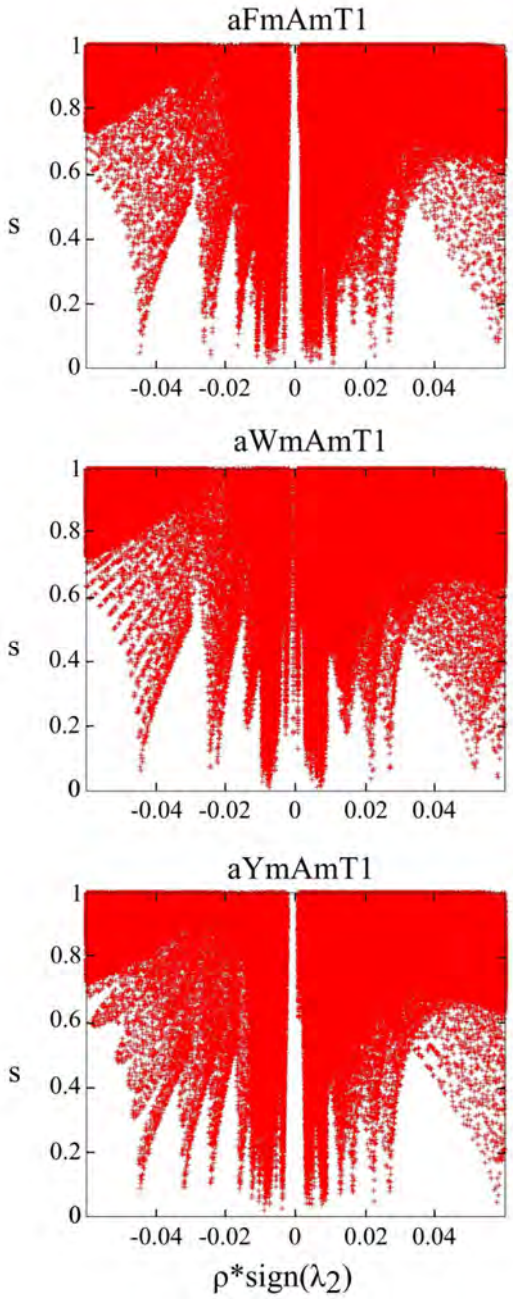


Figure 8.2.72 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the most stable conformations of aFmAmT, aWmAmT and aYmAmT.*



8.3 Aggregation of DNA bases

Figure 8.3.1 *The 10 most stable structures of kk cytosine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

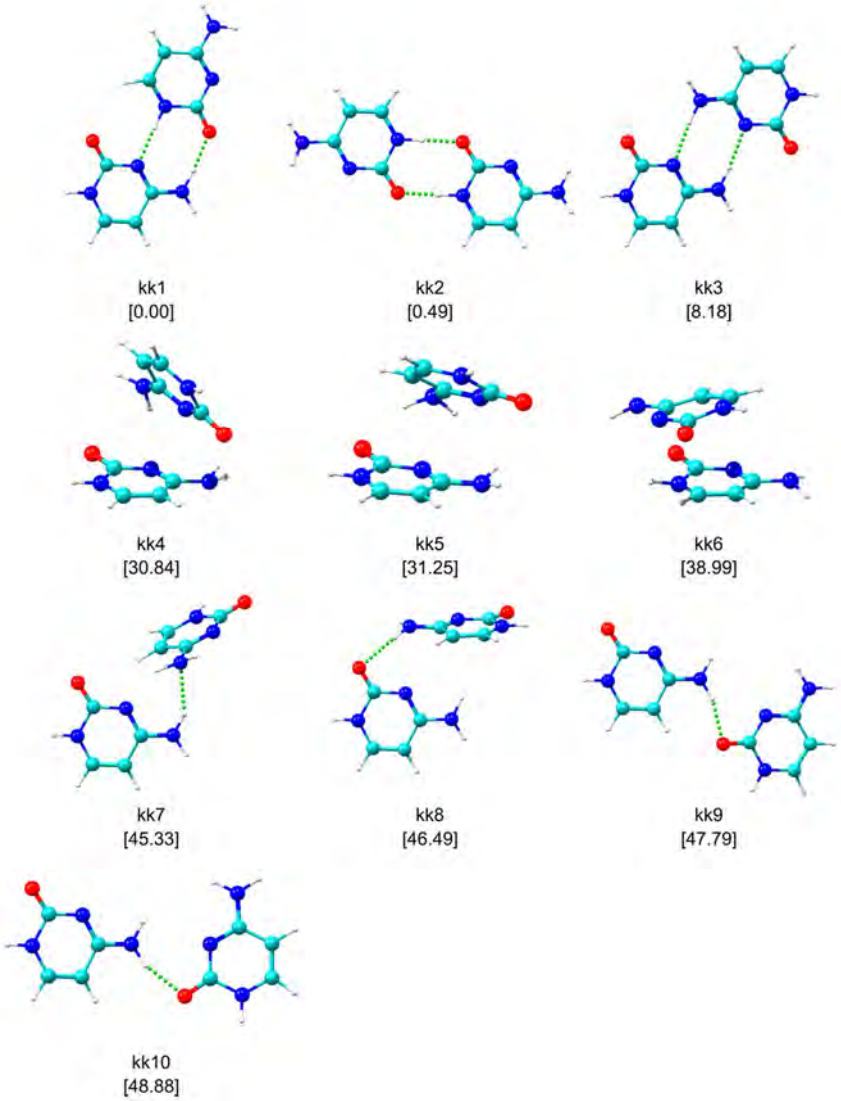


Figure 8.3.2 *The 15 most stable structures of ke cytosine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

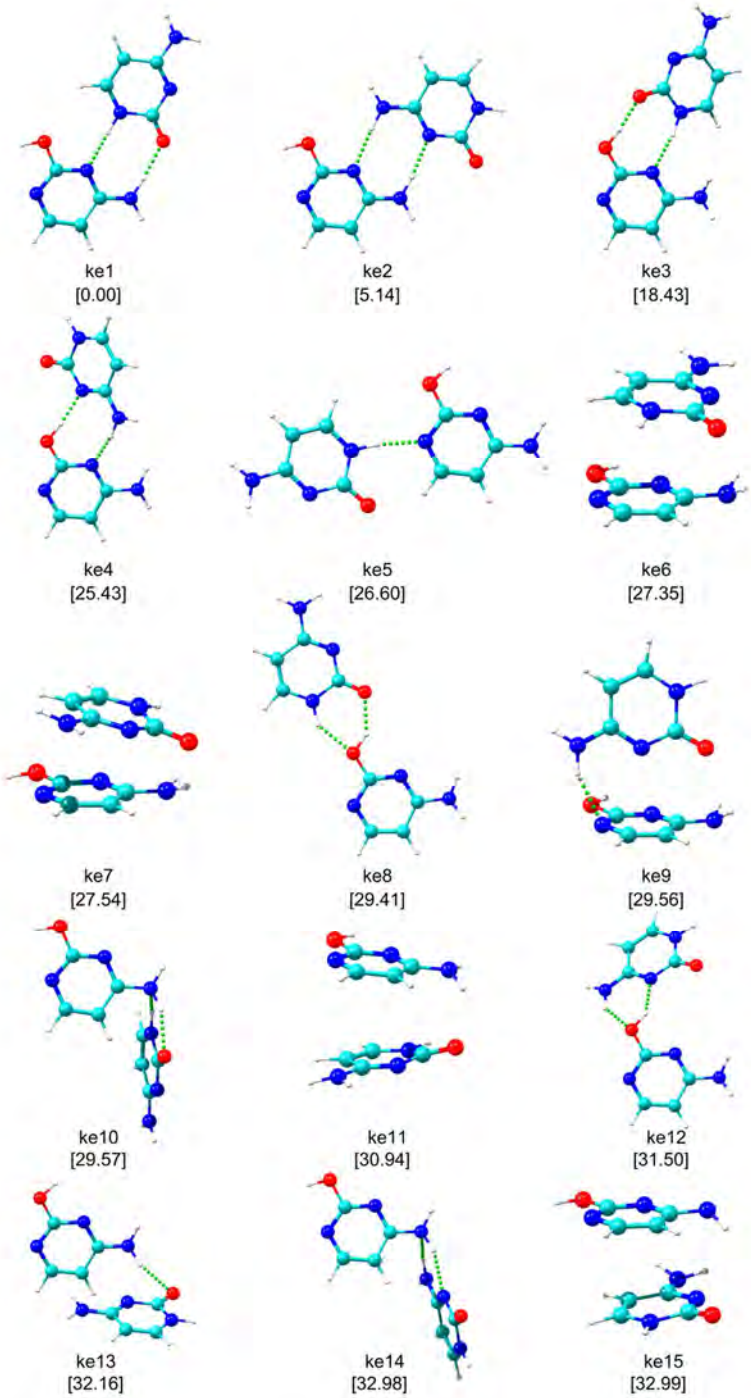


Figure 8.3.3 *The 15 most stable structures of ee cytosine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

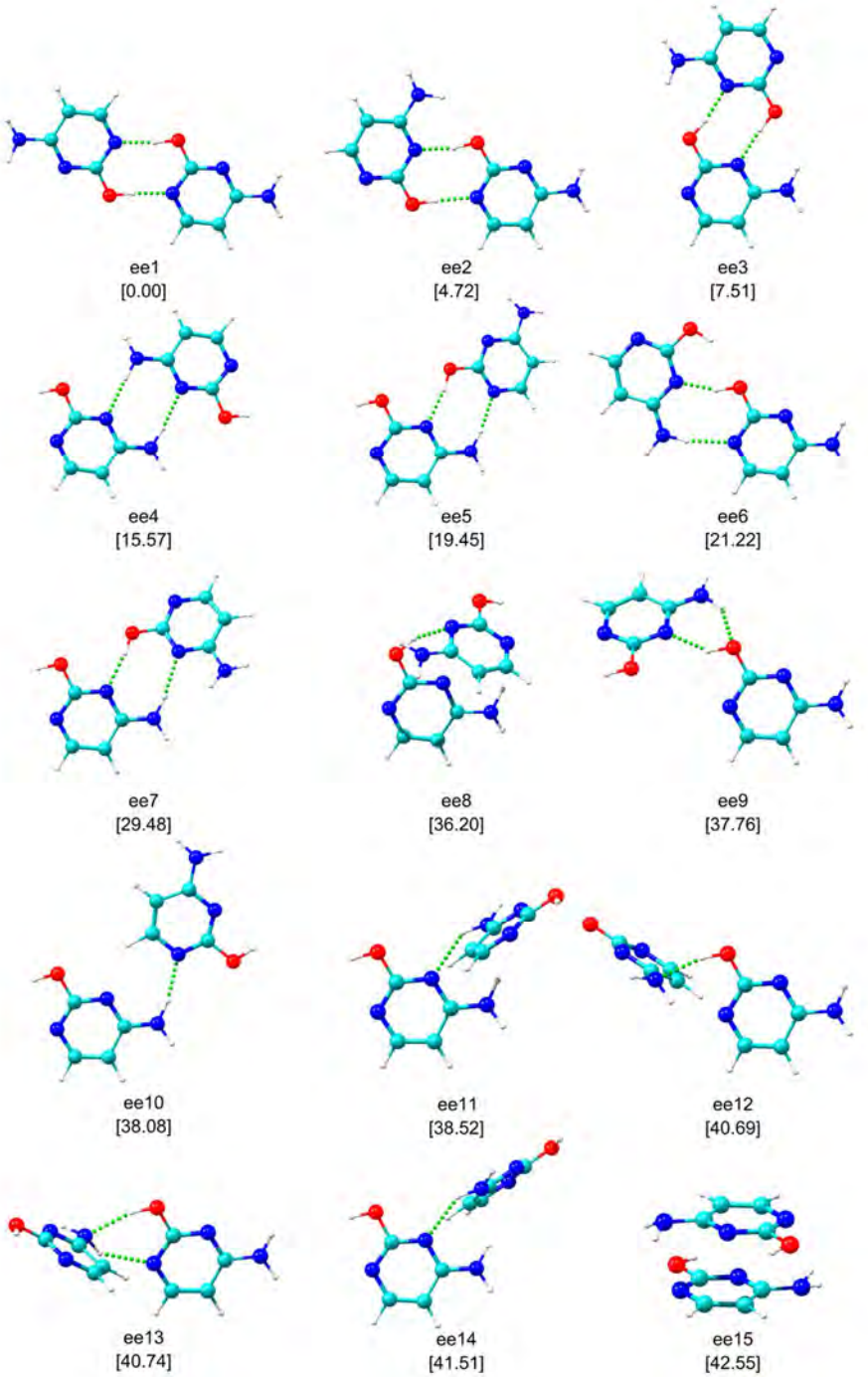


Figure 8.3.4 *Relative Gibbs free energy of the kk dimers of cytosine in the 0-700 K interval. Continous line: planar structure.*

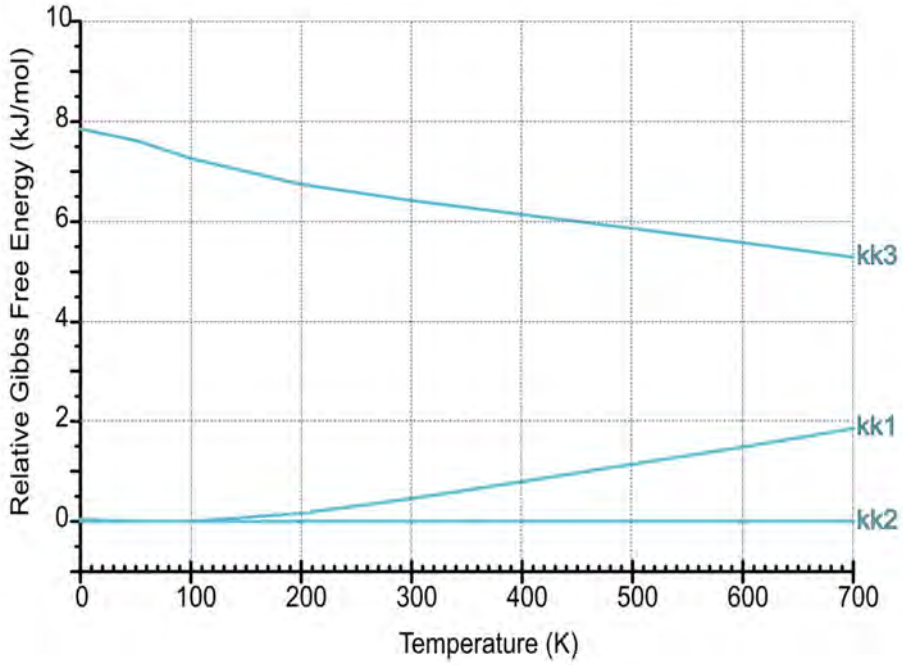


Figure 8.3.5 *Relative Gibbs free energy of the ke dimers of cytosine in the 0-700 K interval. Continous line: planar structure.*

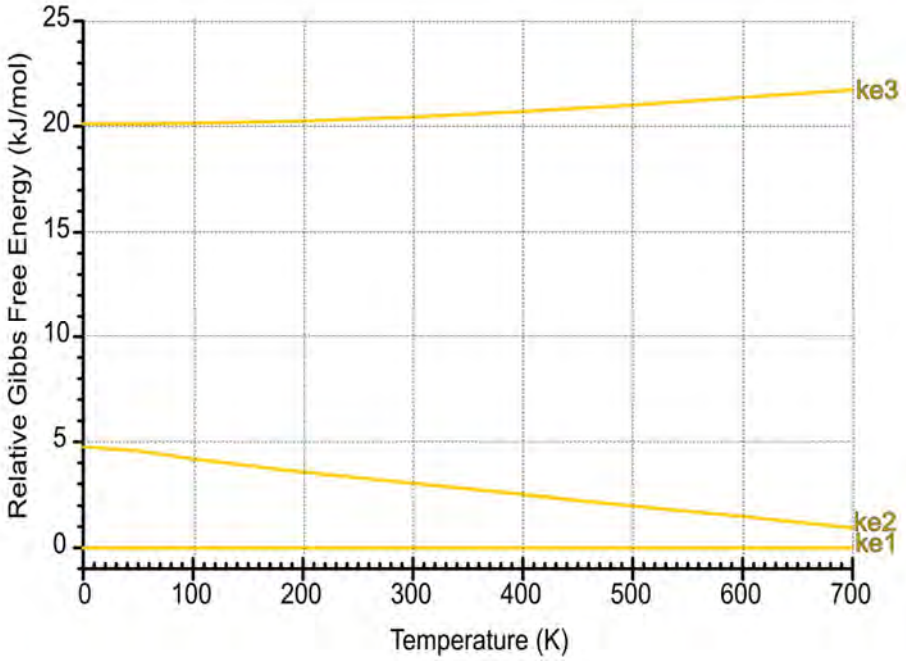


Figure 8.3.6 *Relative Gibbs free energy of the ee dimers of cytosine in the 0-700 K interval. Continous line: planar structure.*

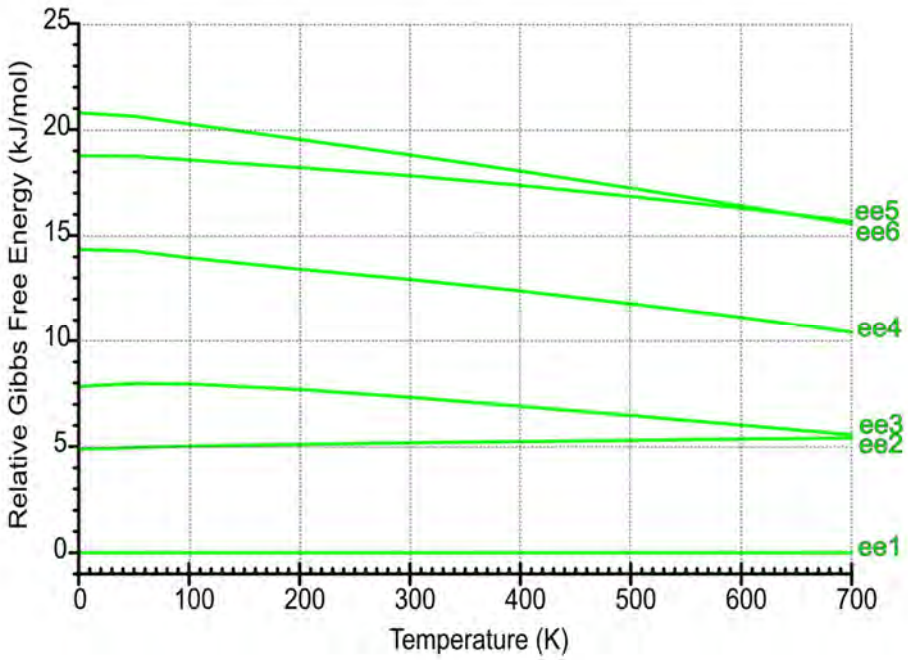


Figure 8.3.7 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered kk dimers of cytosine.*

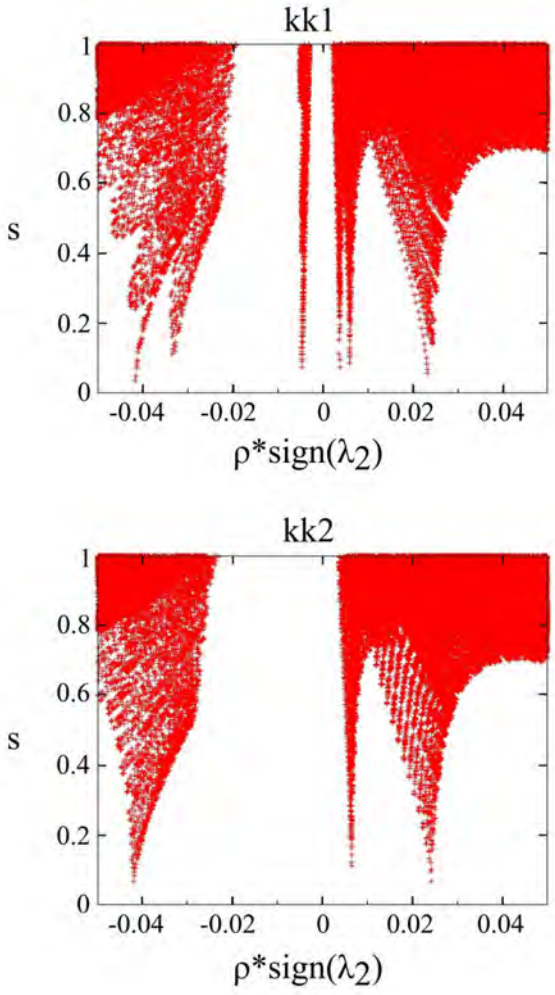


Figure 8.3.8 *The 15 most stable structures of kkk cytosine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

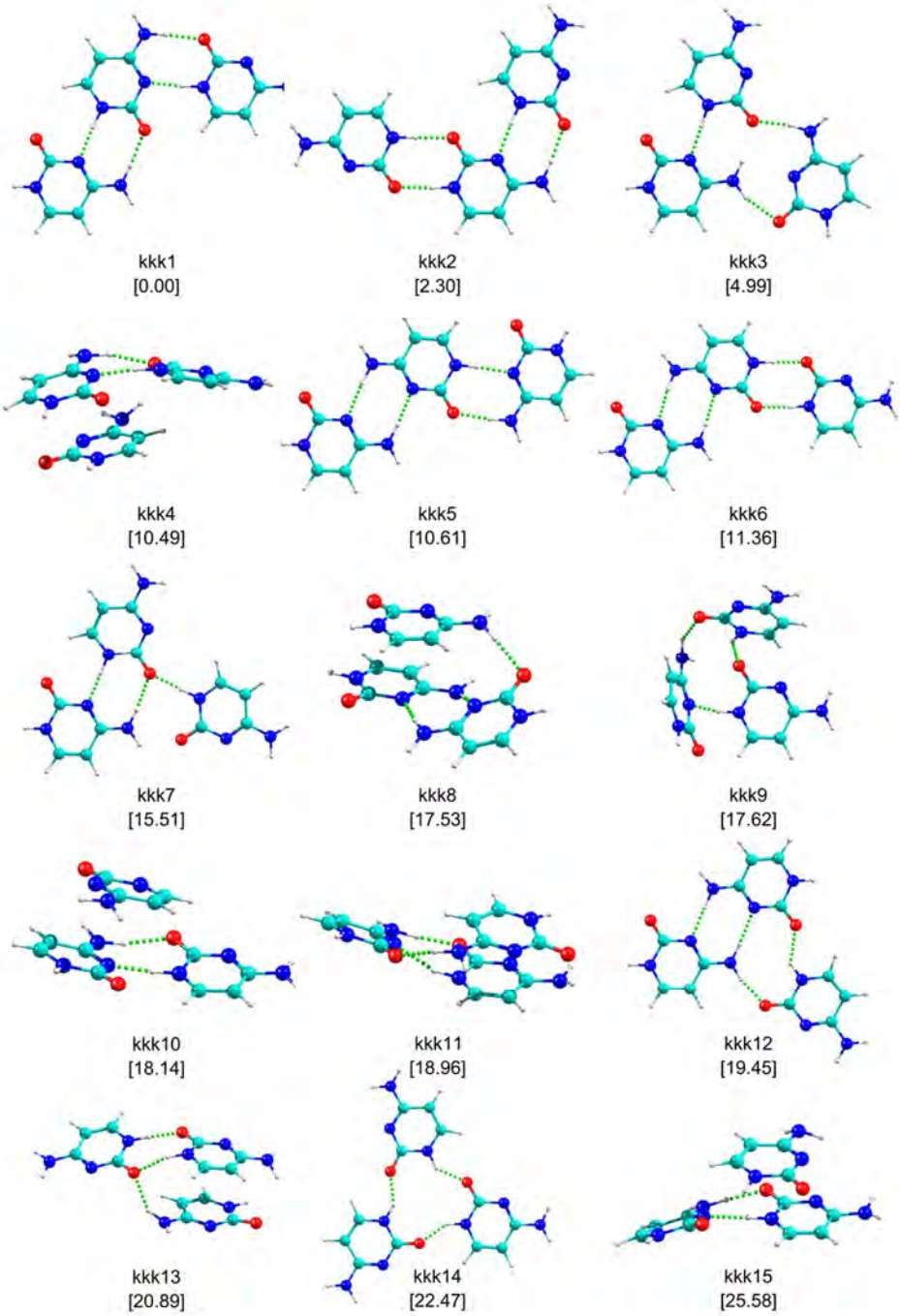


Figure 8.3.9 *The 15 most stable structures of kke cytosine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

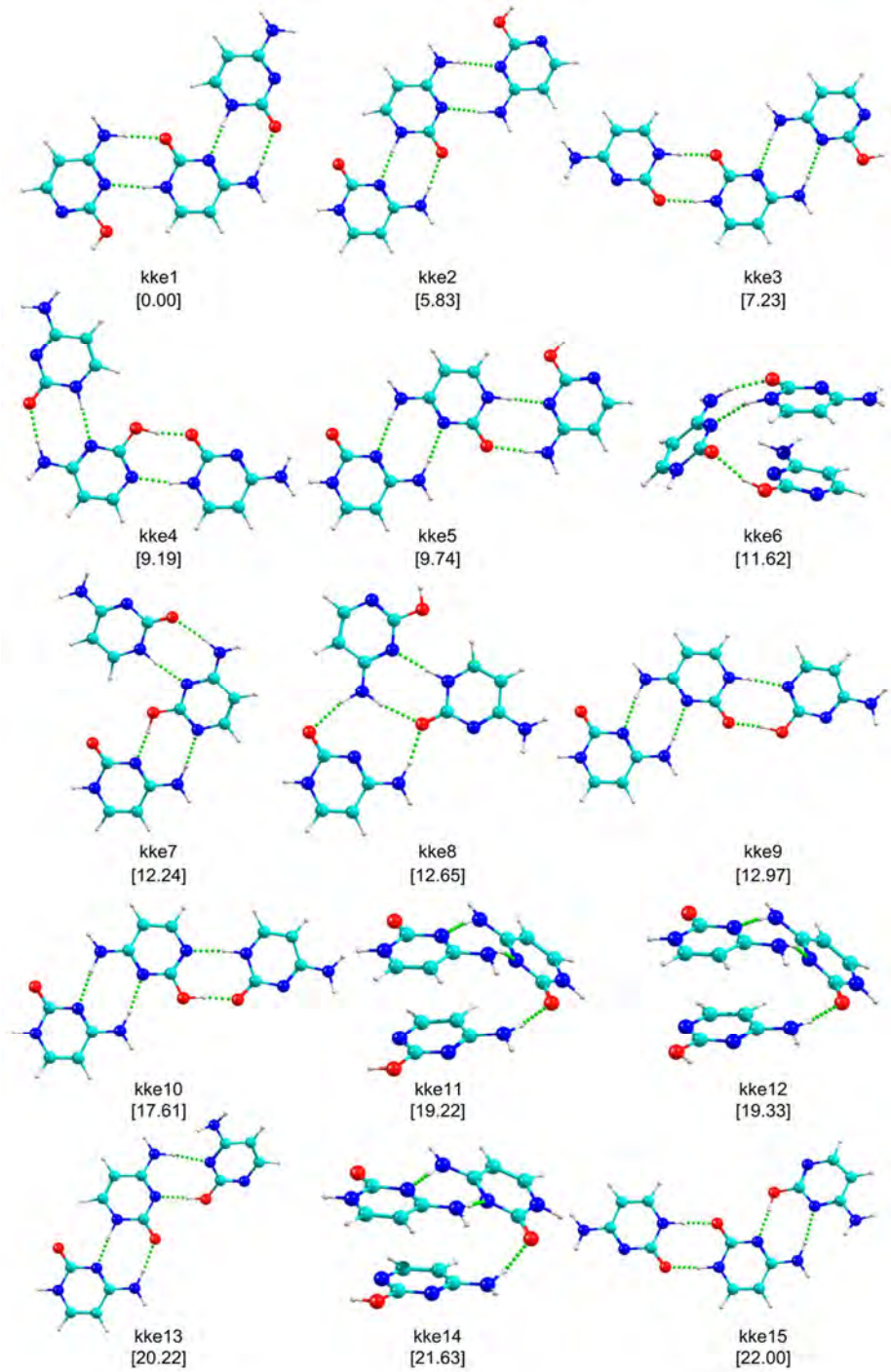


Figure 8.3.10 *The 15 most stable structures of kee cytosine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

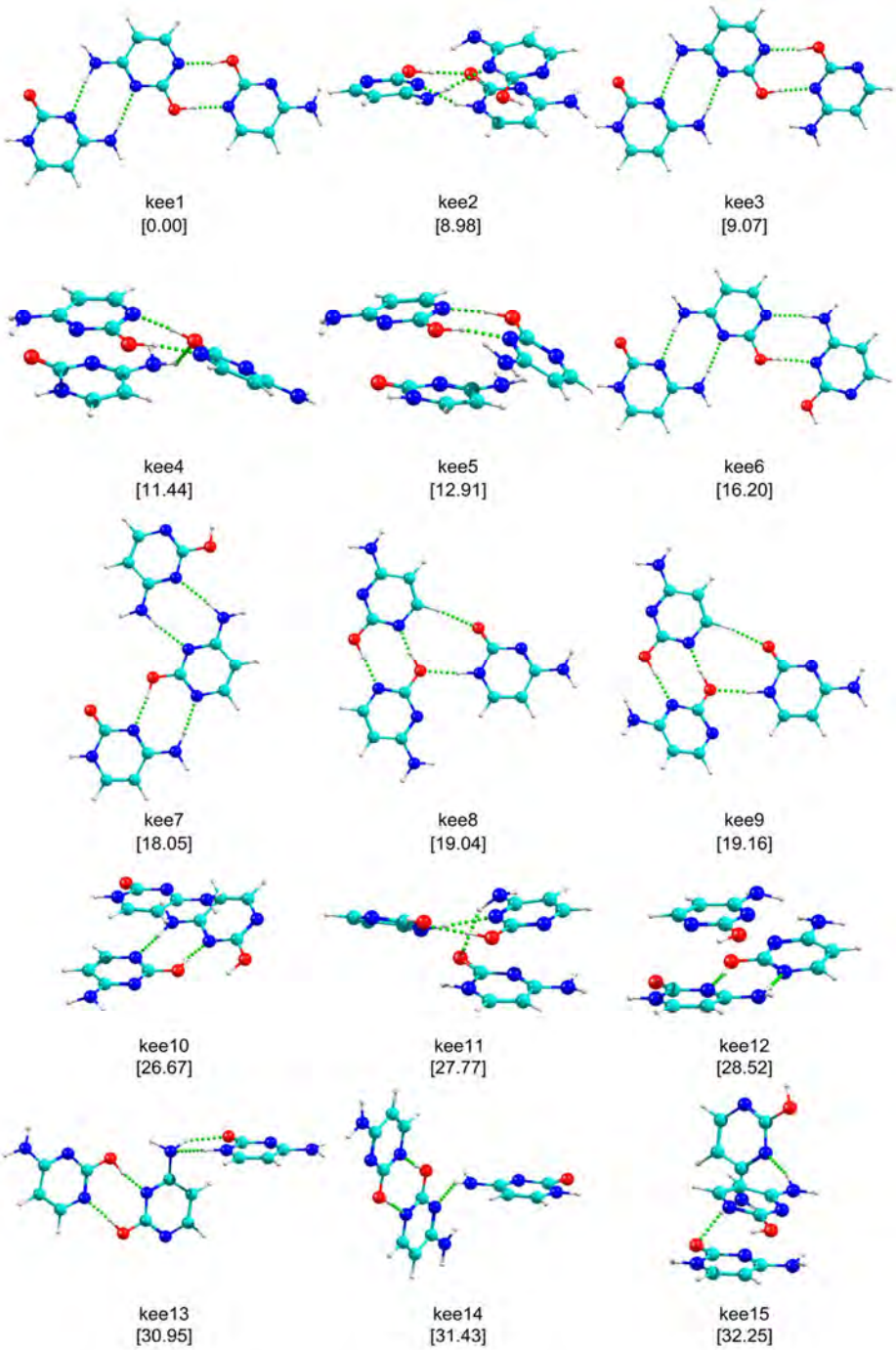


Figure 8.3.11 *Relative Gibbs free energy of the kkk trimers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

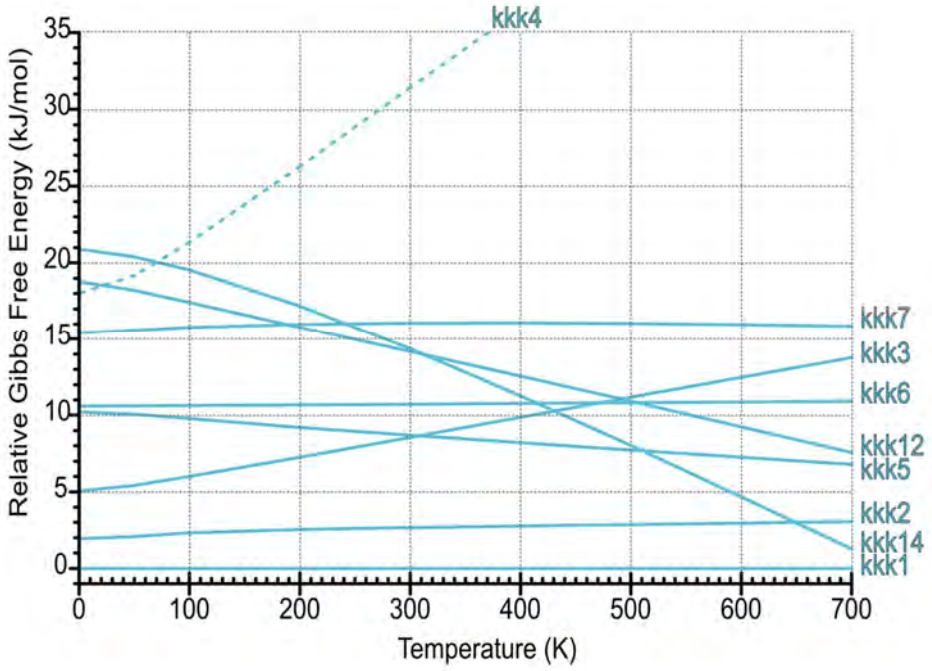


Figure 8.3.12 *Relative Gibbs free energy of the kke trimers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

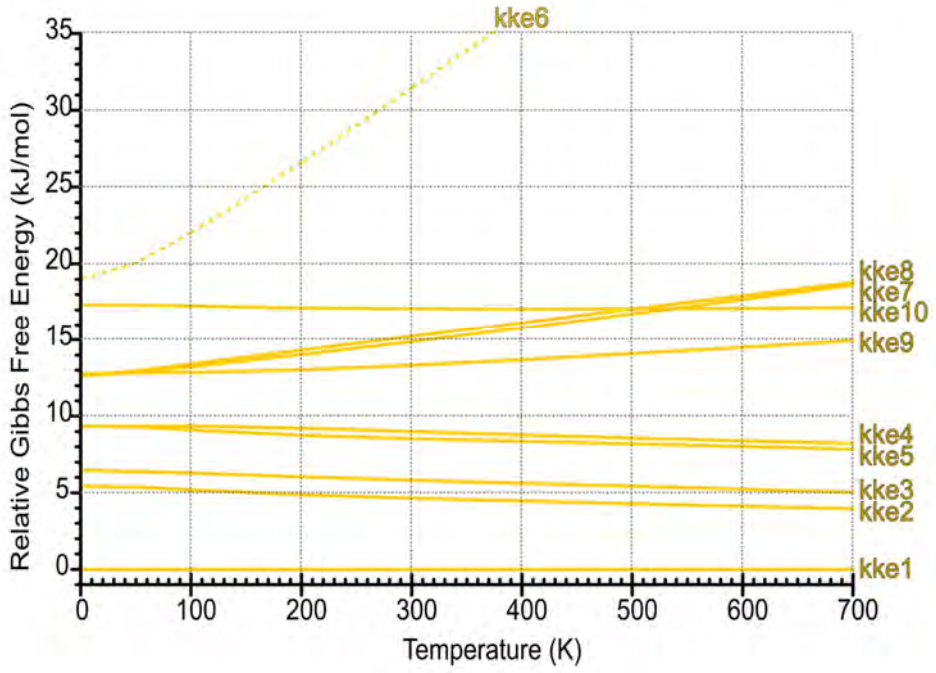


Figure 8.3.13 *Relative Gibbs free energy of the kee trimers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

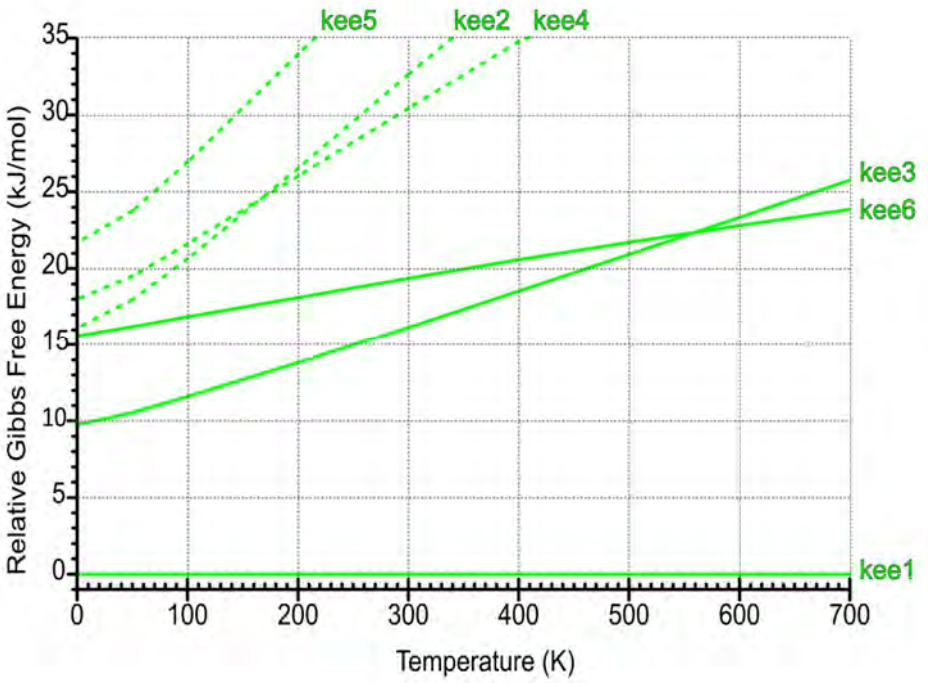


Figure 8.3.14 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered kkk trimers of cytosine.*

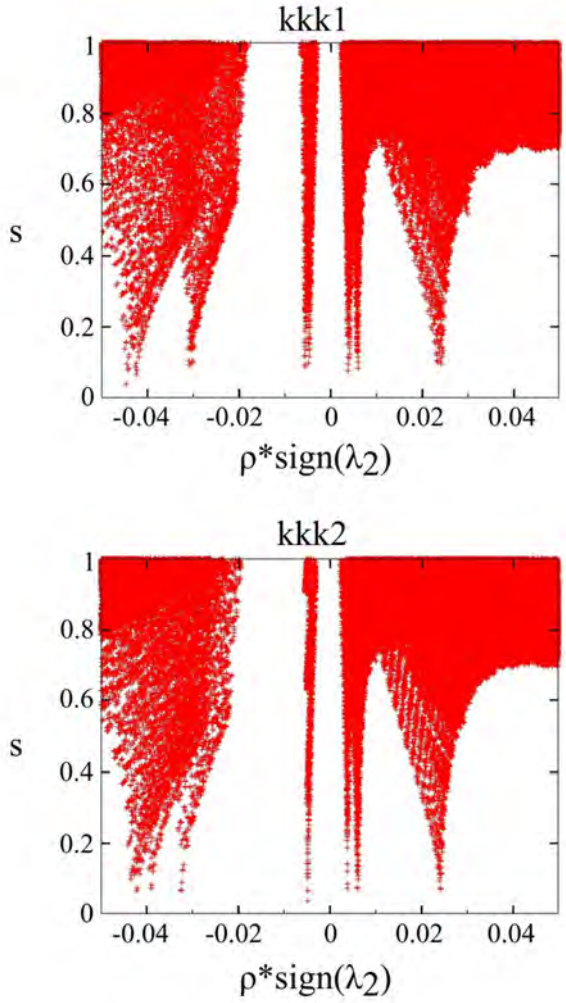


Figure 8.3.15 *IR spectra of trimer of cytosine, scanned with the UV wavelenght fixed at 33588 cm^{-1} (red trace), 33786 cm^{-1}, 33957 cm^{-1} and 34069 cm^{-1} (blak trace). The positions of the band has been jutted out with different color bars.*

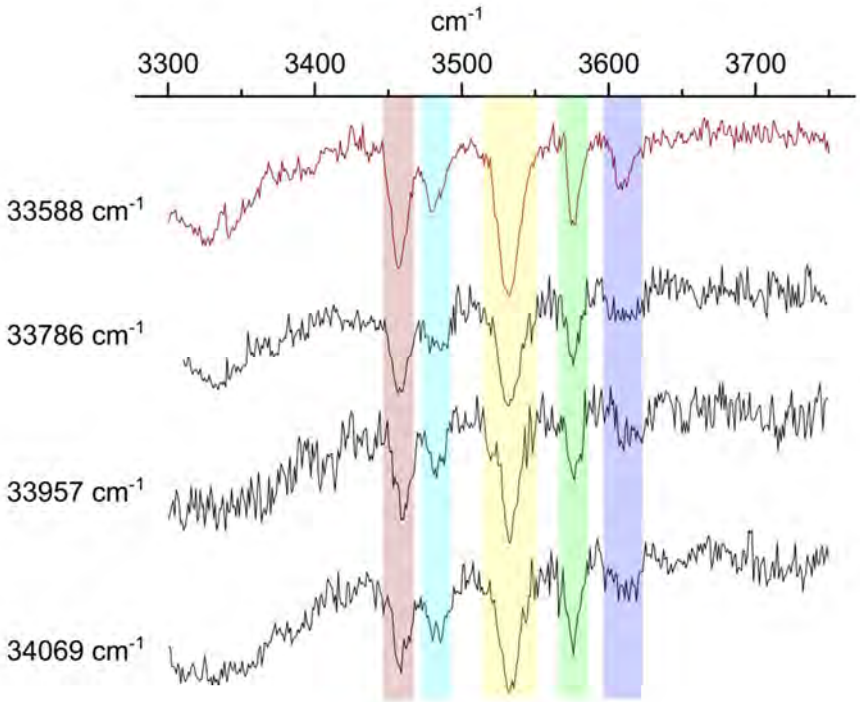
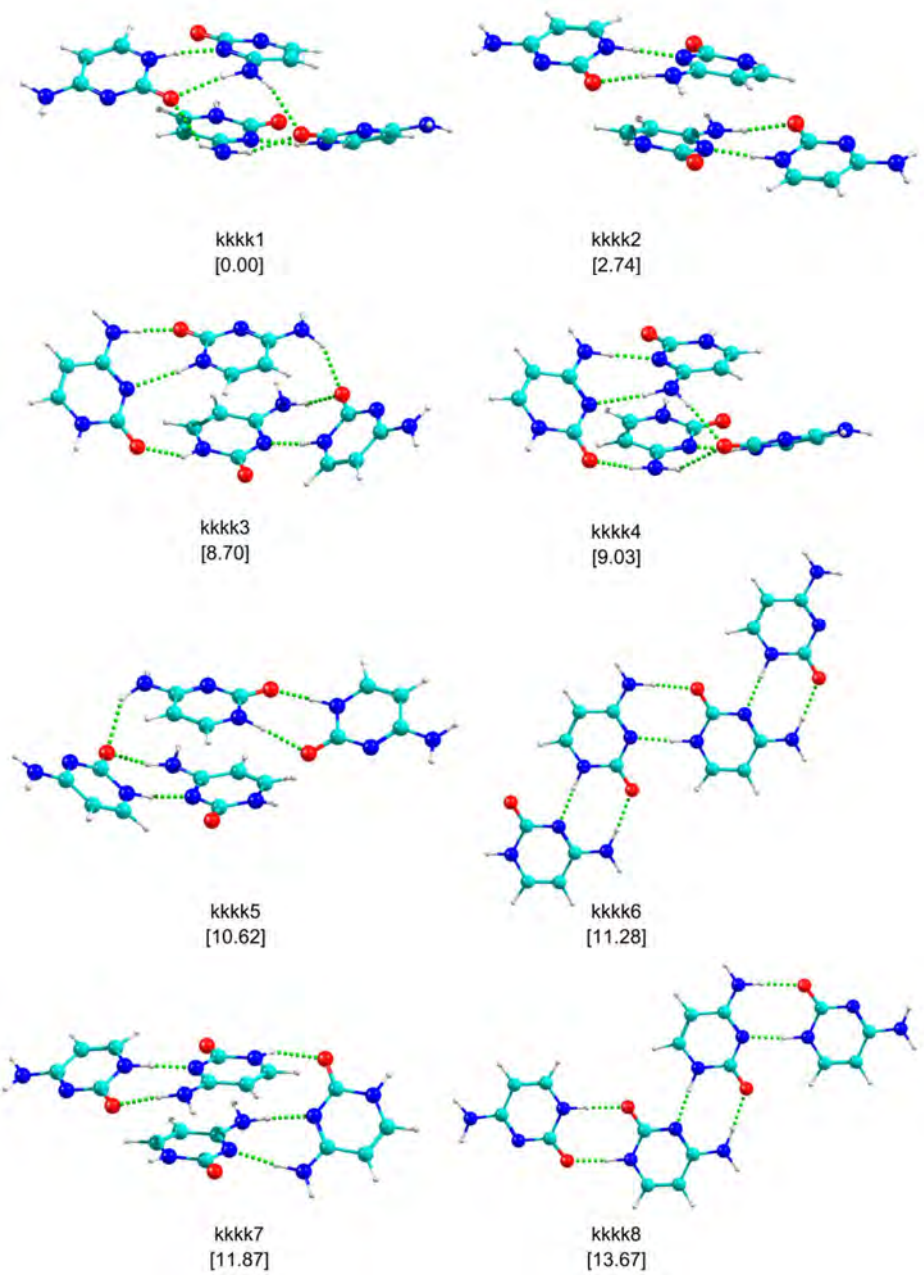
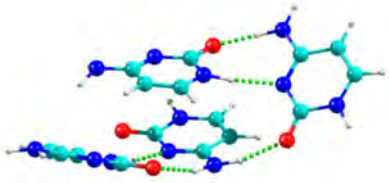
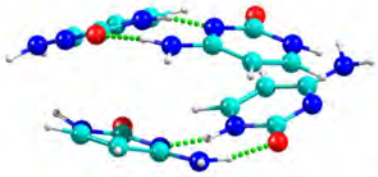


Figure 8.3.16 *The 15 most stable structures of kkkk cytosine tetramers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

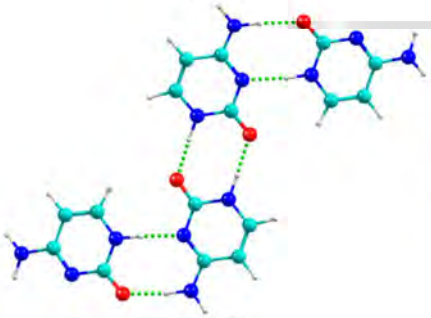




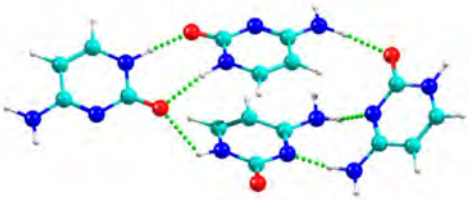
kkkk9
[14.31]



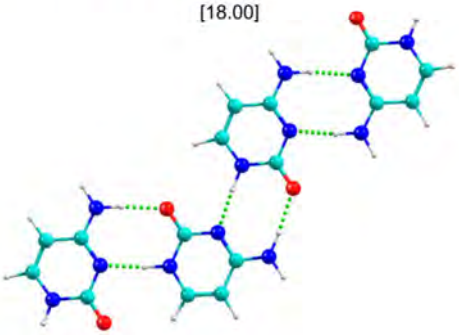
kkkk10
[14.67]



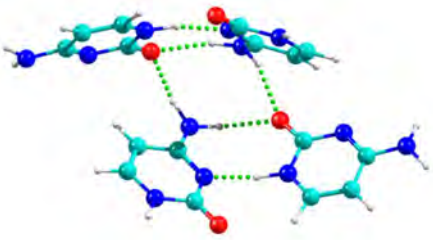
kkkk11
[18.00]



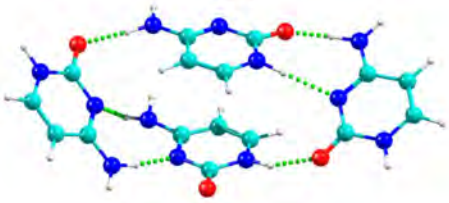
kkkk12
[19.63]



kkkk13
[19.85]

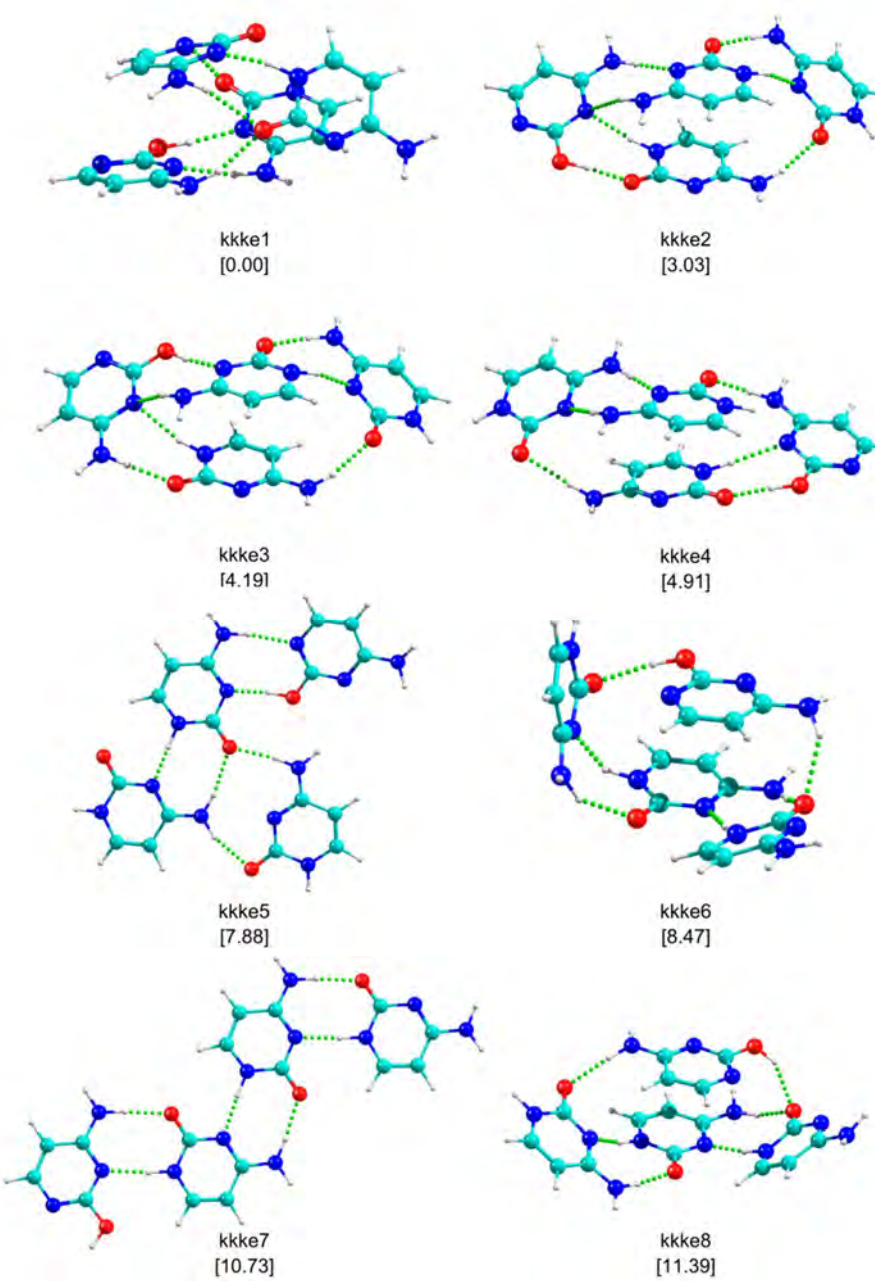


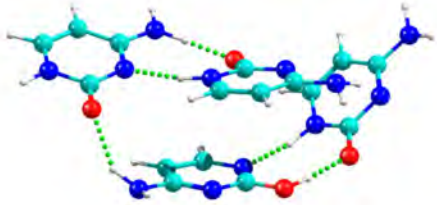
kkkk14
[20.64]



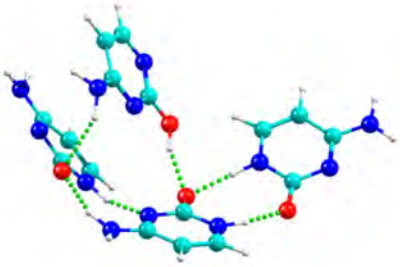
kkkk15
[21.21]

Figure 8.3.17 *The 15 most stable structures of kkke cytosine tetramers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

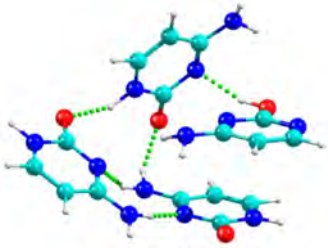




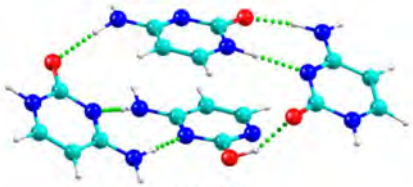
kkke9
[13.21]



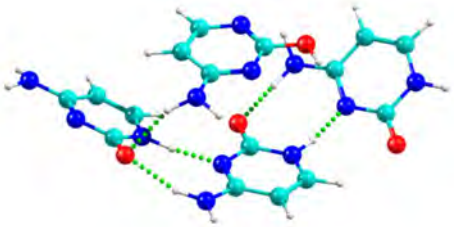
kkke10
[14.72]



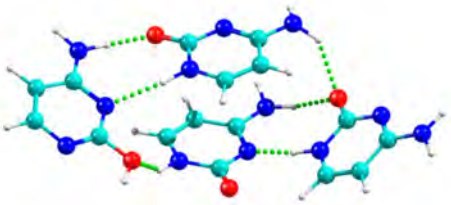
kkke11
[14.88]



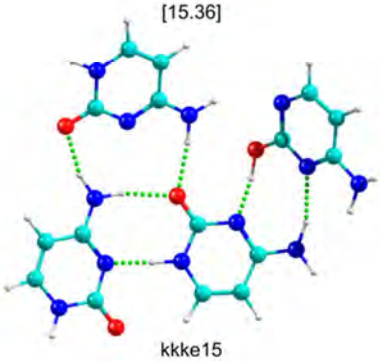
kkke12
[15.25]



kkke13
[15.36]

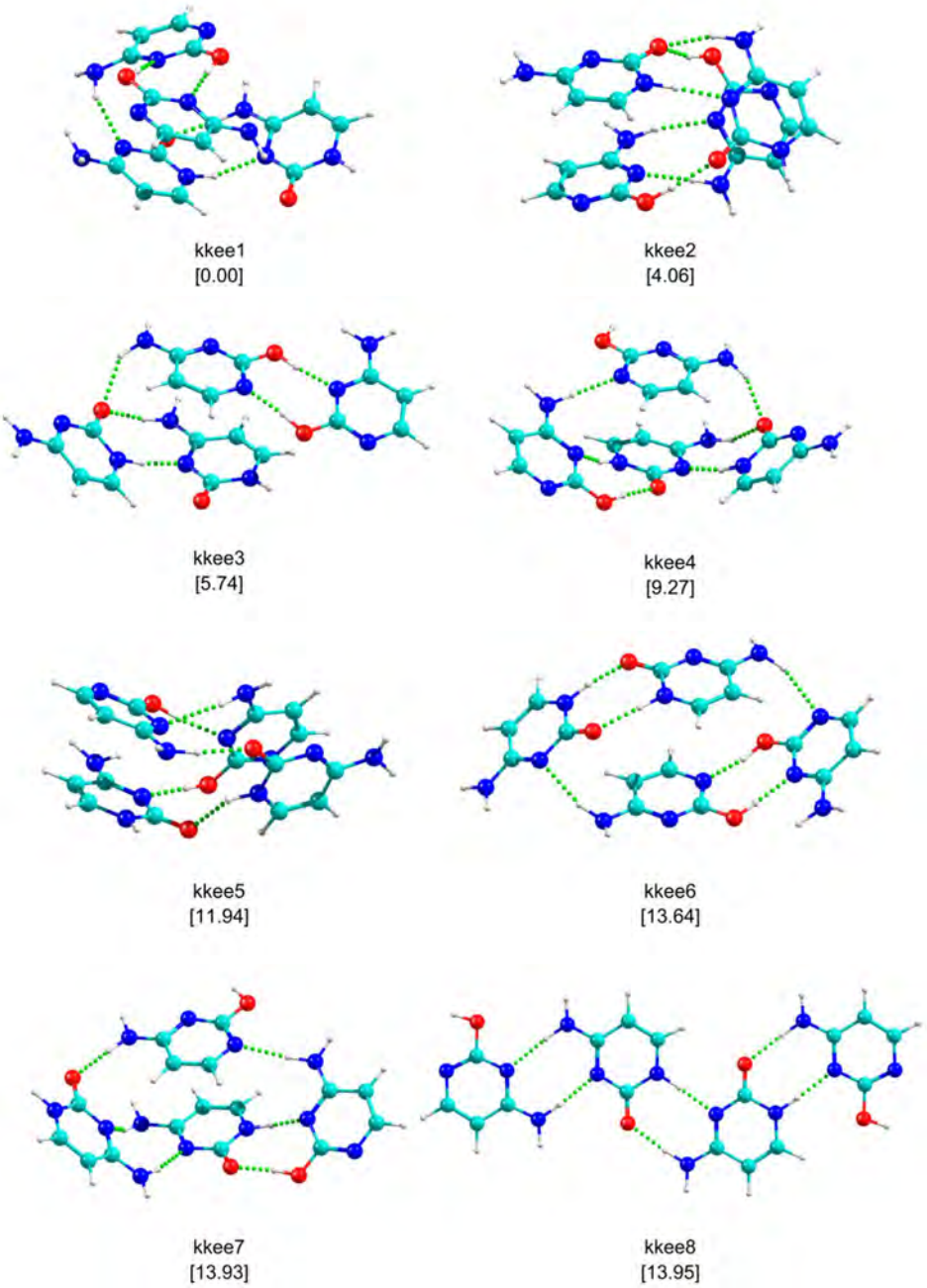


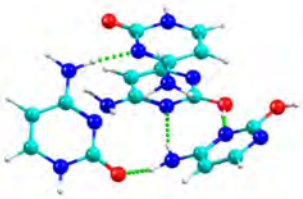
kkke14
[15.43]



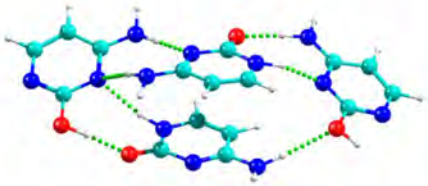
kkke15
[16.27]

Figure 8.3.18 *The 15 most stable structures of kkee cytosine tetramers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

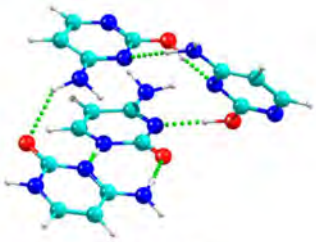




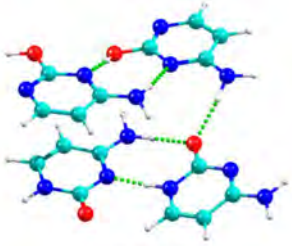
kkee9
[14.62]



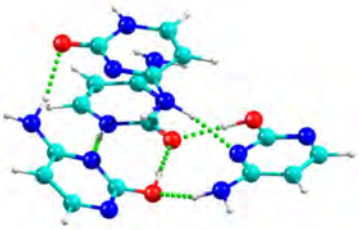
kkee10
[15.64]



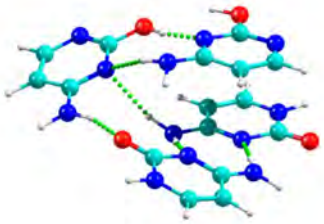
kkee11
[15.66]



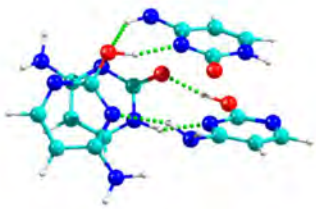
kkee12
[16.18]



kkee13
[17.32]

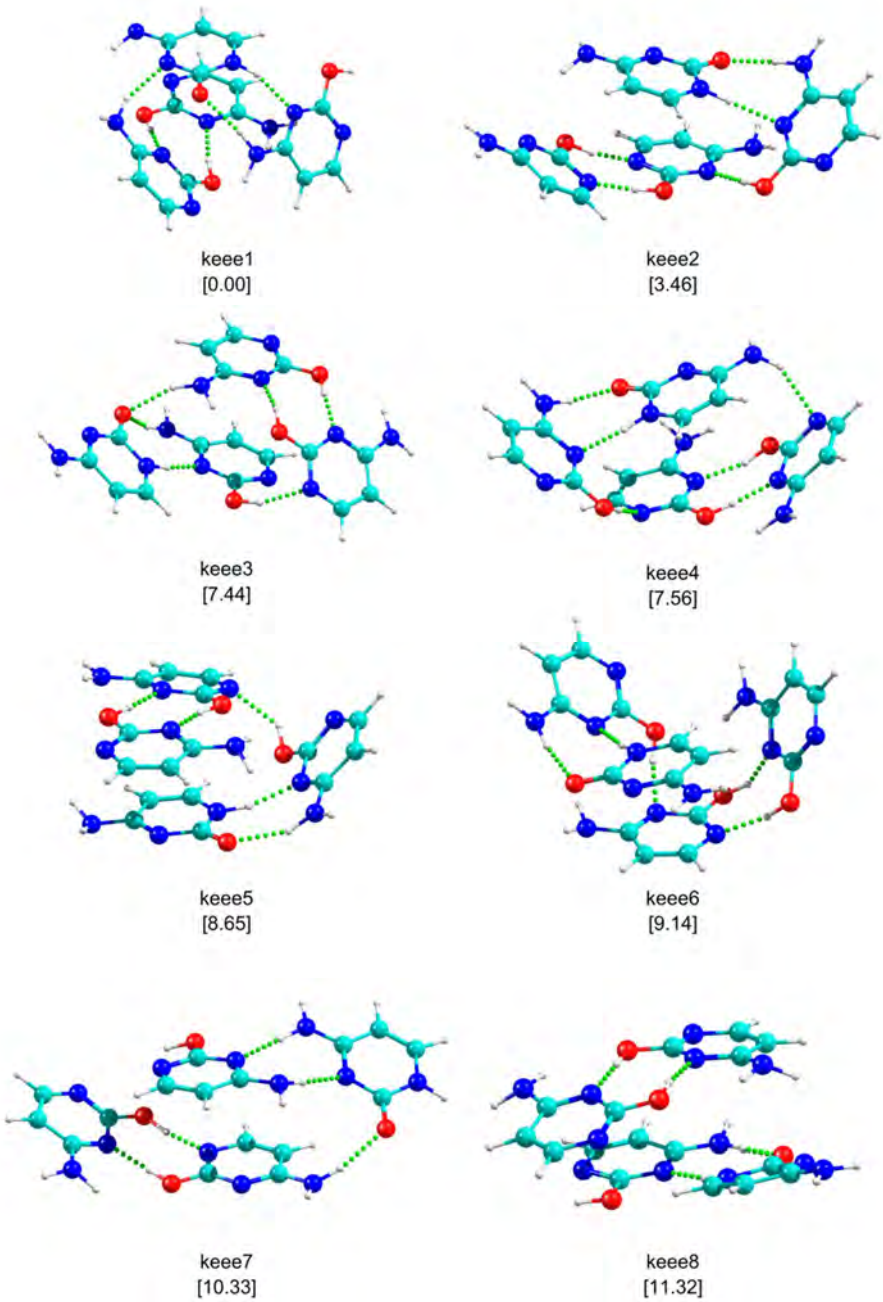


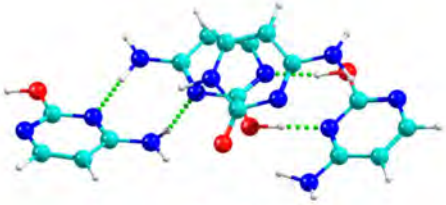
kkee14
[17.83]



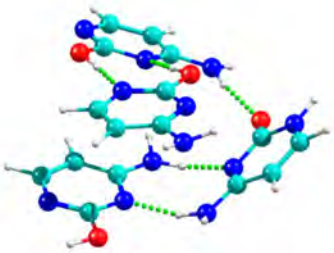
kkee15
[18.11]

Figure 8.3.19 *The 15 most stable structures of keee cytosine tetramers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

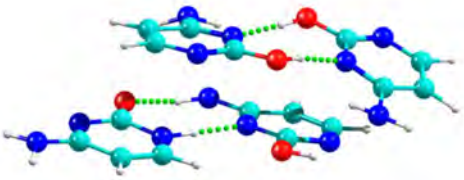




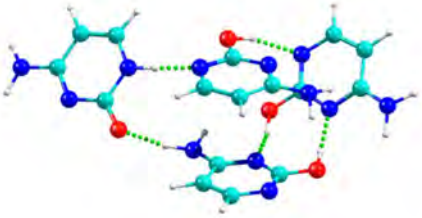
keee9
[11.69]



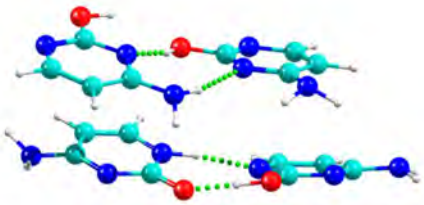
keee10
[12.43]



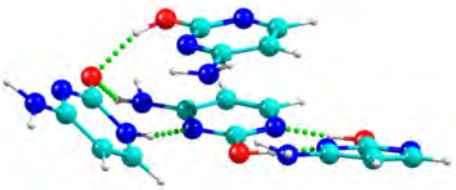
keee11
[13.49]



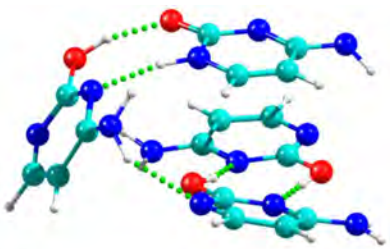
keee12
[15.59]



keee13
[19.75]



keee14
[19.81]



keee15
[21.24]

Figure 8.3.20 *Relative Gibbs free energy of the kkkk tetramers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

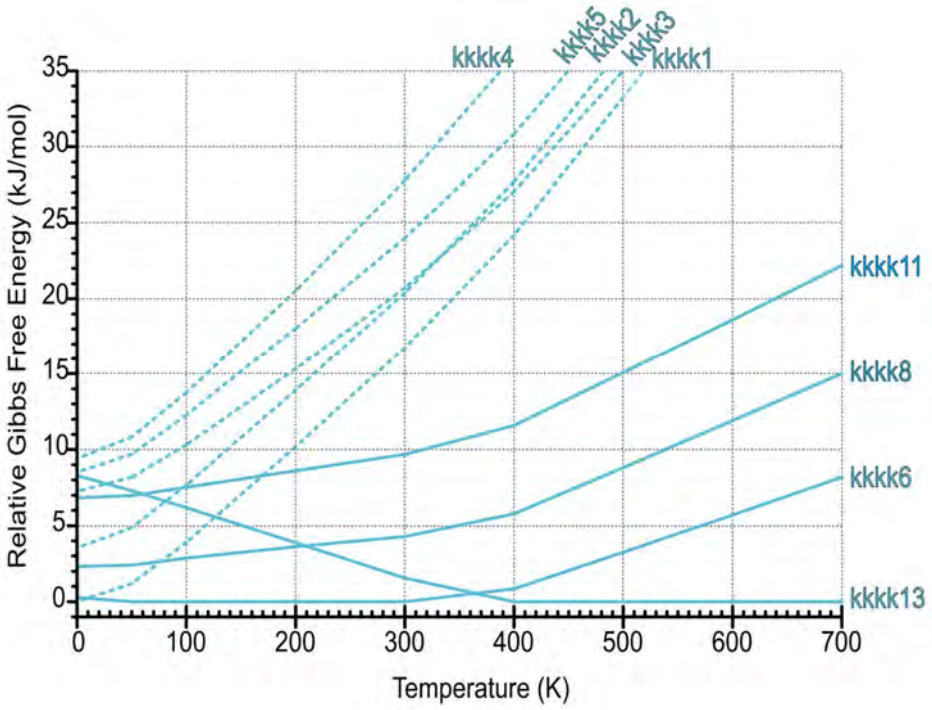


Figure 8.3.21 *Relative Gibbs free energy of the kkke tetramers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

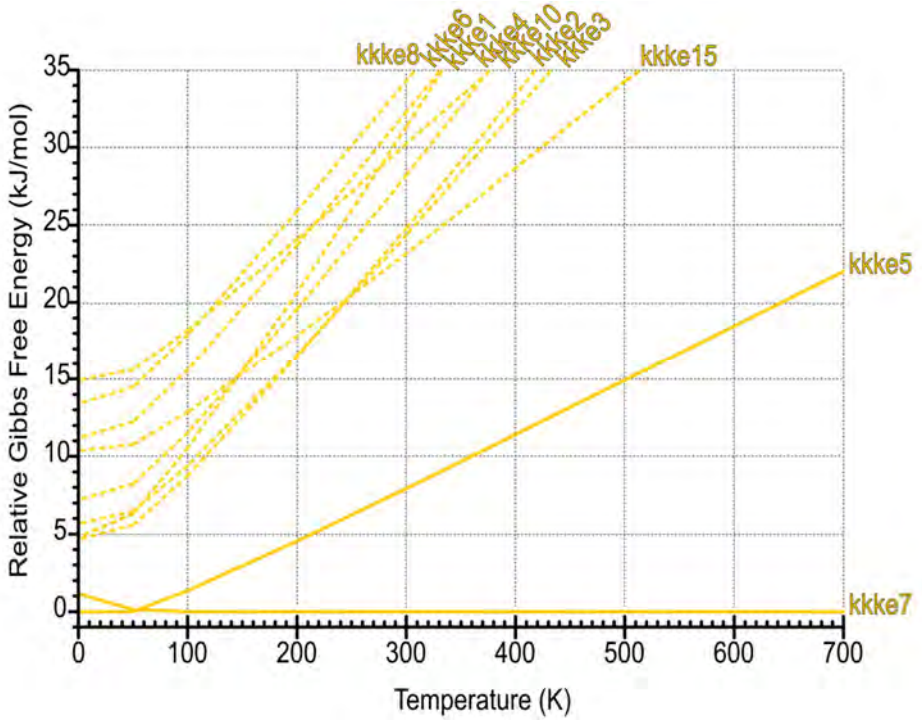


Figure 8.3.22 *Relative Gibbs free energy of the kkee tetramers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

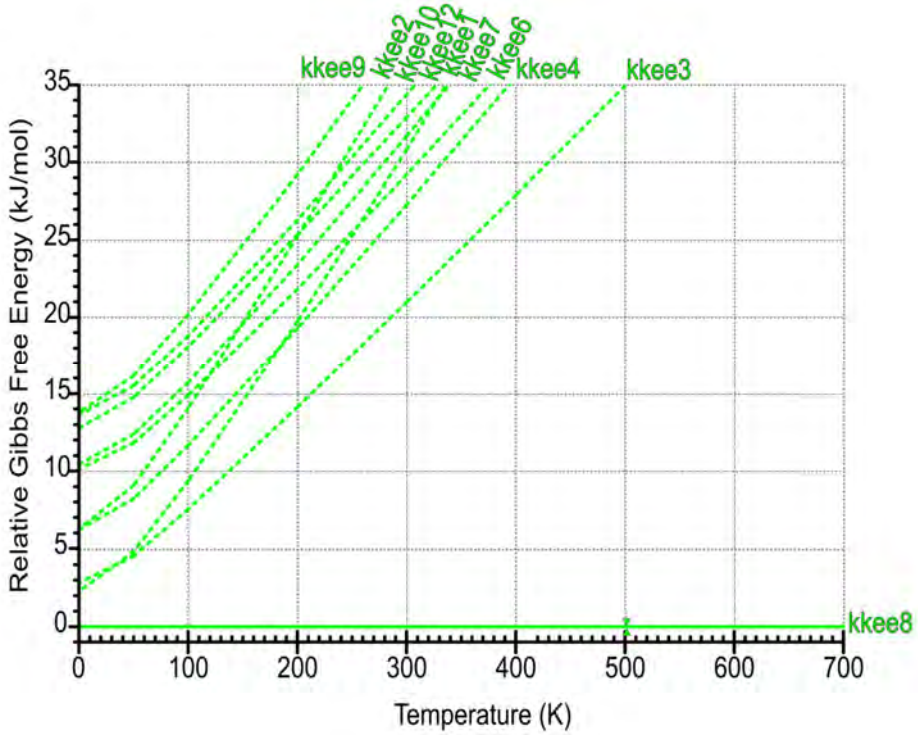


Figure 8.3.23 *Relative Gibbs free energy of the keee tetramers of cytosine in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

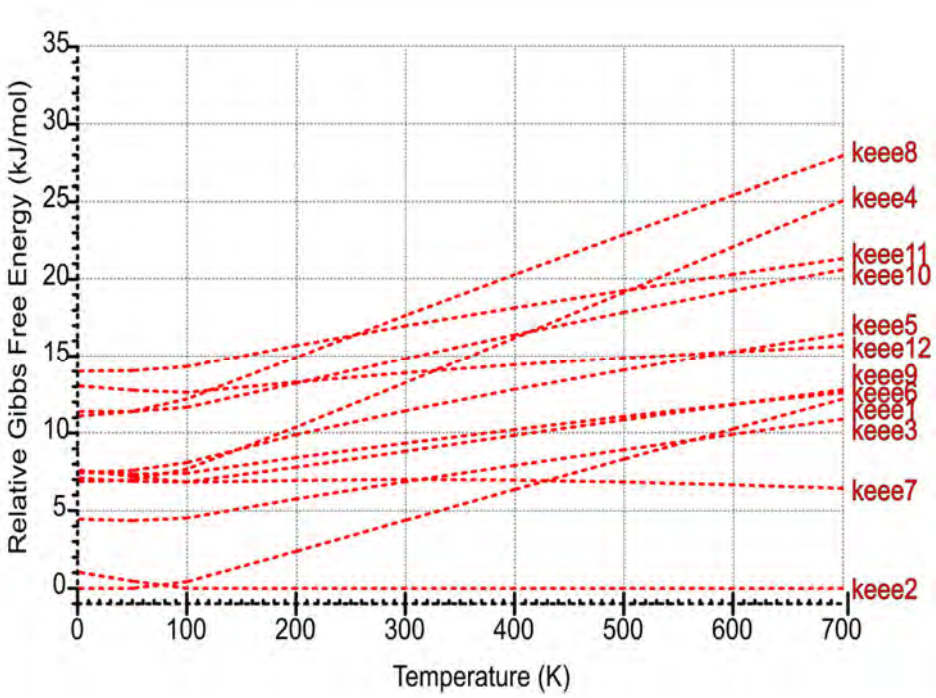


Figure 8.3.24 *2D-NCI plots (s vs. ρ*sign(λ2)) for the considered kkkk tetramers of cytosine.*

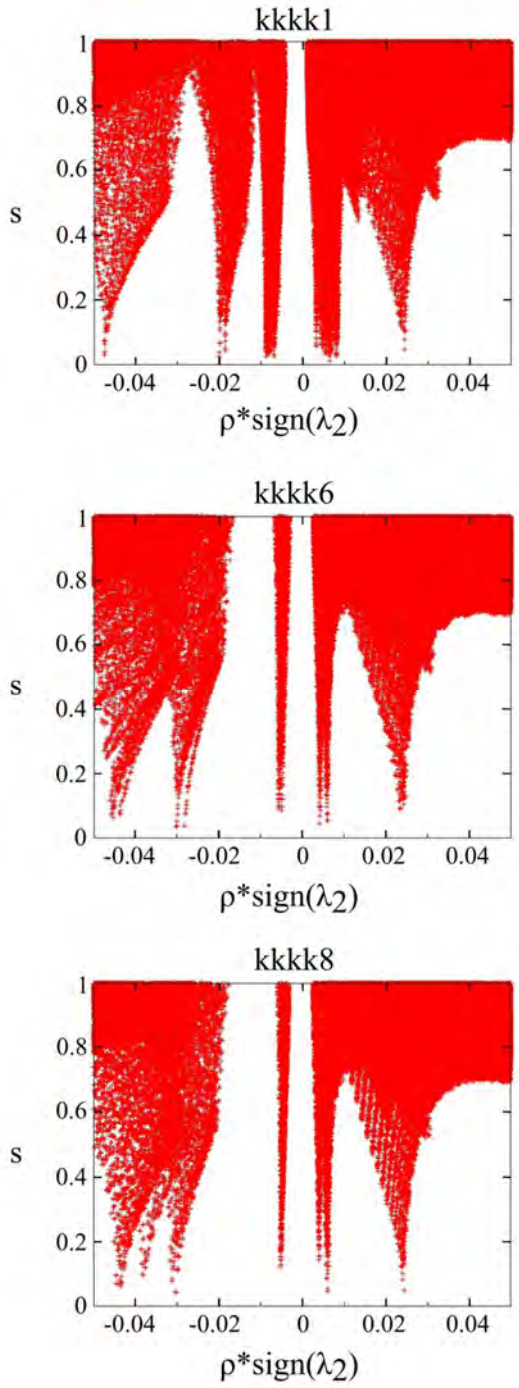


Figure 8.3.25 *Relative Gibbs free energy of some selected calculated species of cytosine A) monomer; B) dimer; C) trimer; D) tetramer in the 0-700 K interval,assuming a non-equilibrium system with a rotational temperature around T=4K.*

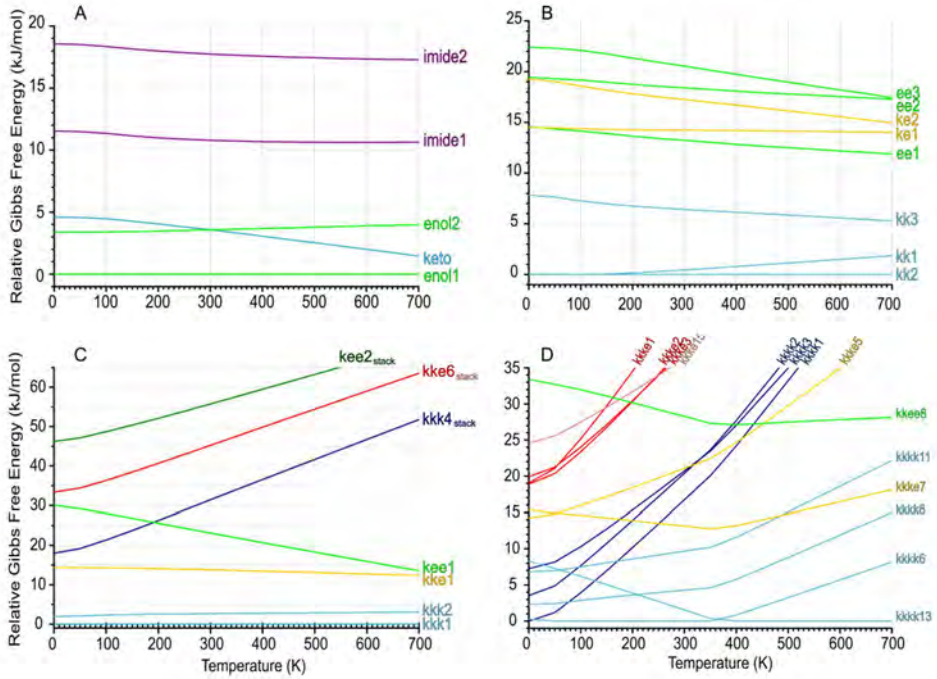


Figure 8.3.26 *The 14 most stable structures of 1-methylcytosine dimers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

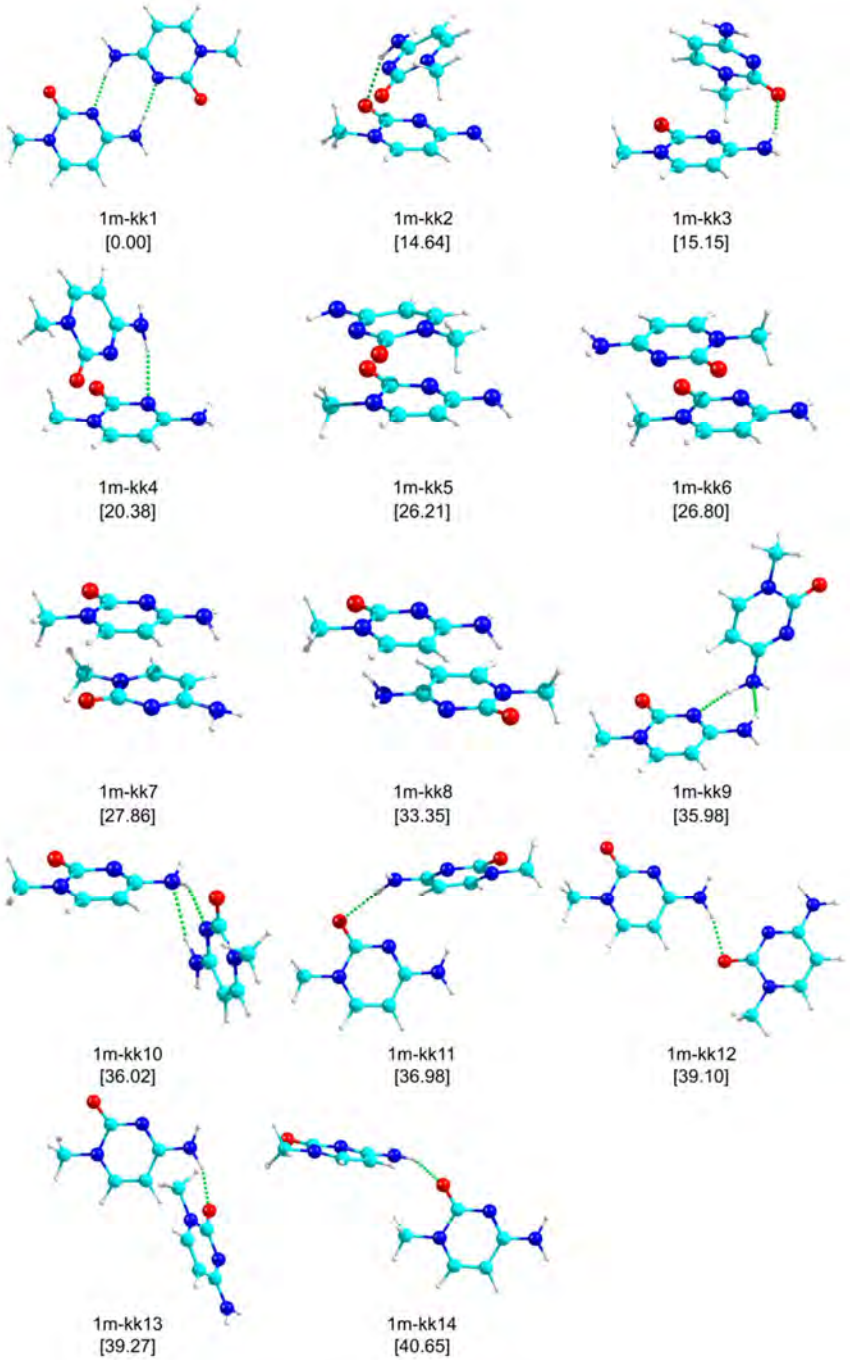


Figure 8.3.27 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered kk dimer of 1-methylcytosine.*

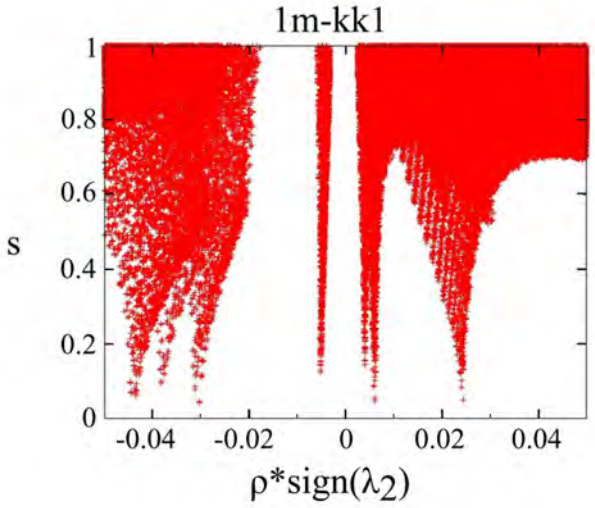


Figure 8.3.28 *The 15 most stable structures of 1-methylcytosine trimers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

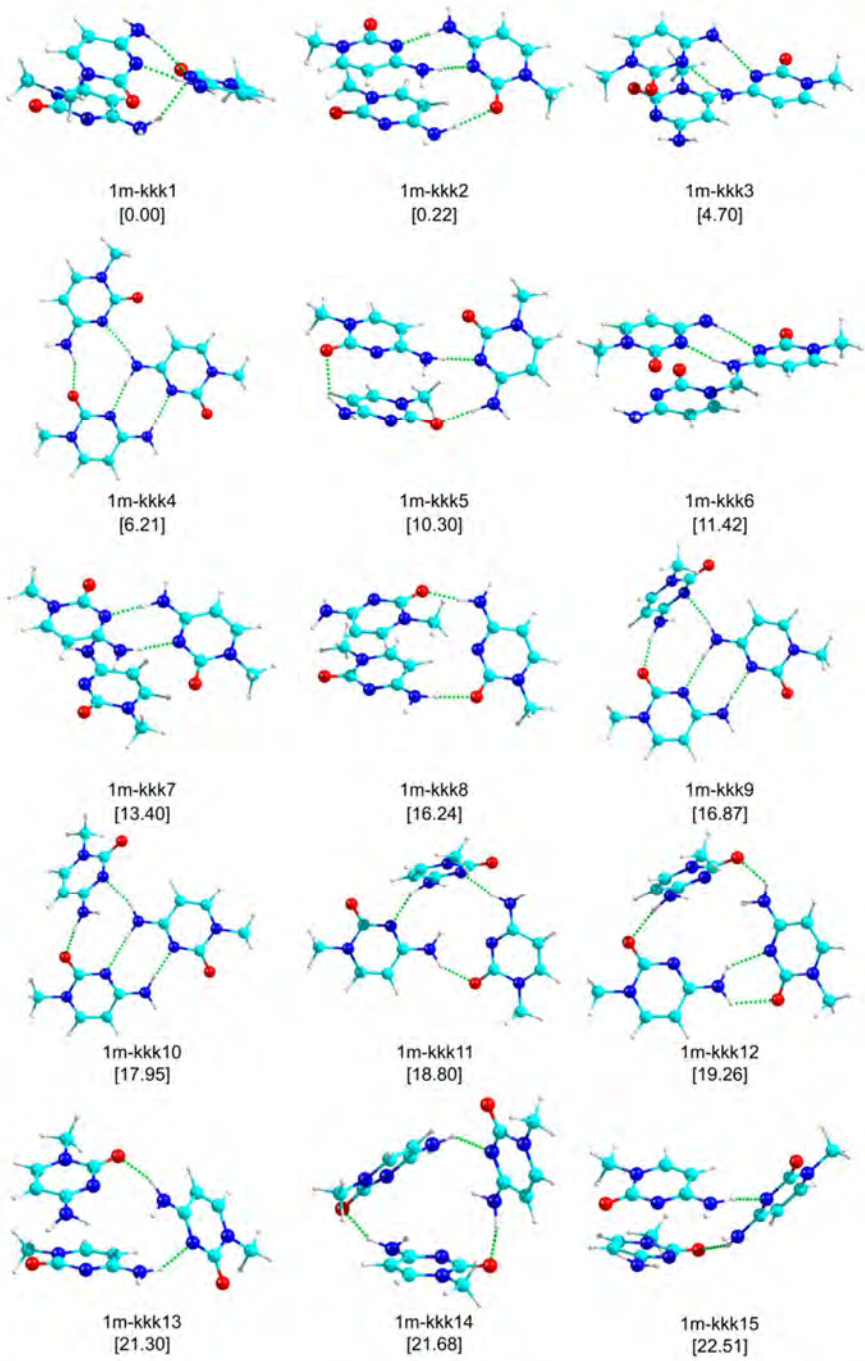


Figure 8.3.29 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered kkk trimer of 1-methylcytosine.*

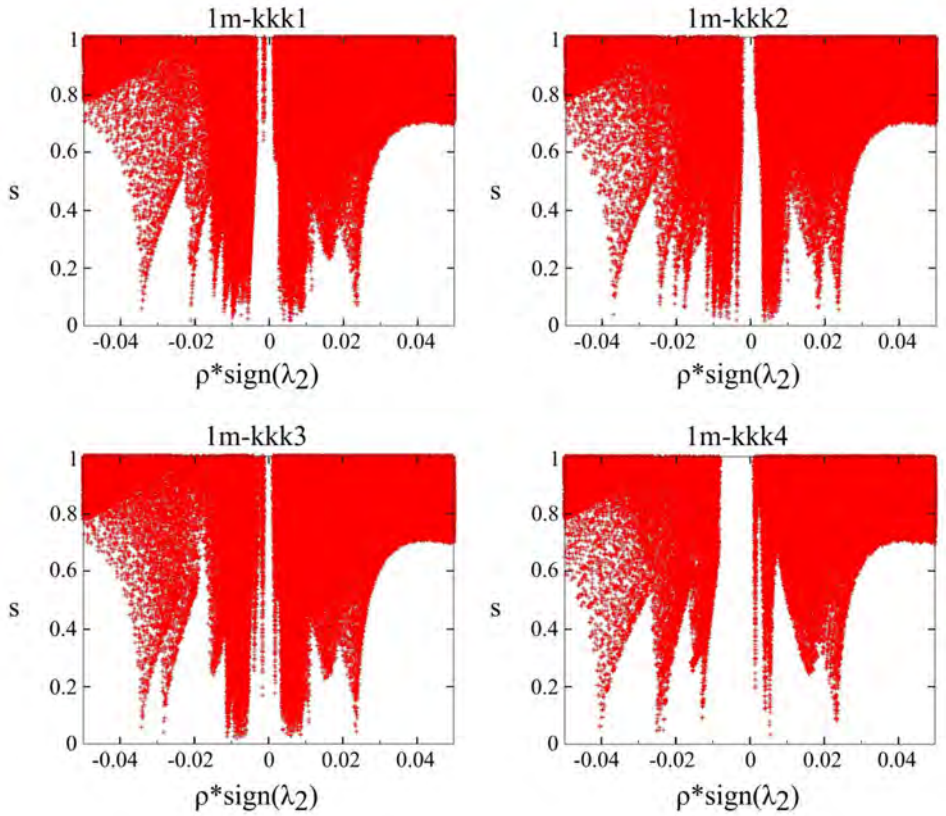
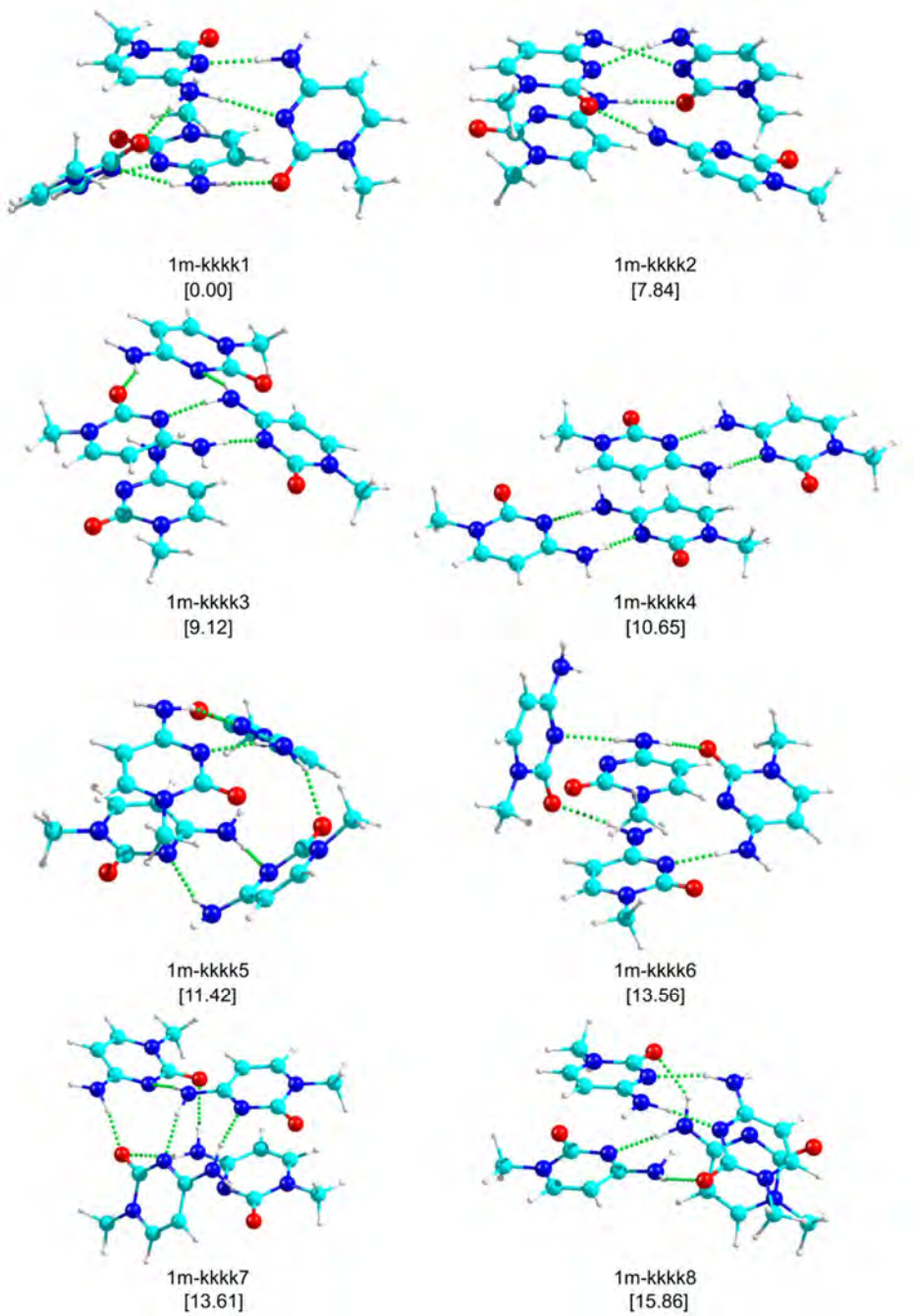
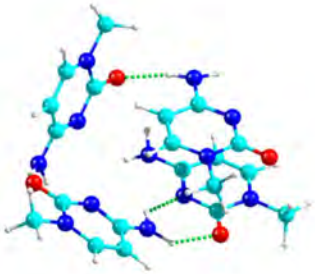
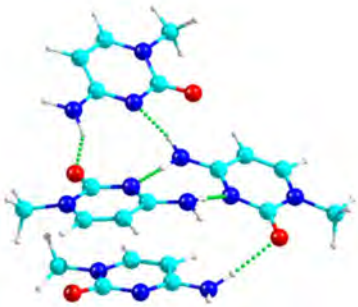


Figure 8.3.30 *The 15 most stable structures of 1-methylcytosine tetramers at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. The first planar structures (1m-kkkk38) was also shown. ZPE correction was applied to the energy values.*

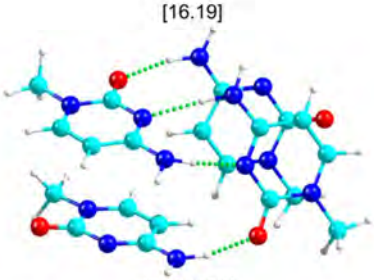




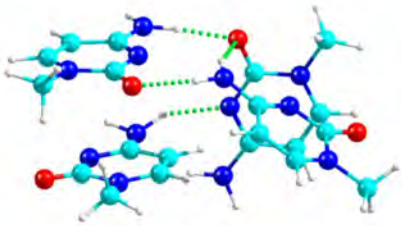
1m-kkkk9
[16.19]



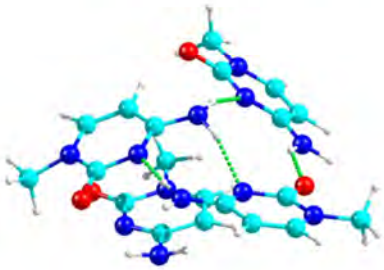
1m-kkkk10
[16.81]



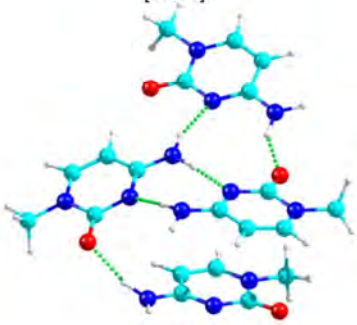
1m-kkkk11
[16.92]



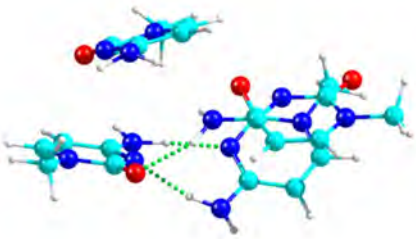
1m-kkkk12
[17.53]



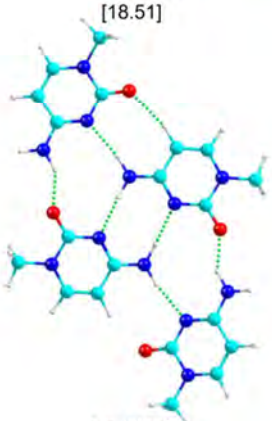
1m-kkkk13
[18.29]



1m-kkkk14
[18.51]



1m-kkkk15
[19.60]



1m-kkkk38
[26.97]

Figure 8.3.31 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered kkkk tetramer of 1-methylcytosine.*

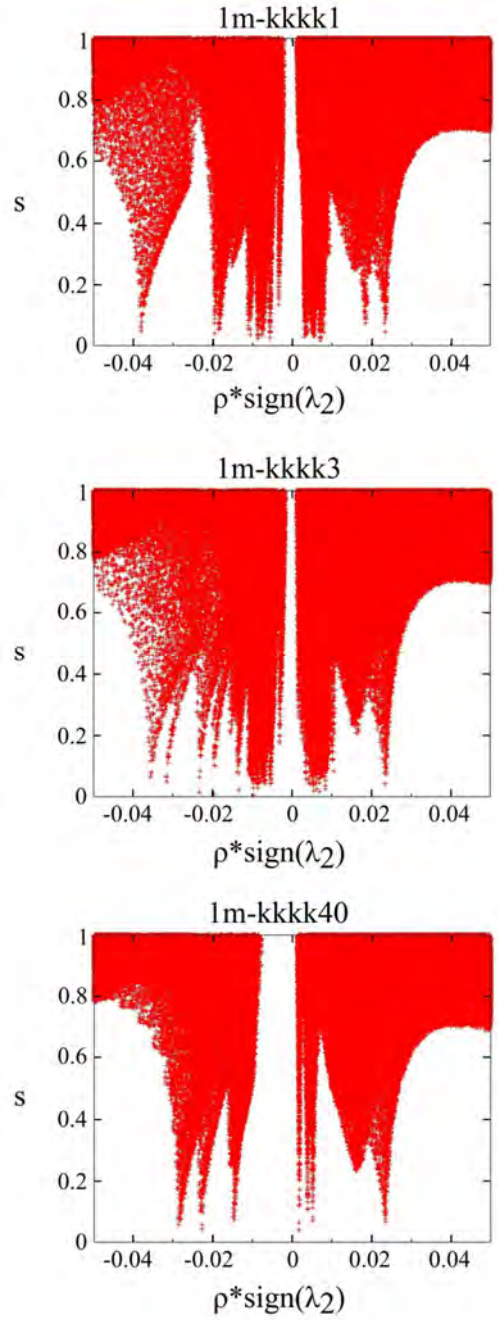


Figure 8.3.32 *The 6 most stable structures of 9k9k guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

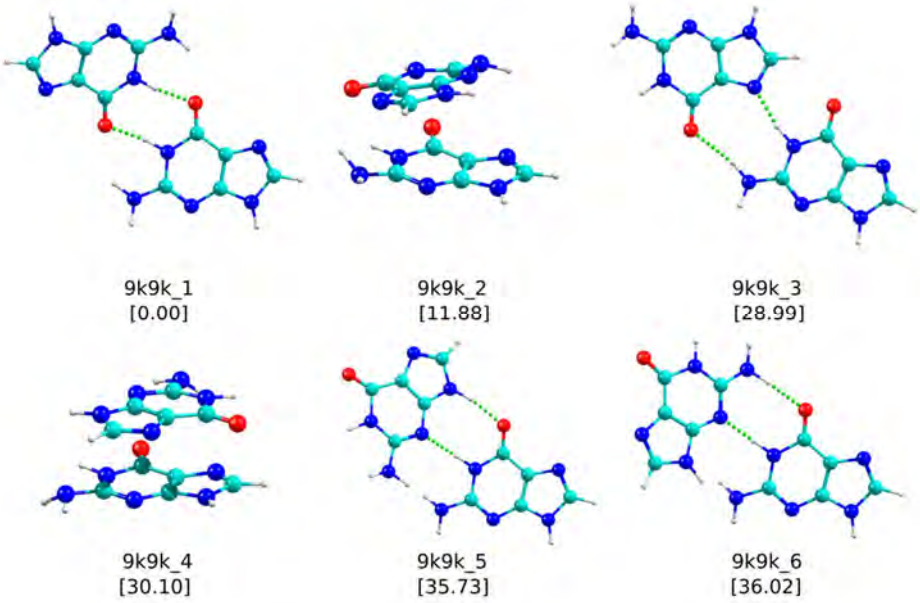


Figure 8.3.33 *The 13 most stable structures of 9k7k guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

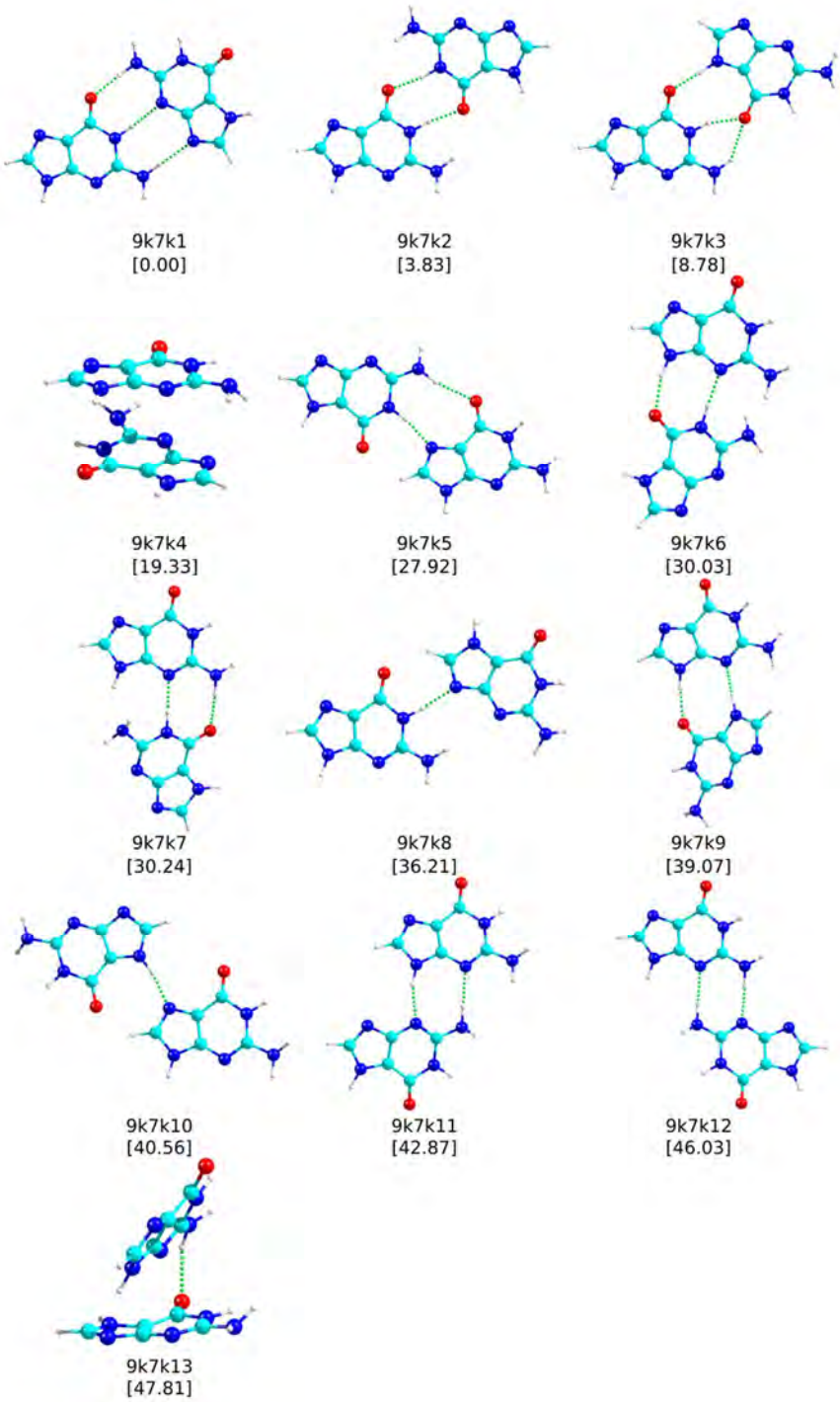


Figure 8.3.34 *The 15 most stable structures of 9k7ik guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

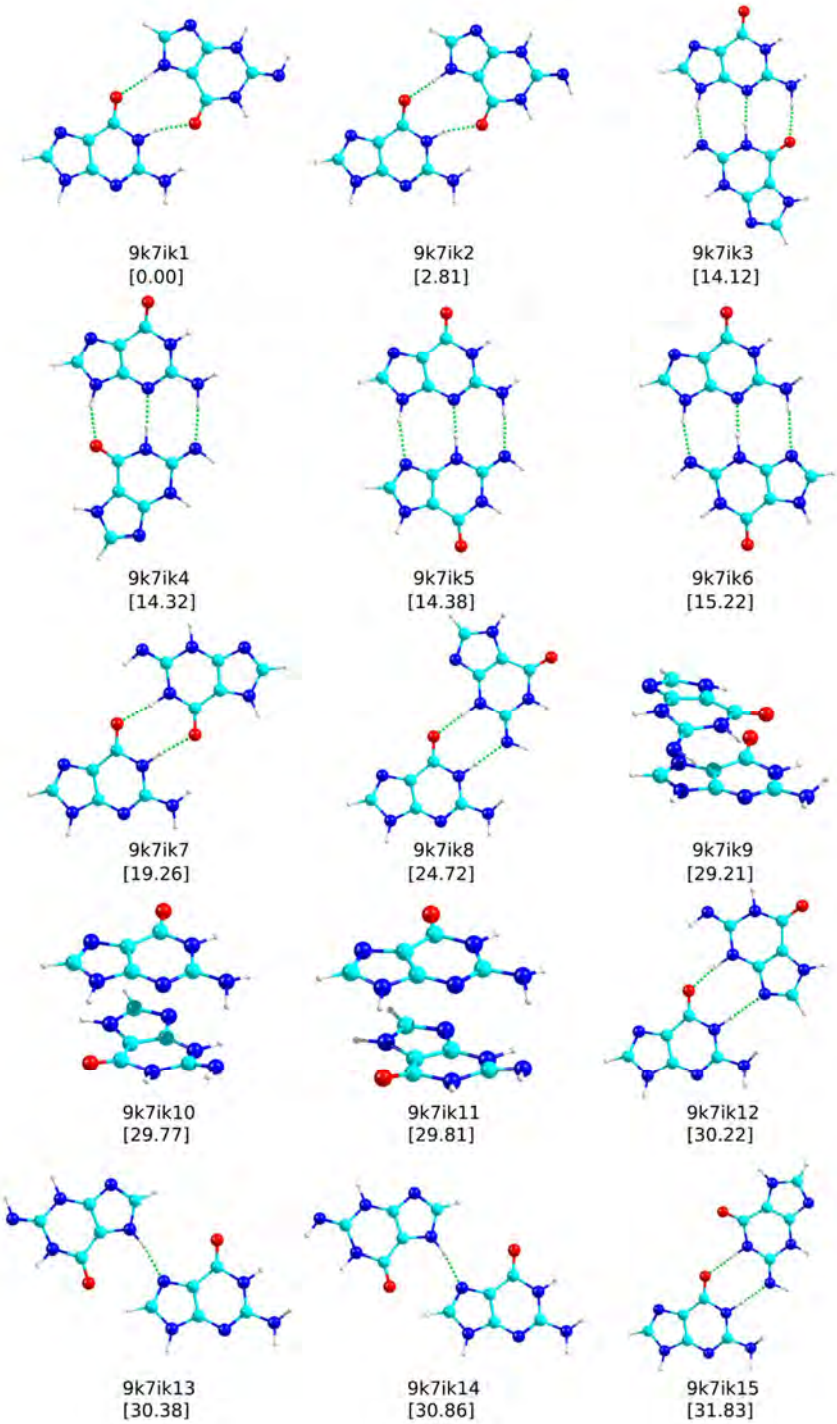


Figure 8.3.35 *The 15 most stable structures of 9k9e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

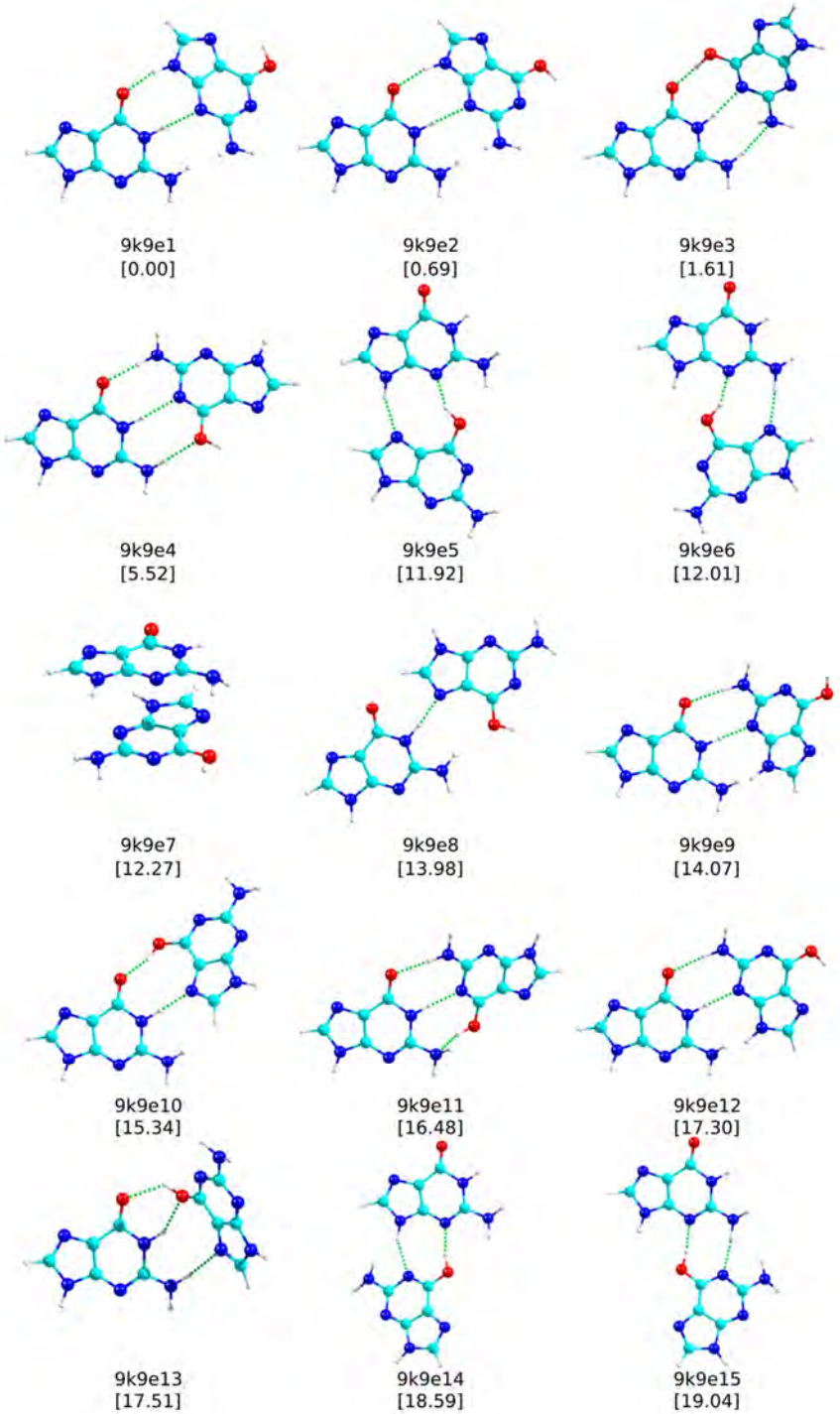


Figure 8.3.36 *The 15 most stable structures of 9k7e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

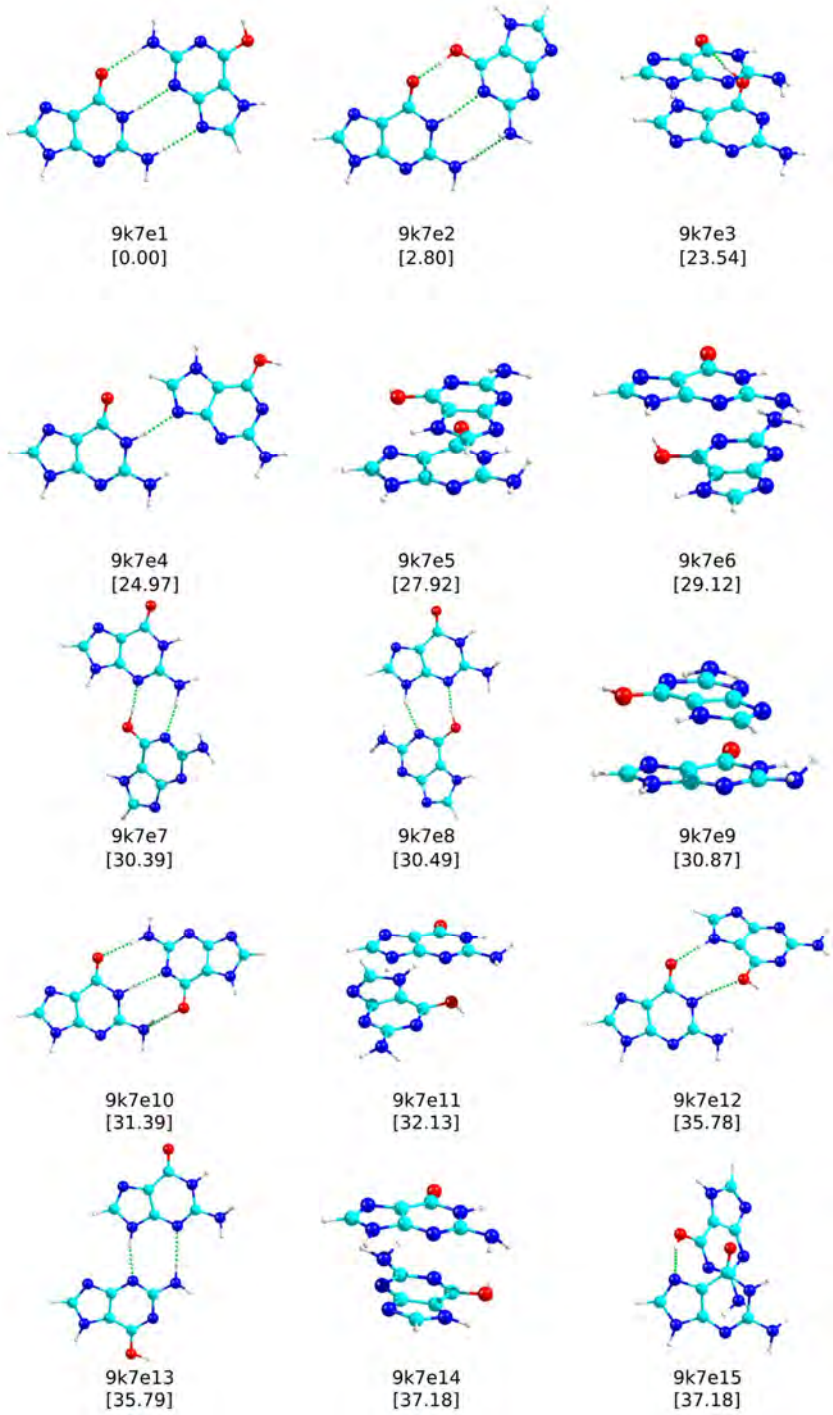


Figure 8.3.37 *The 13 most stable structures of 7k7k guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

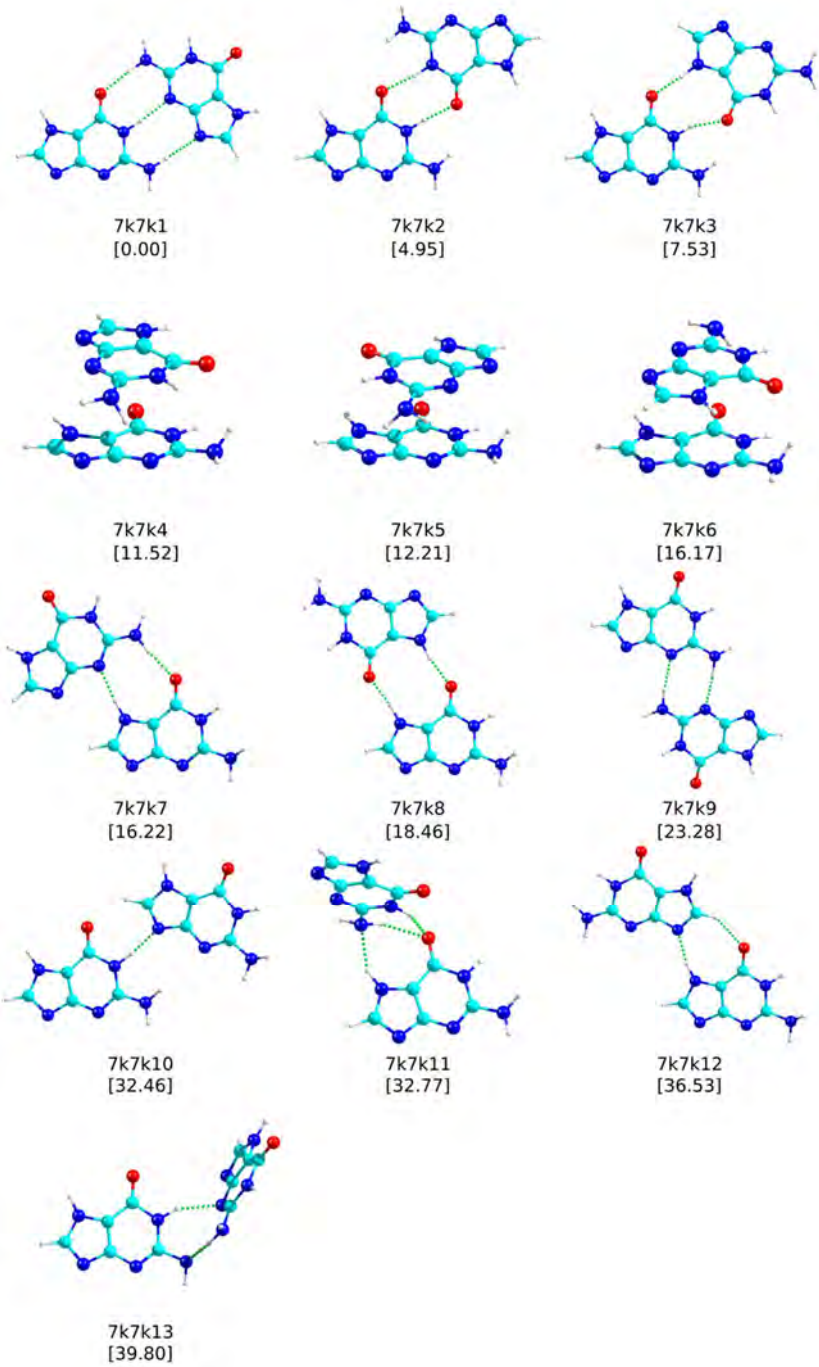


Figure 8.3.38 *The 15 most stable structures of 7k9e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

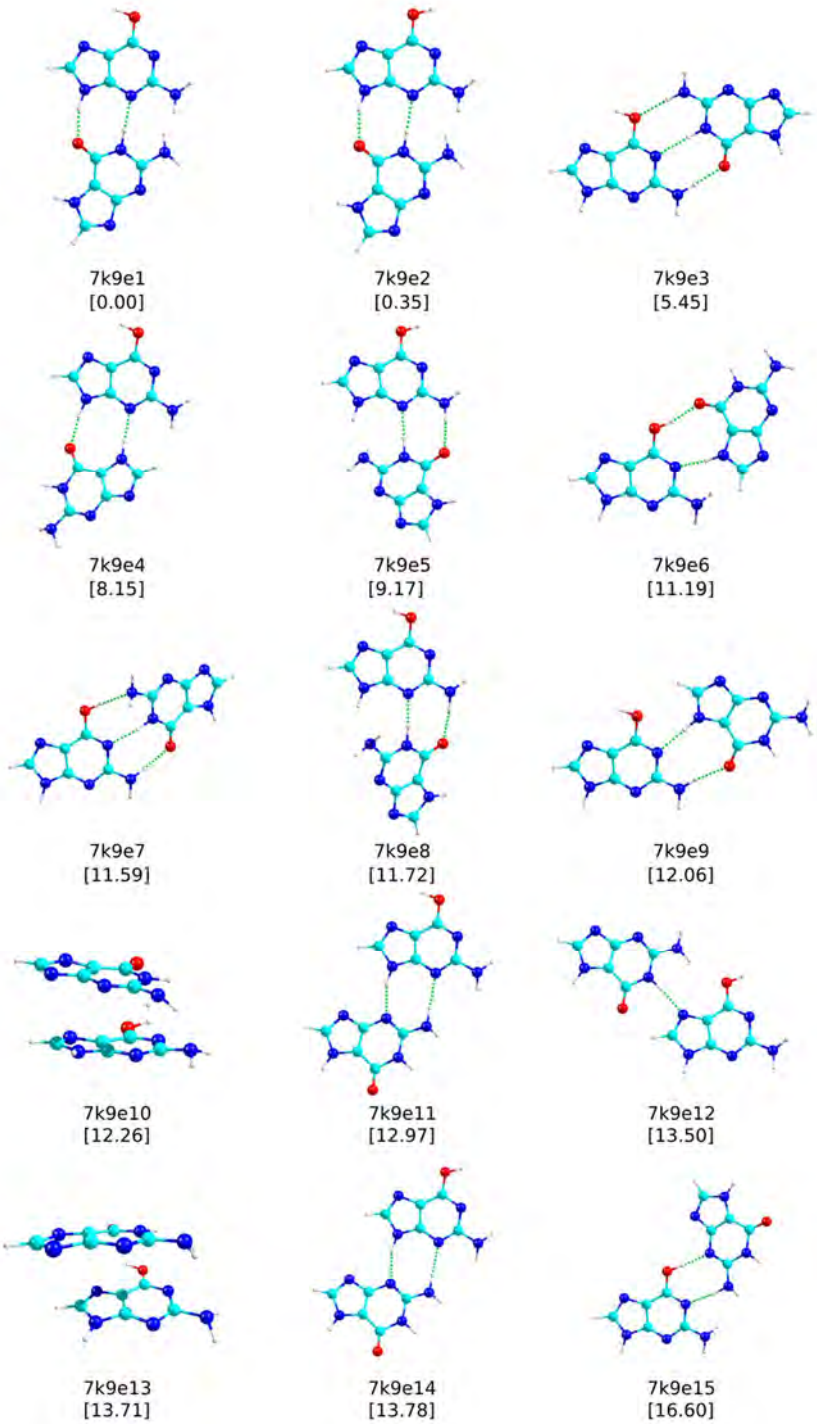


Figure 8.3.39 *The 15 most stable structures of 7k7e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

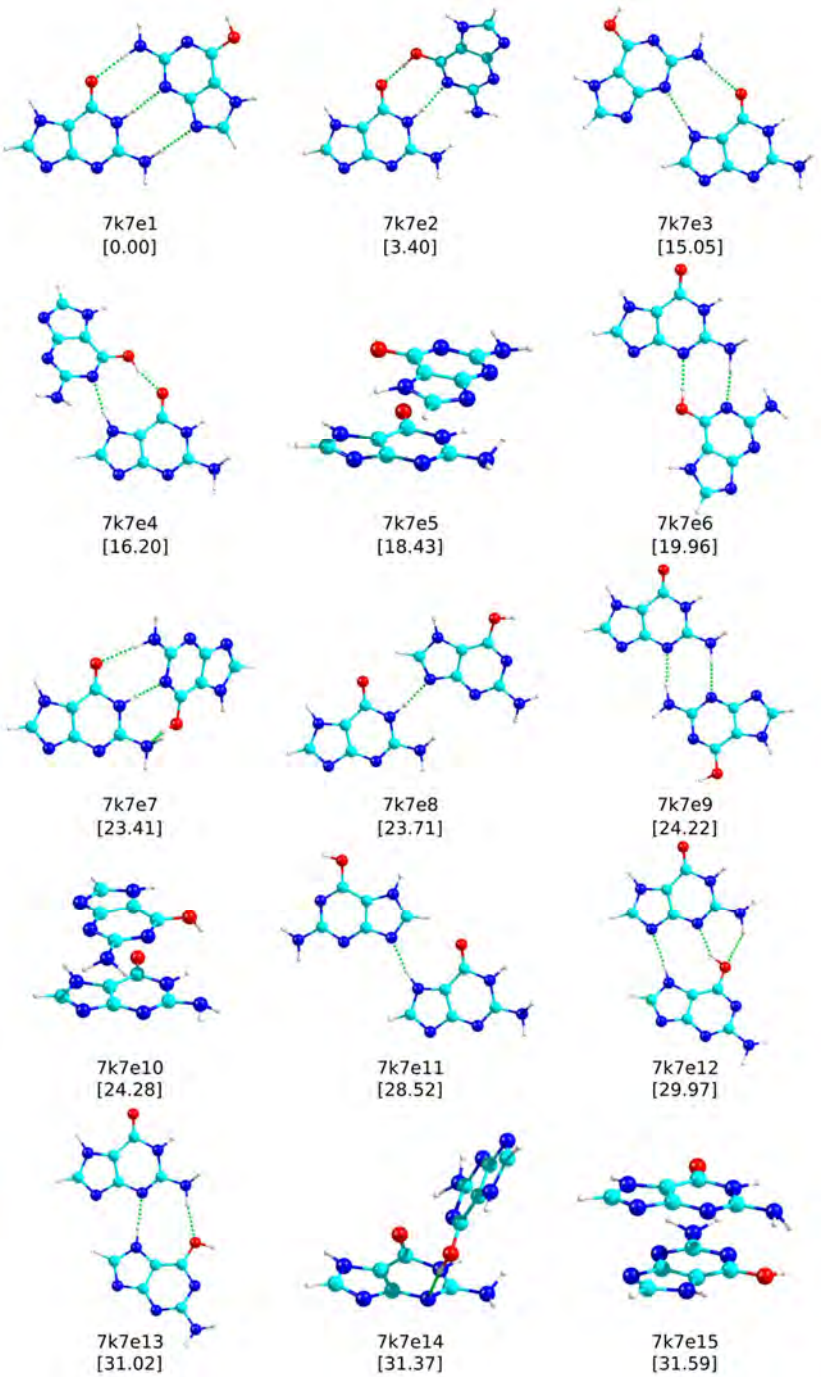


Figure 8.3.40 *The 15 most stable structures of 7k7ik guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

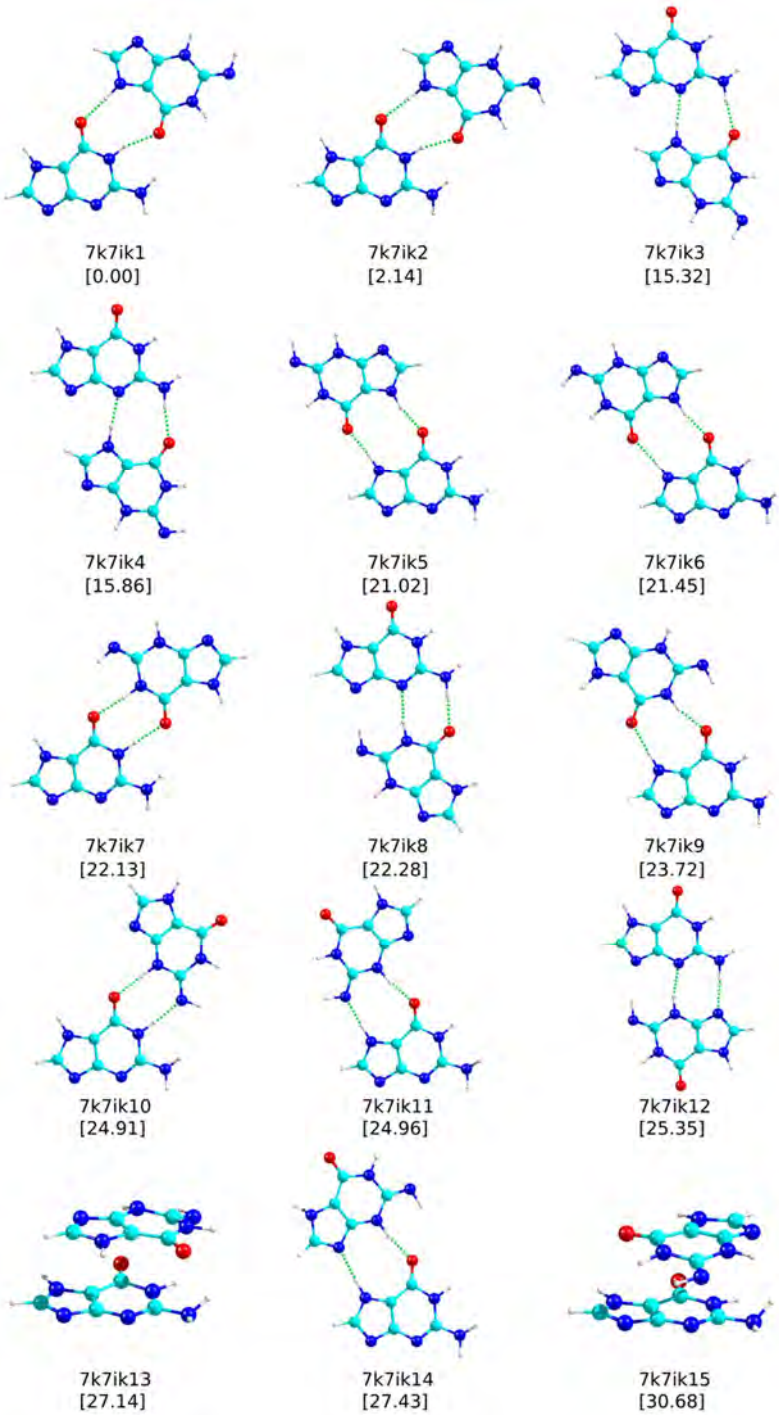


Figure 8.3.41 *The 15 most stable structures of 9e9e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

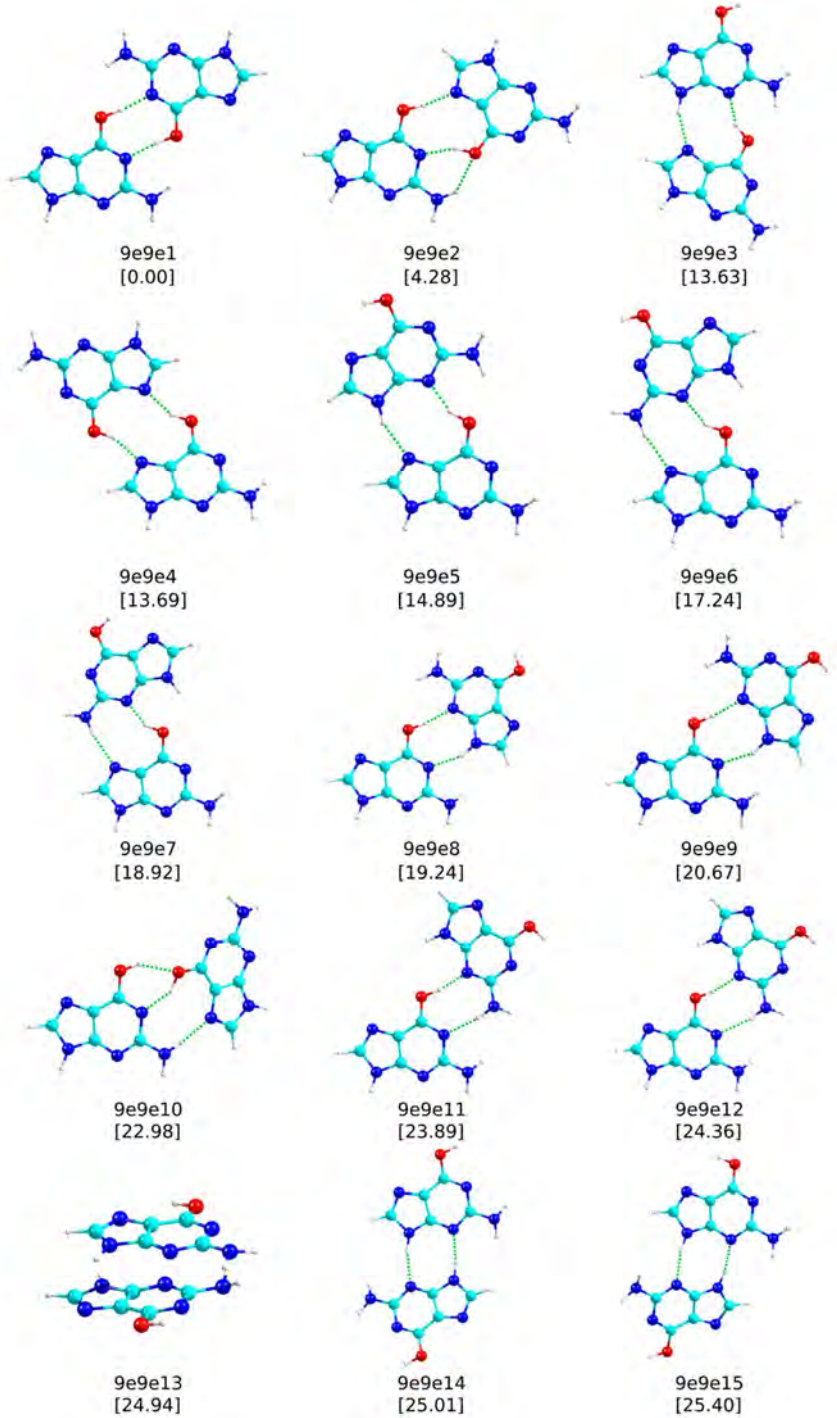


Figure 8.3.42 *The 15 most stable structures of 9e7e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

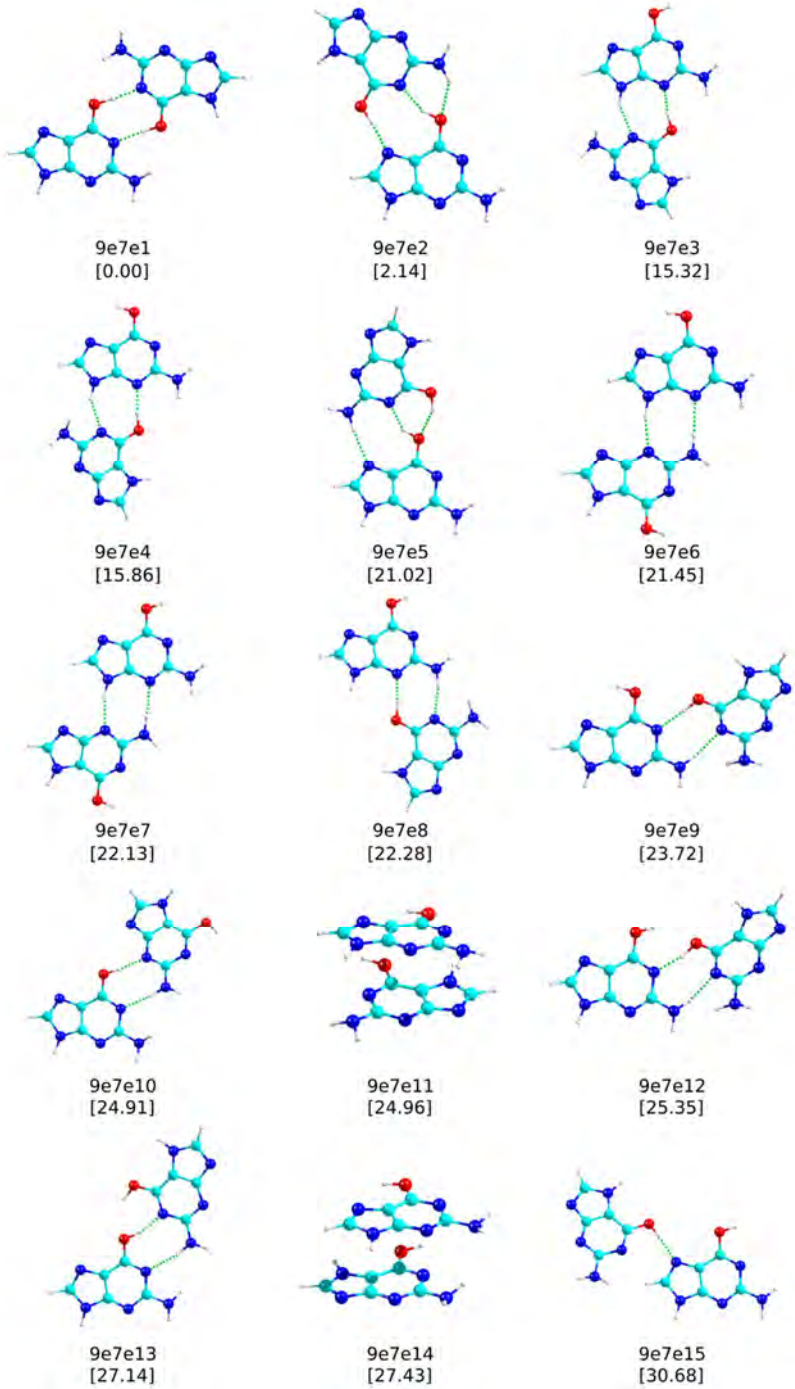


Figure 8.3.43 *The 15 most stable structures of 9e7ik guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

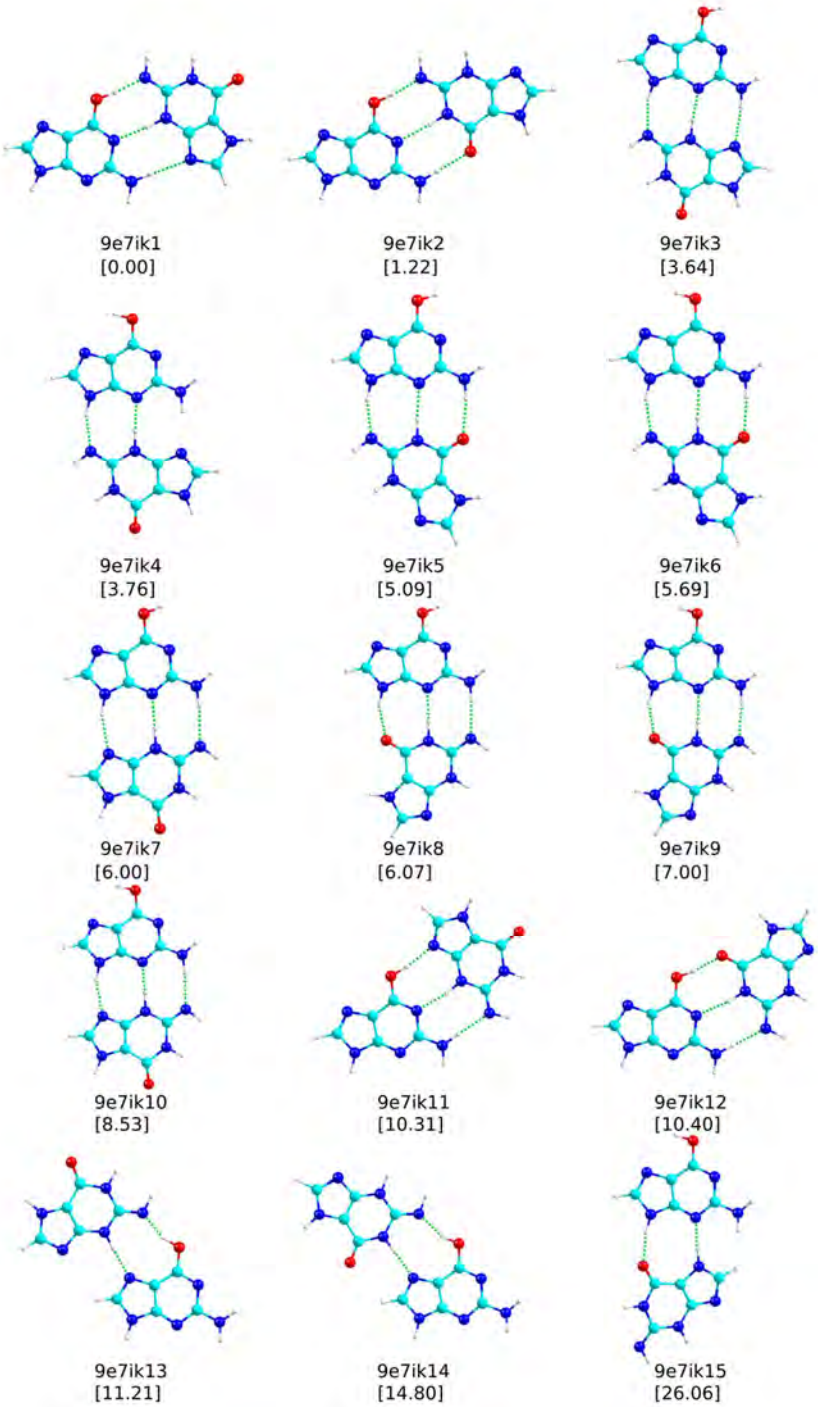


Figure 8.3.44 *The 15 most stable structures of 7e7e guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

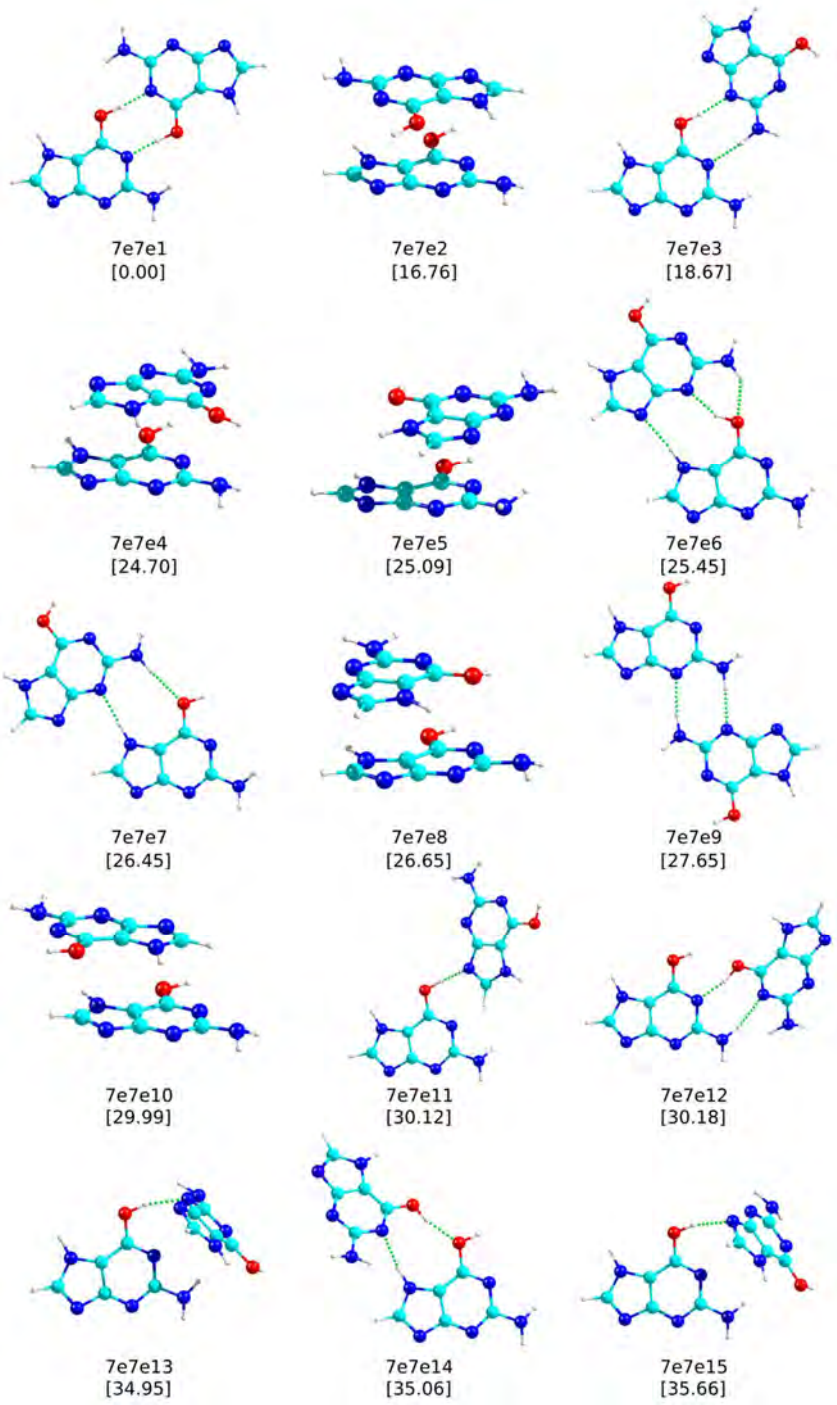


Figure 8.3.45 *The 15 most stable structures of 7e7ik guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

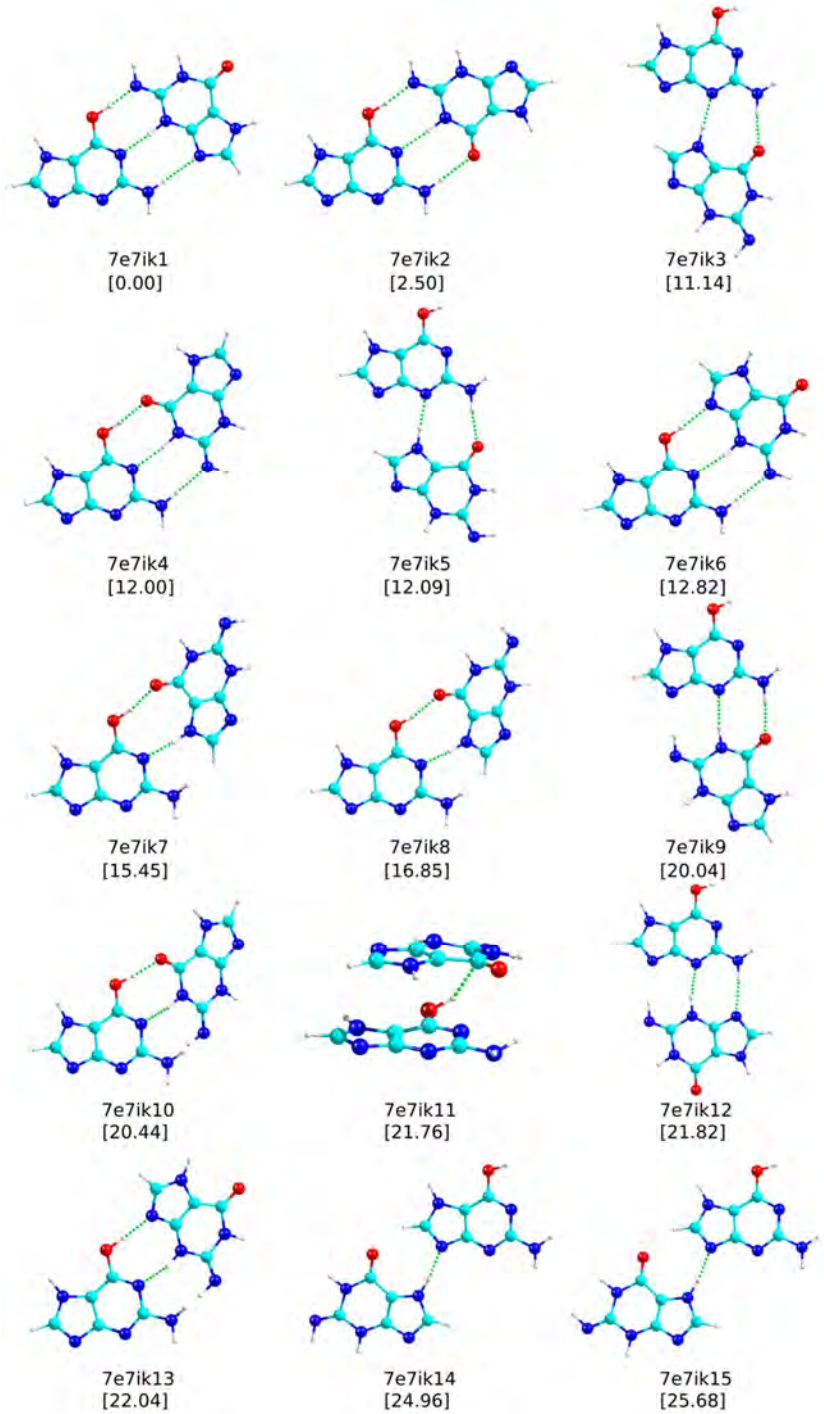


Figure 8.3.46 *The 15 most stable structures of 7ik7ik guanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

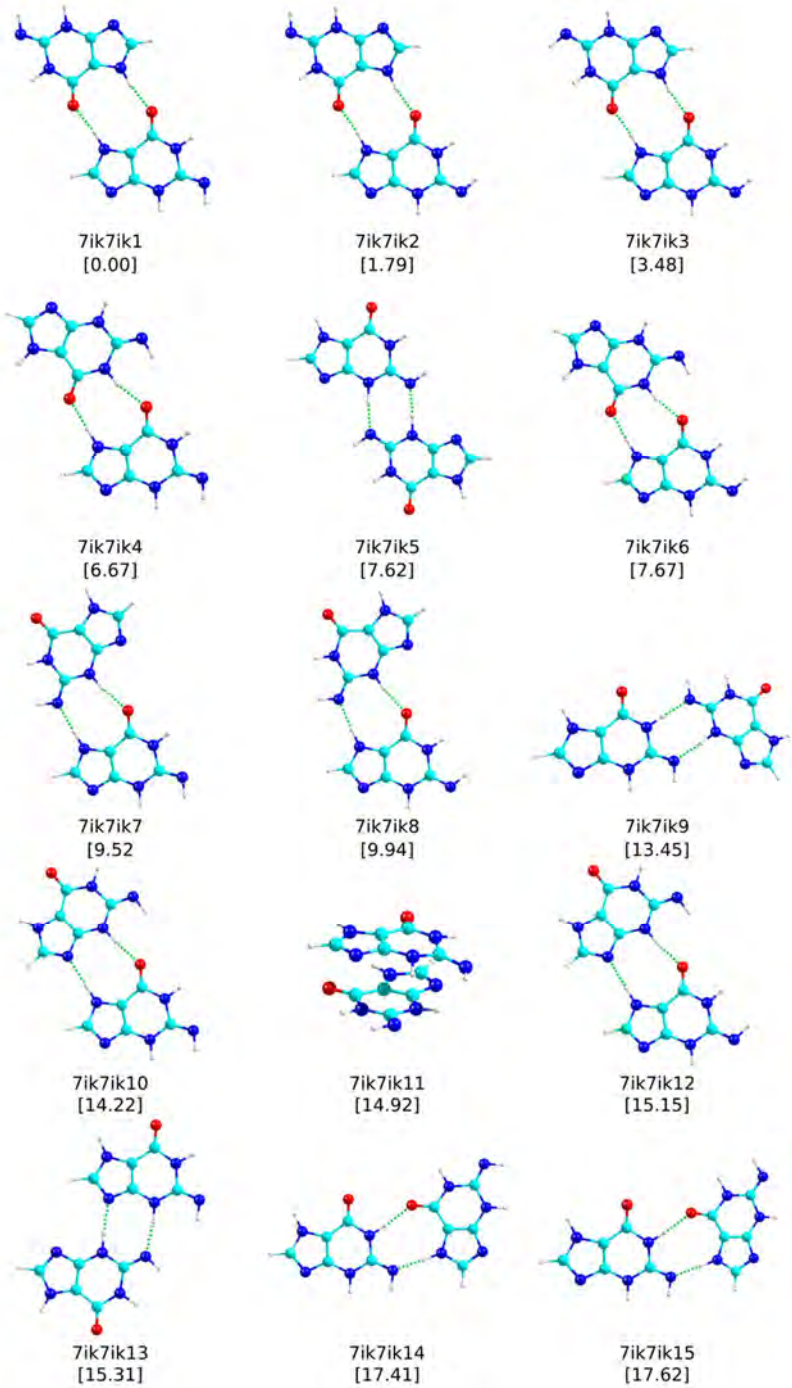


Figure 8.3.47 *Relative Gibbs free energy of the 9k9k guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

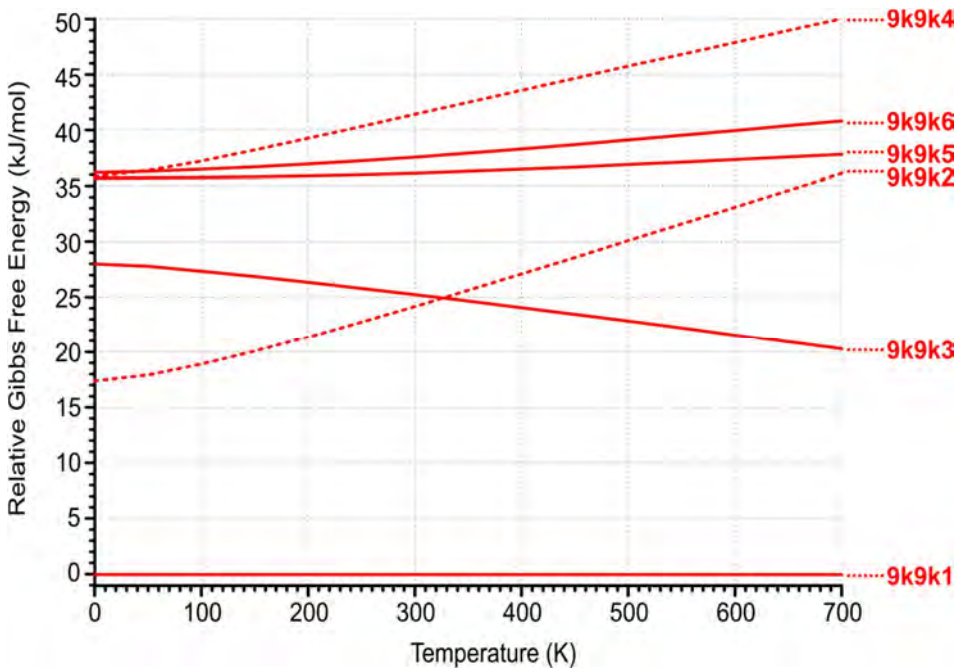


Figure 8.3.48 *Relative Gibbs free energy of the 9k7k guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

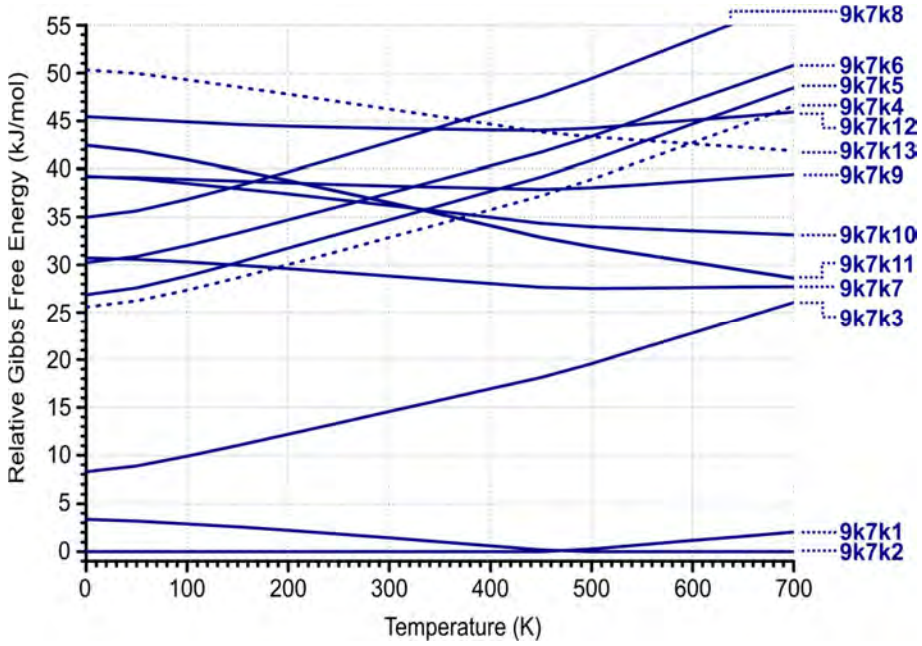


Figure 8.3.49 *Relative Gibbs free energy of the 9k7ik guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

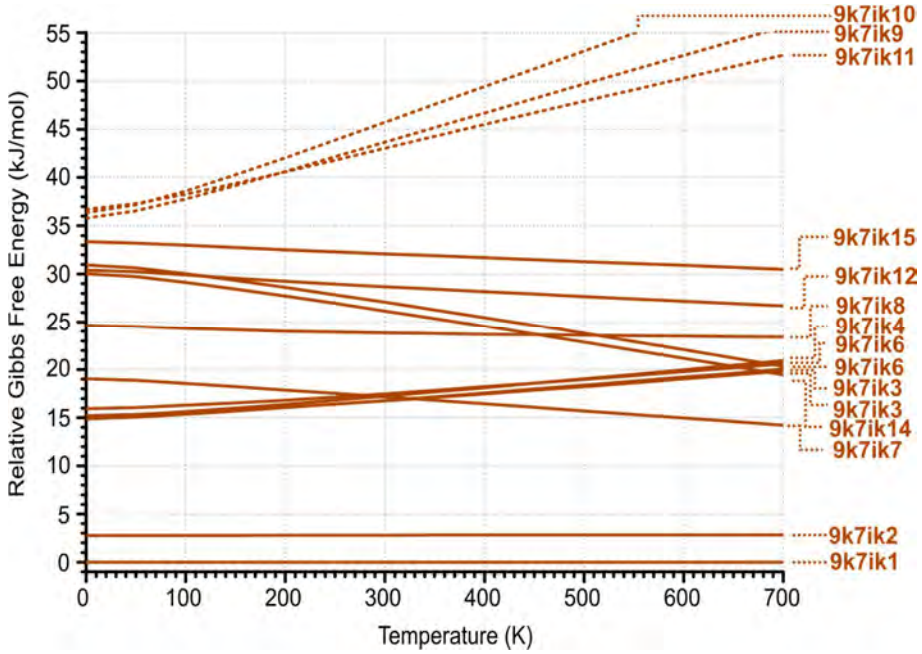


Figure 8.3.50 *Relative Gibbs free energy of the 9k9e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

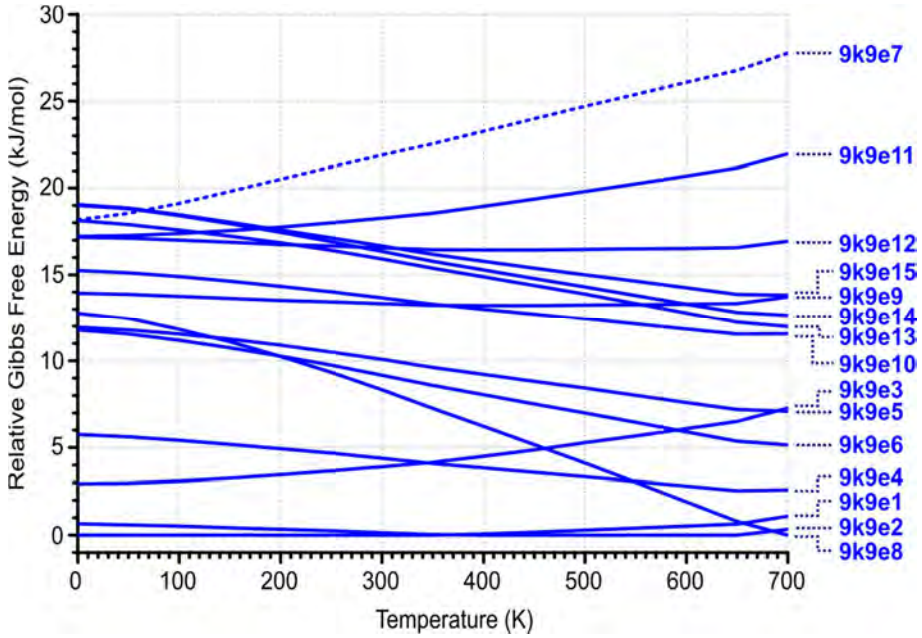


Figure 8.3.51 *Relative Gibbs free energy of the 9k7e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

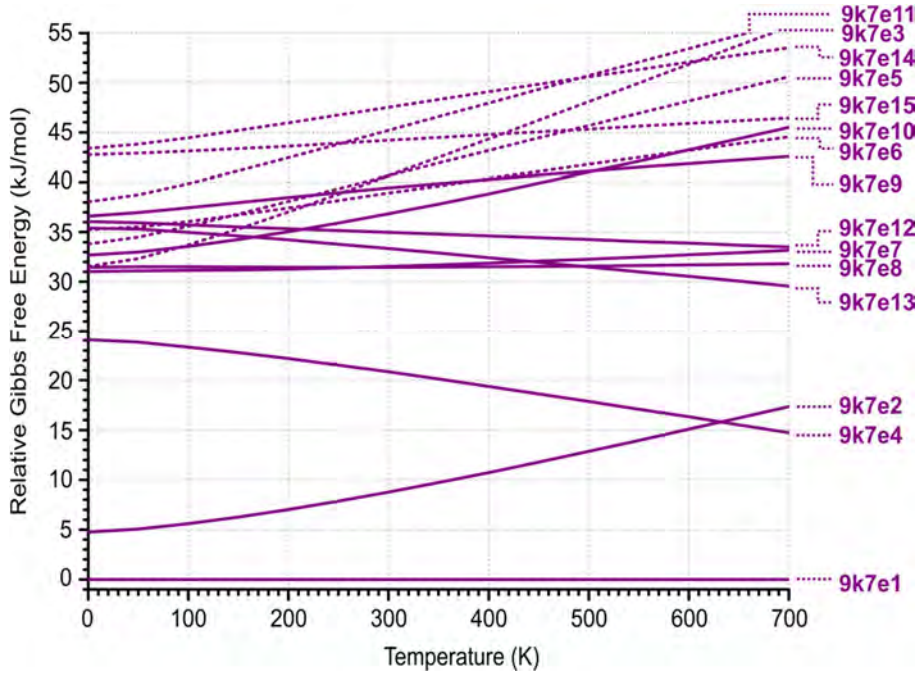


Figure 8.3.52 *Relative Gibbs free energy of the 7k7k guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

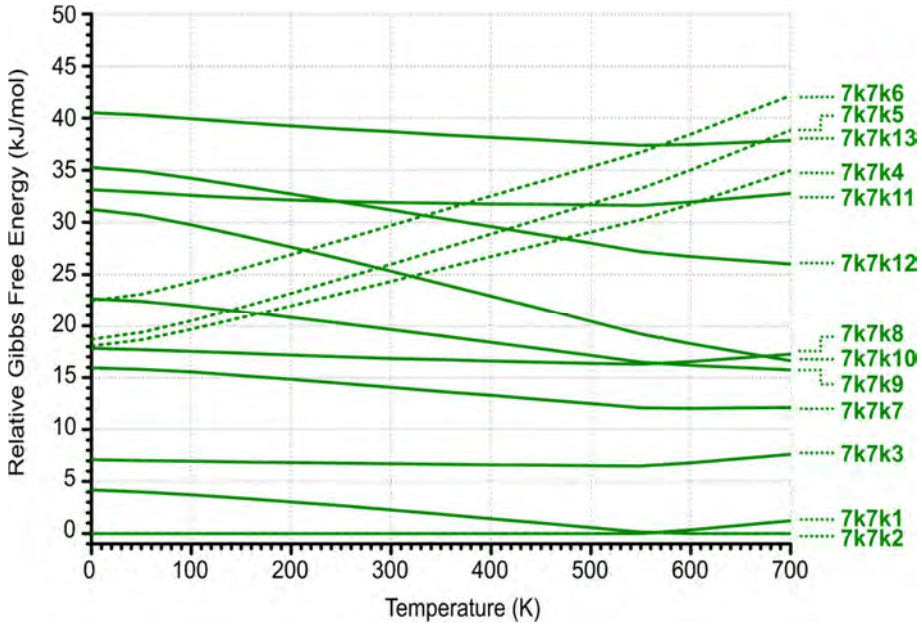


Figure 8.3.53 *Relative Gibbs free energy of the 7k9e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

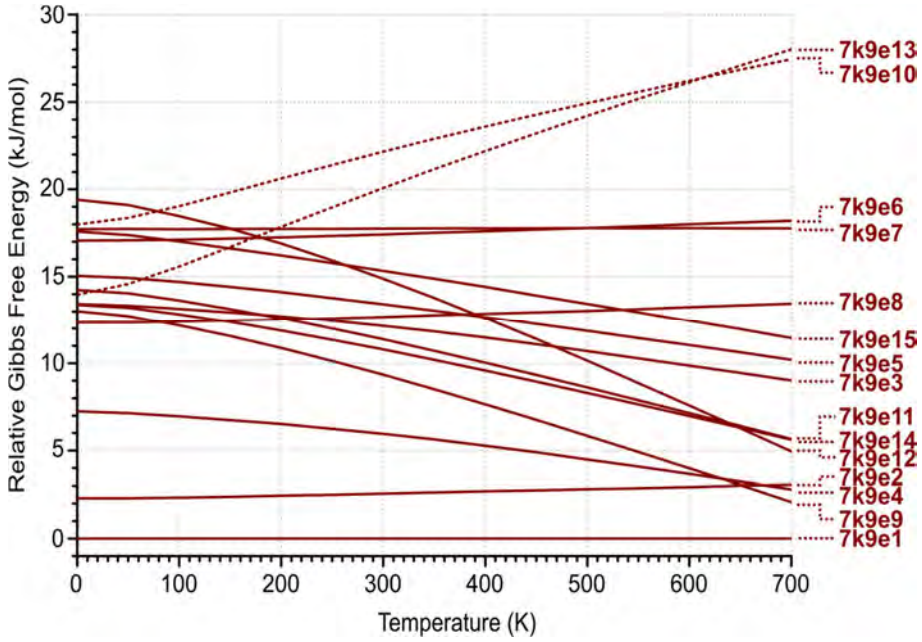


Figure 8.3.54 *Relative Gibbs free energy of the 7k7e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

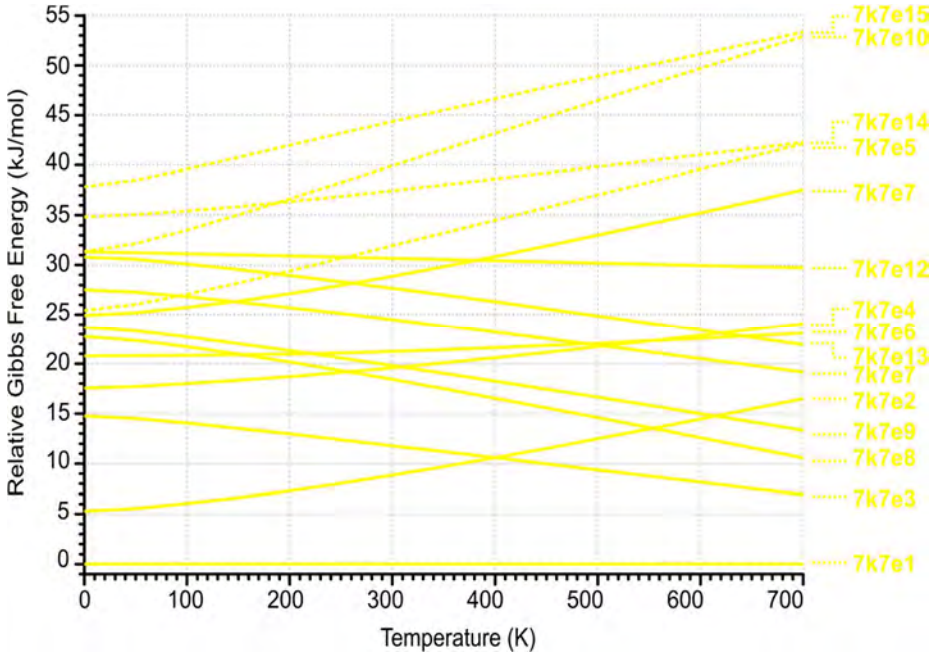


Figure 8.3.55 *Relative Gibbs free energy of the 7k7ik guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

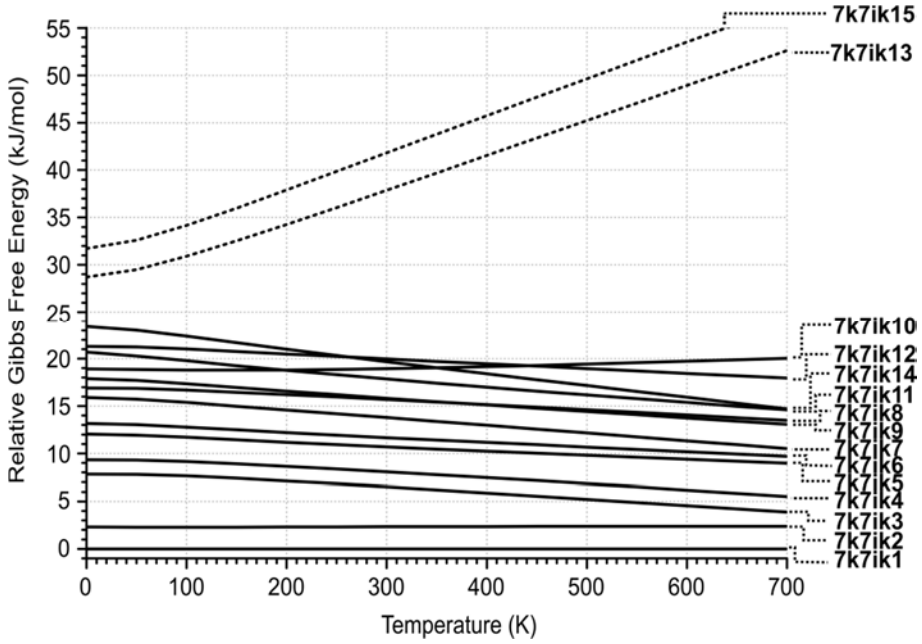


Figure 8.3.56 *Relative Gibbs free energy of the 9e9e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

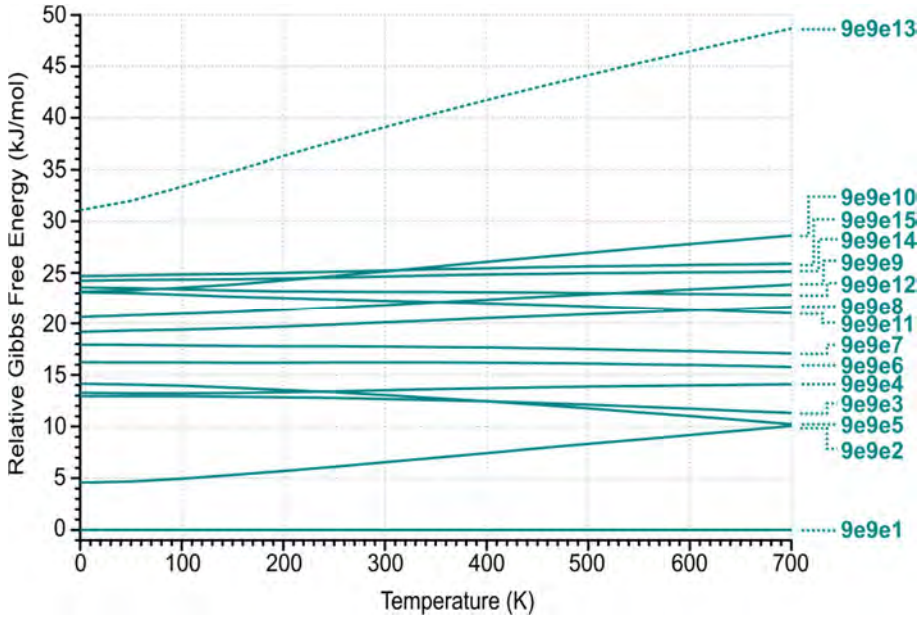


Figure 8.3.57 *Relative Gibbs free energy of the 9e7e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

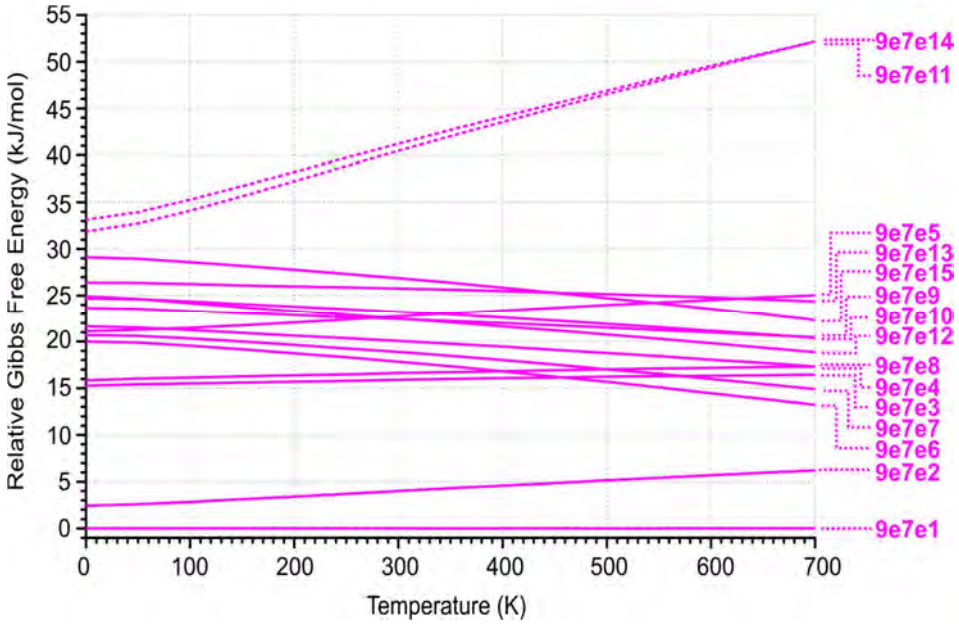


Figure 8.3.58 *Relative Gibbs free energy of the 9e7ik guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

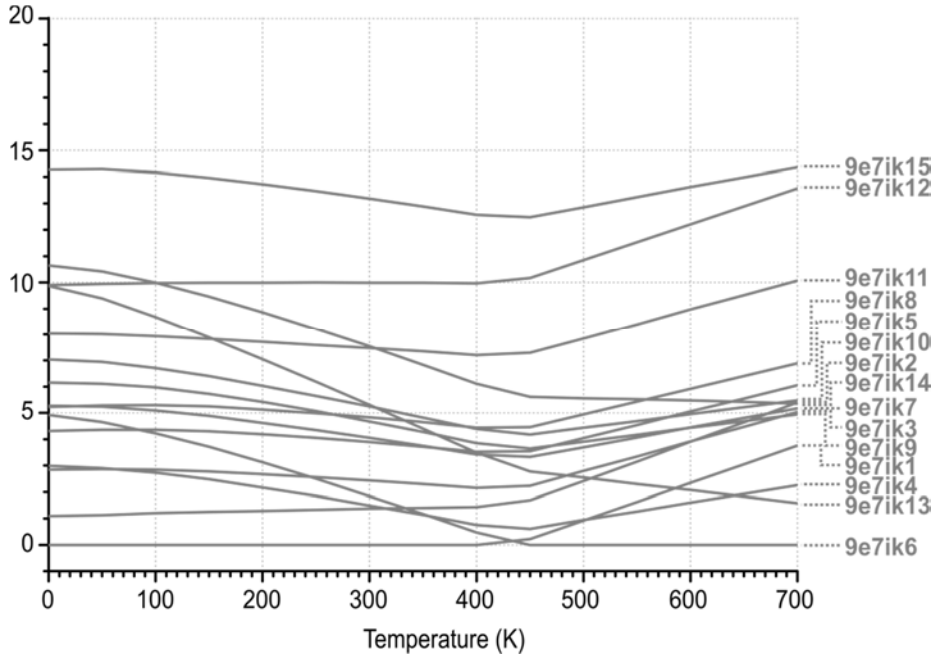


Figure 8.3.59 *Relative Gibbs free energy of the 7e7e guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

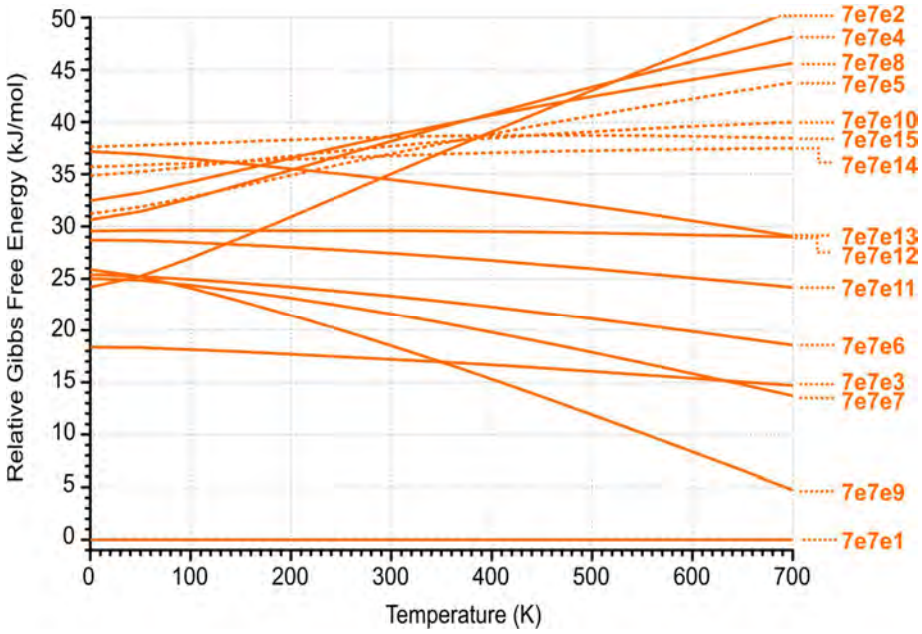


Figure 8.3.60 *Relative Gibbs free energy of the 7e7ik guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

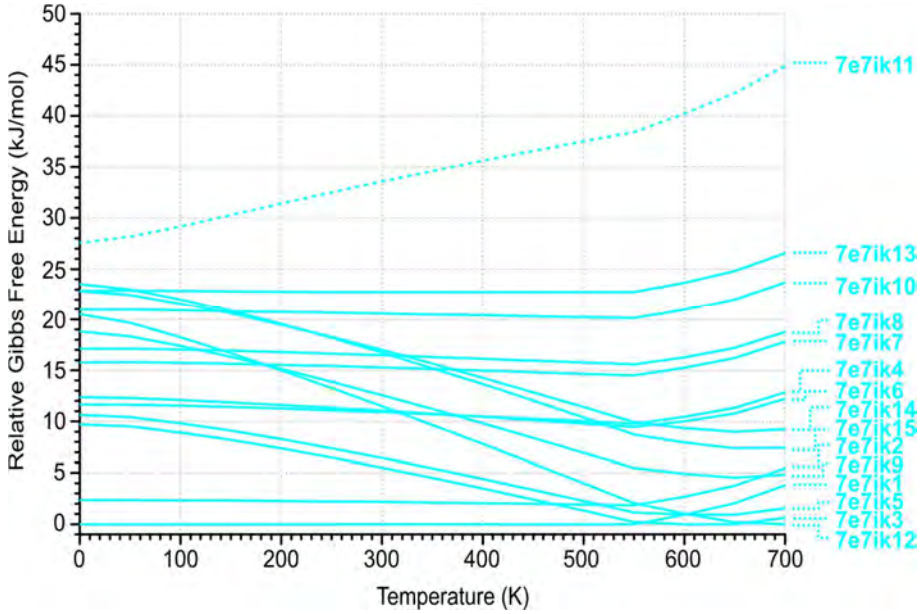


Figure 8.3.61 *Relative Gibbs free energy of the 7ik7ik guanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

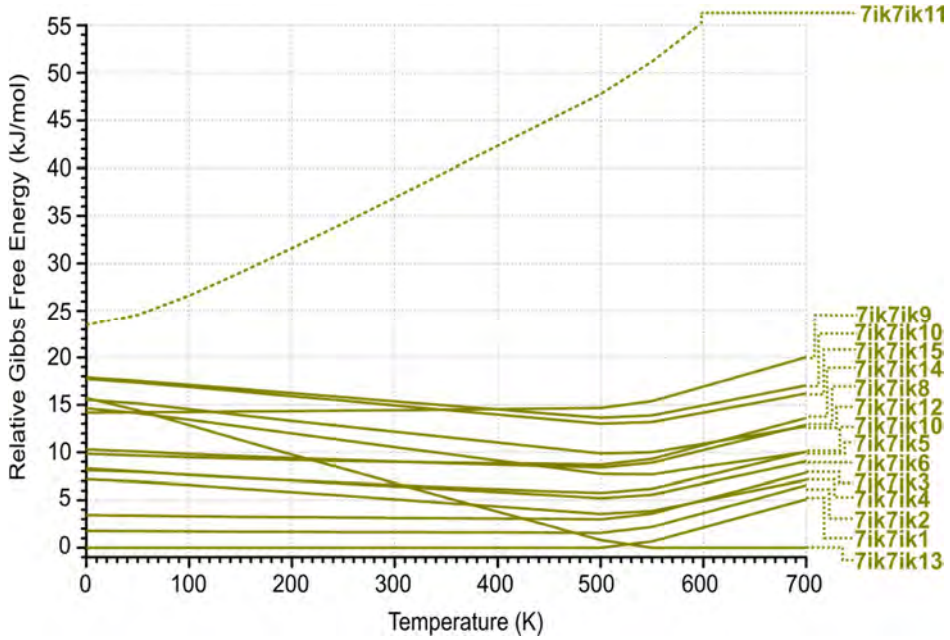


Figure 8.3.62 *2D-NCI plots (s vs. ρ*sign(λ₂)) for the considered dimers of guanine.*

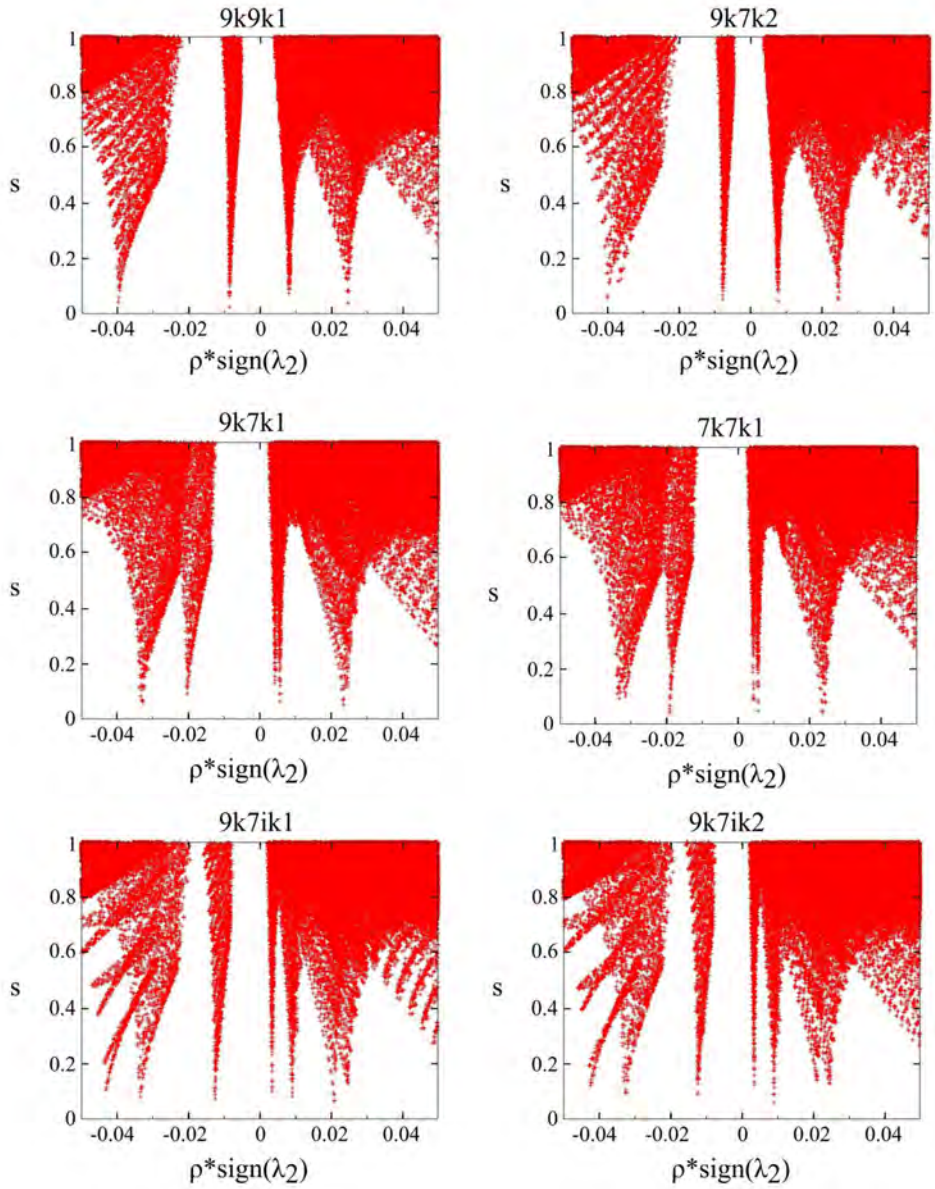


Figure 8.3.63 *IR spectra of dimer of guanine, scanned with the UV wavelenght fixed at 33108 cm^{-1}(blue trace), 33734 cm^{-1} and 34063 cm^{-1}(blak trace). The positions of the band has been jutted out with different color bars.*

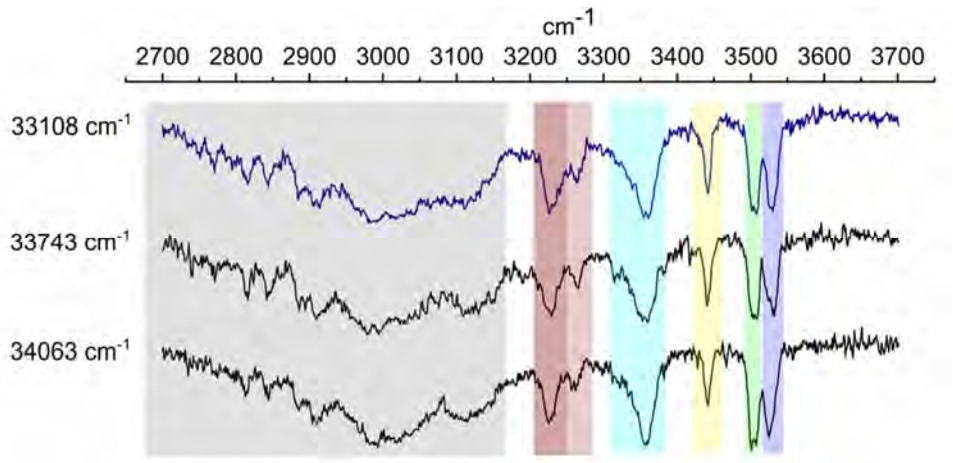
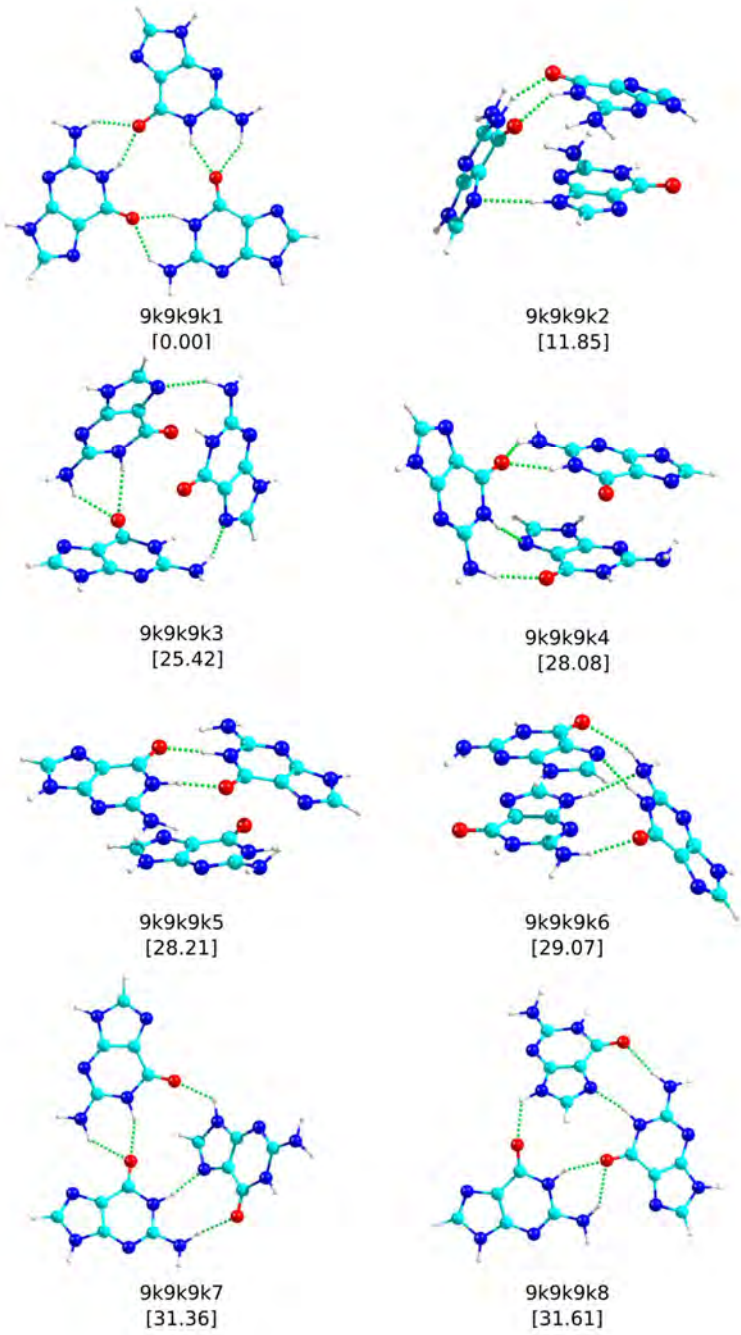
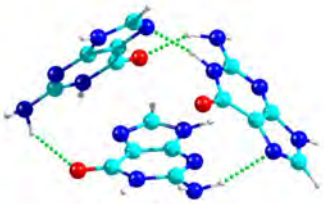
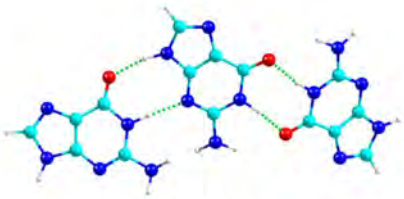


Figure 8.3.64 *The 16 most stable structures of 9k9k9k guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

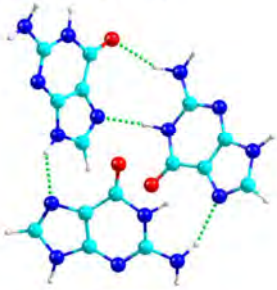




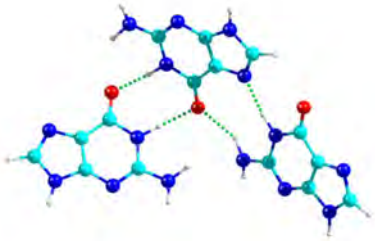
9k9k9k9
[32.99]



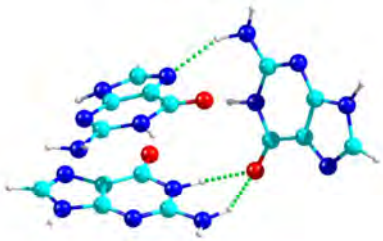
9k9k9k10
[39.45]



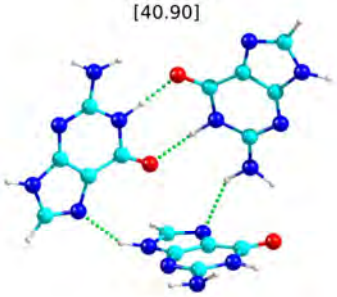
9k9k9k11
[39.92]



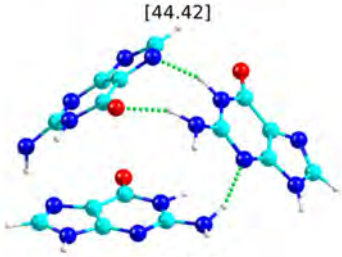
9k9k9k12
[40.90]



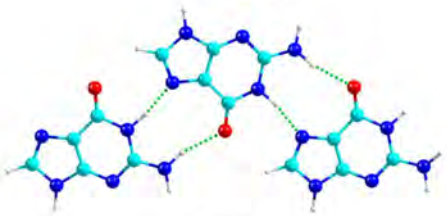
9k9k9k13
[44.42]



9k9k9k14
[45.67]

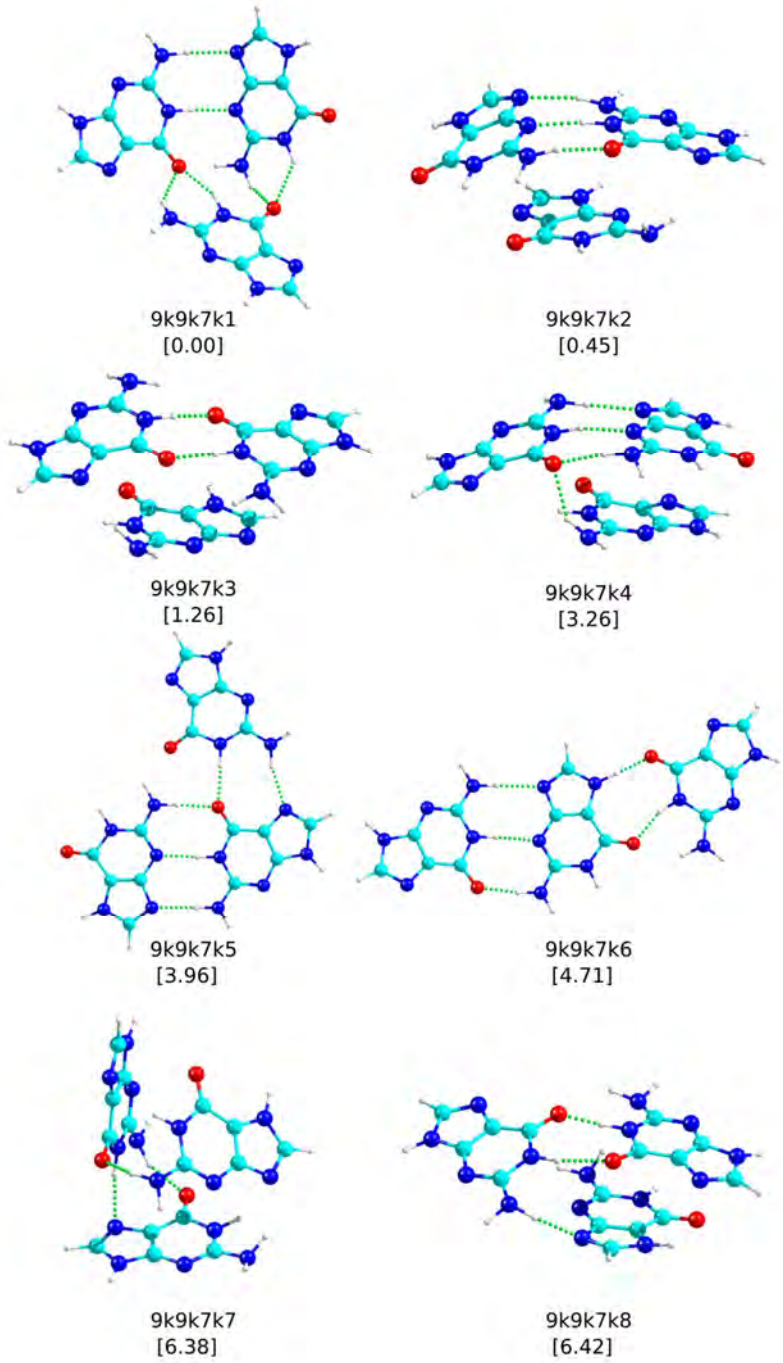


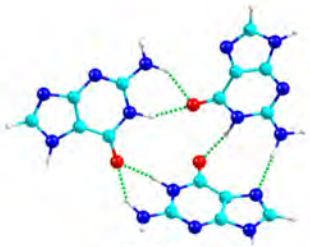
9k9k9k15
[45.93]



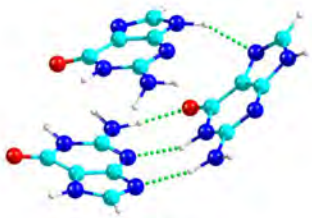
9k9k9k16
[59.34]

Figure 8.3.65 *The 15 most stable structures of 9k9k7k guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

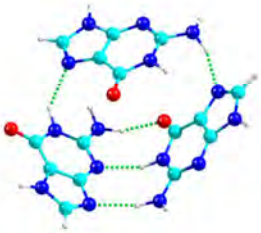




9k9k7k9
[7.62]

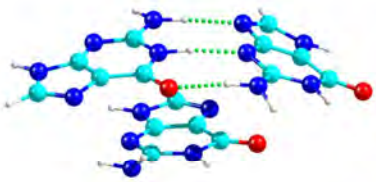


9k9k7k10
[7.94]

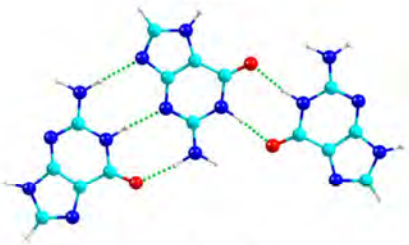


9k9k7k11
[8.49]

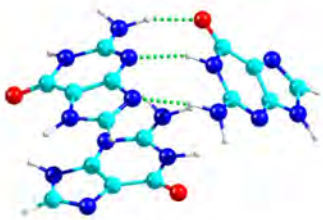
9k9k7k12
[9.35]



9k9k7k13
[9.50]

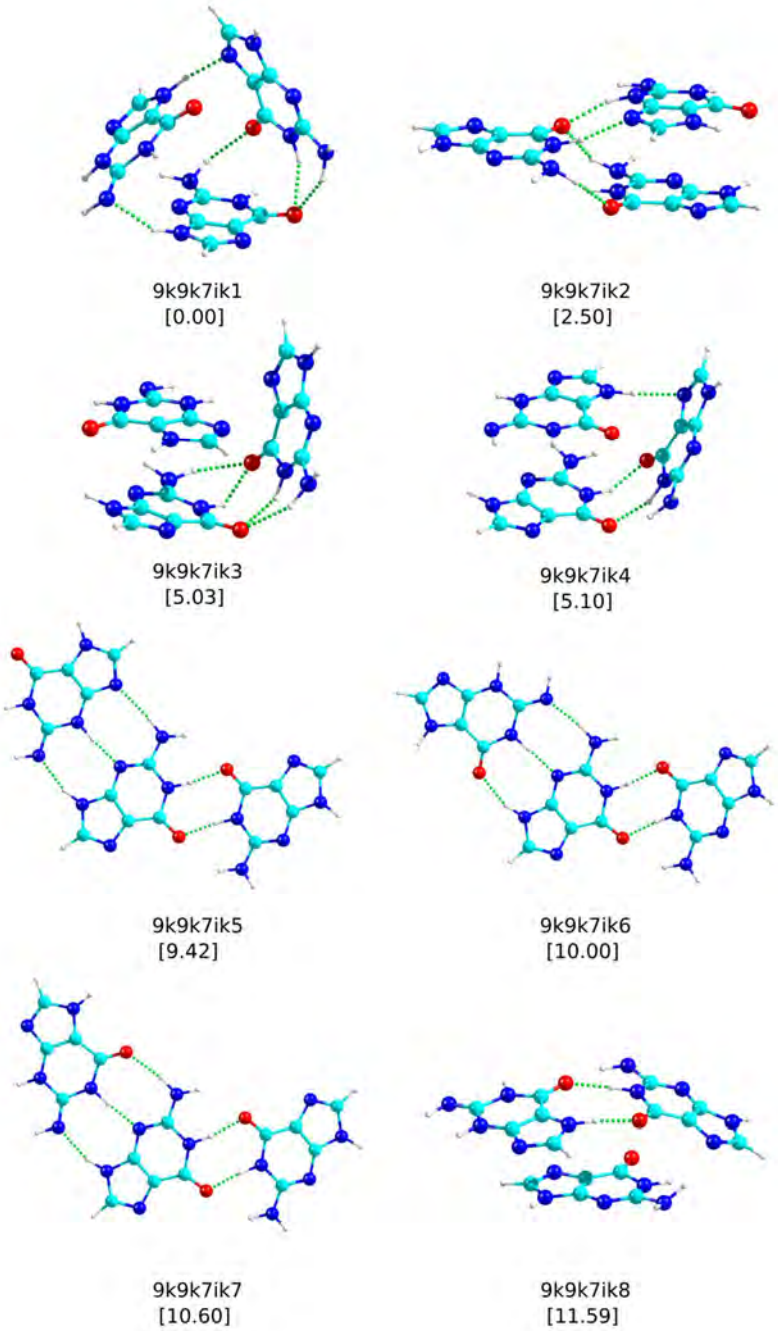


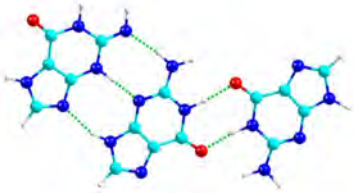
9k9k7k14
[9.92]



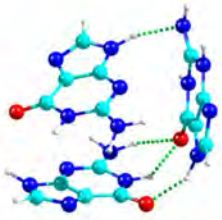
9k9k7k15
[10.98]

Figure 8.3.66 *The 15 most stable structures of 9k9k7ik guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

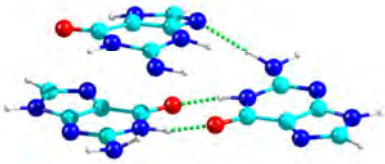




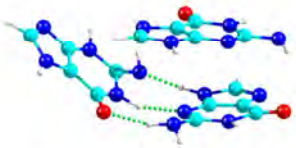
9k9k7ik9
[11.63]



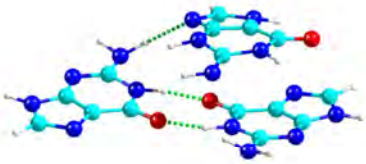
9k9k7ik10
[11.77]



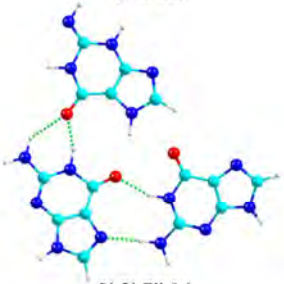
9k9k7ik11
[12.93]



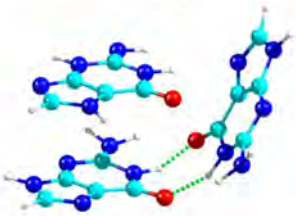
9k9k7ik12
[13.19]



9k9k7ik13
[14.03]

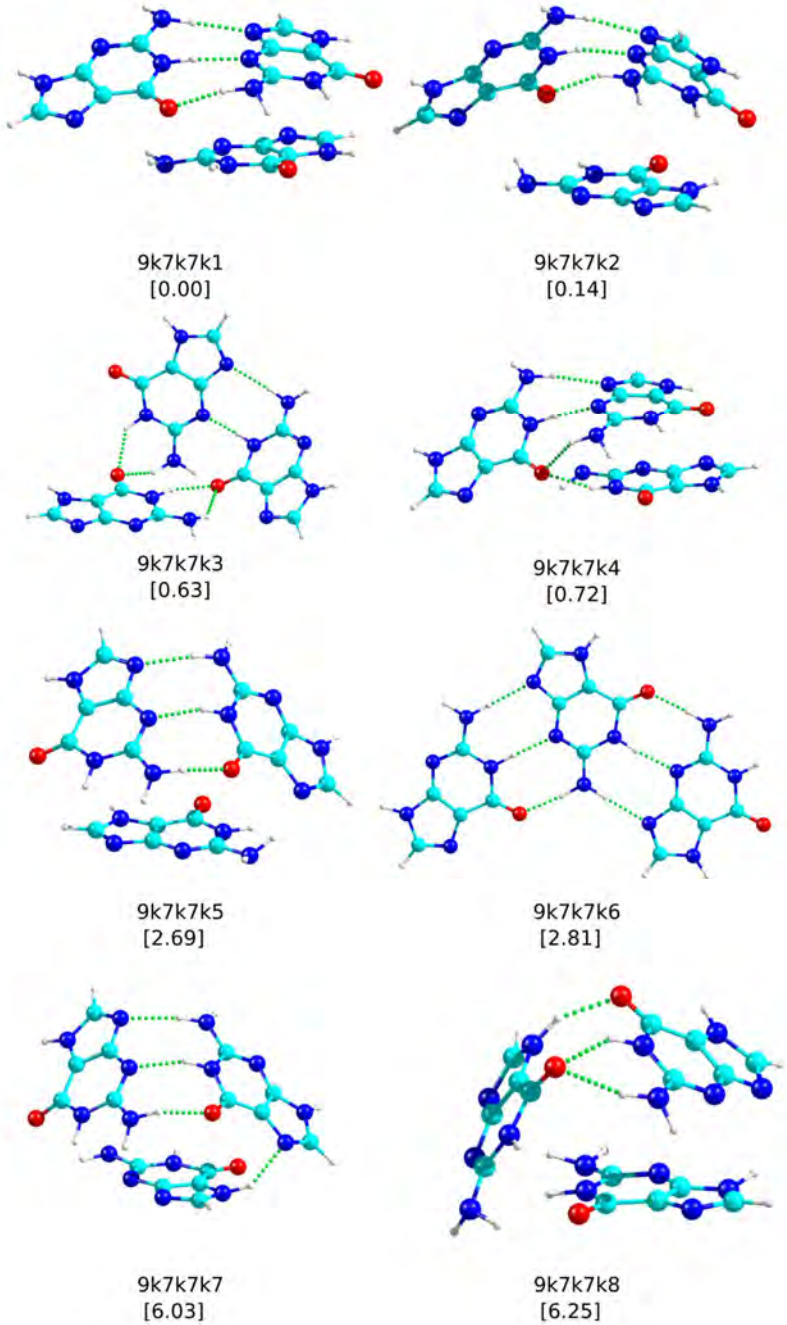


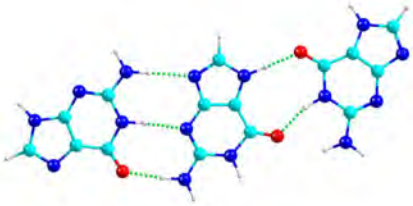
9k9k7ik14
[16.28]



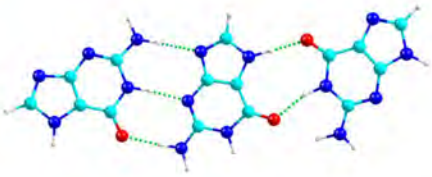
9k9k7ik15
[17.42]

Figure 8.3.67 *The 15 most stable structures of 9k7k7k guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

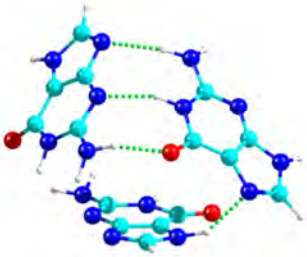




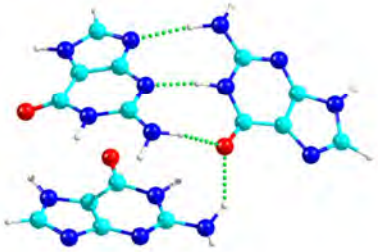
9k7k7k9
[6.44]



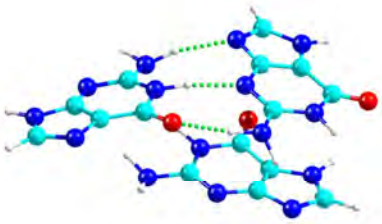
9k7k7k10
[7.03]



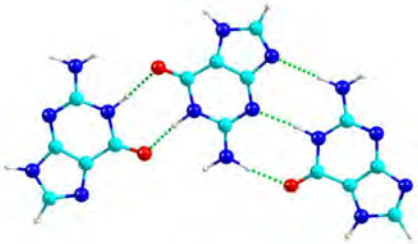
9k7k7k11
[7.05]



9k7k7k12
[7.92]



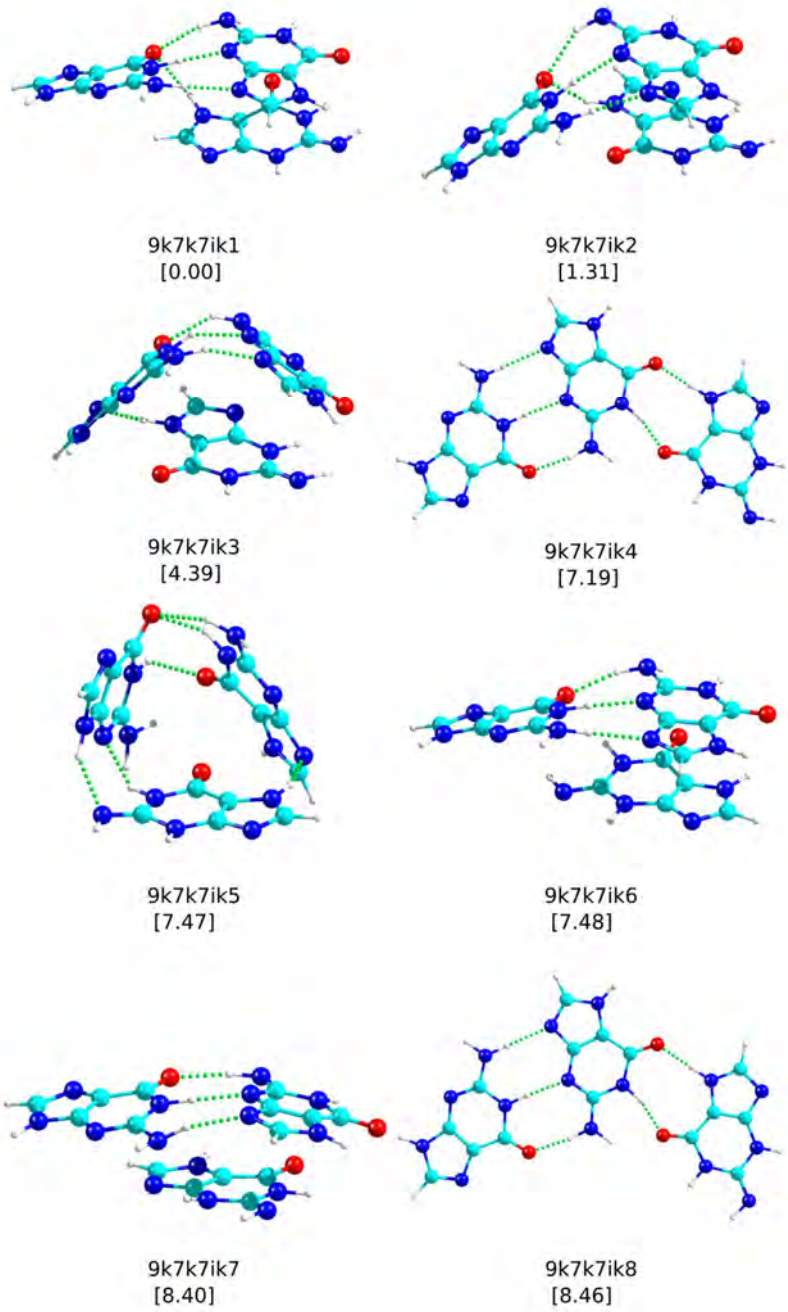
9k7k7k13
[8.10]

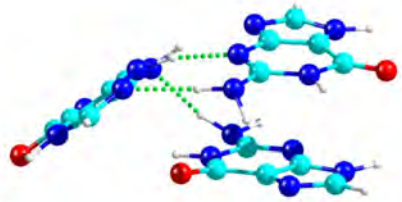


9k7k7k14
[9.06]

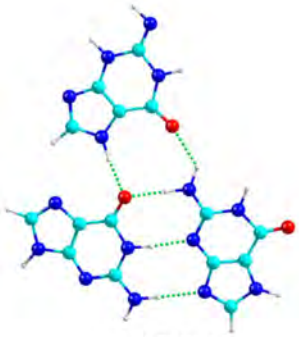
9k7k7k15
[9.95]

Figure 8.3.68 *The 15 most stable structures of 9k7k7ik guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

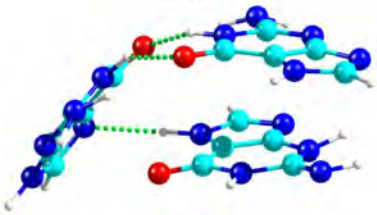




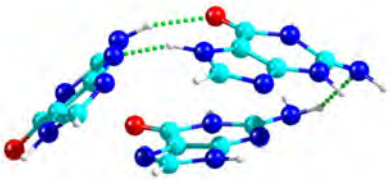
9k7k7ik9
[9.54]



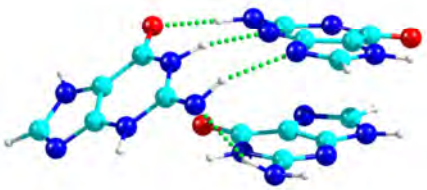
9k7k7ik10
[9.61]



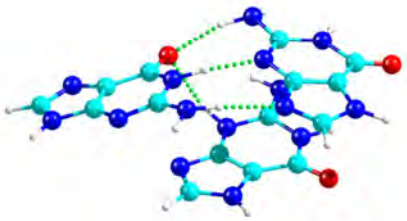
9k7k7ik11
[10.06]



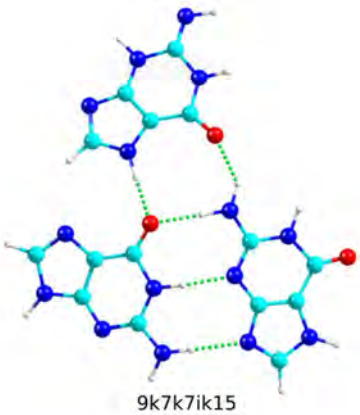
9k7k7ik12
[10.69]



9k7k7ik13
[10.71]

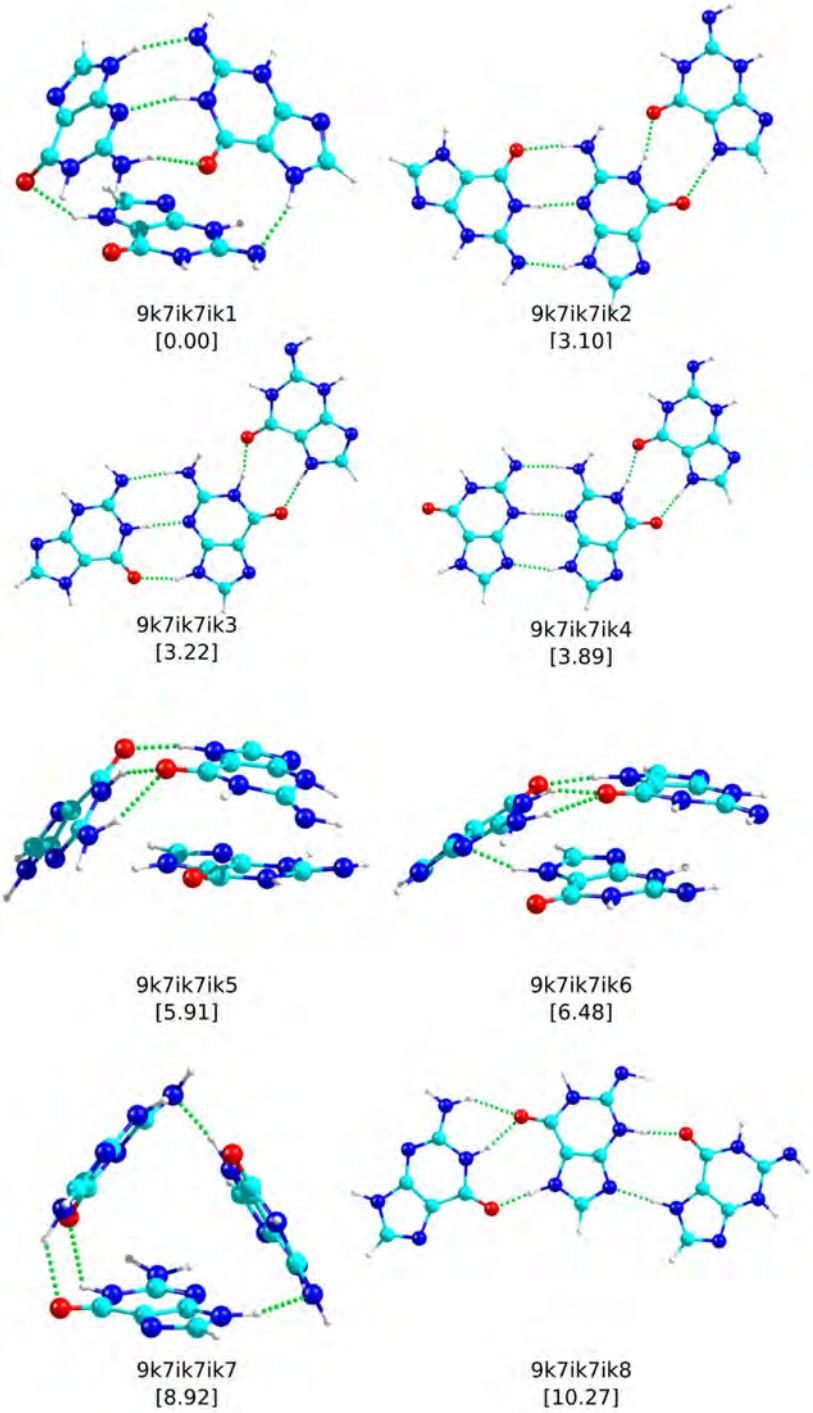


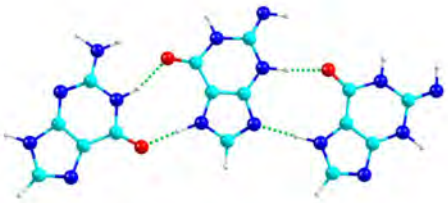
9k7k7ik14
[11.08]



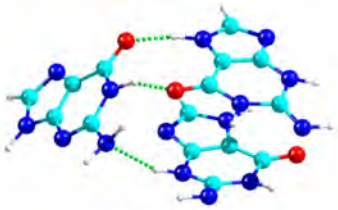
9k7k7ik15
[11.97]

Figure 8.3.69 *The 15 most stable structures of 9k7ik7ik guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

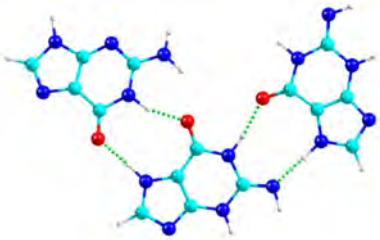




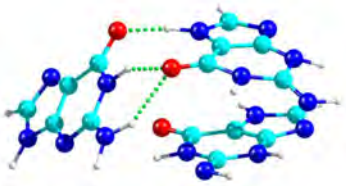
9k7ik7ik9
[11.13]



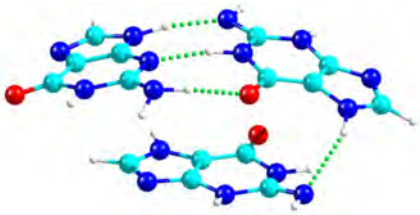
9k7ik7ik10
[11.32]



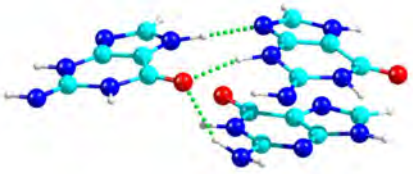
9k7ik7ik11
[13.50]



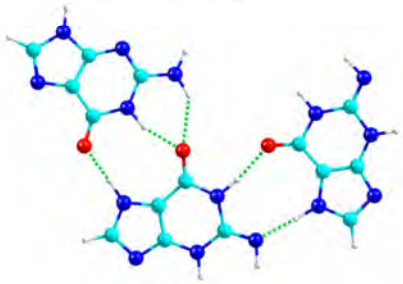
9k7ik7ik12
[14.09]



9k7ik7ik13
[15.46]

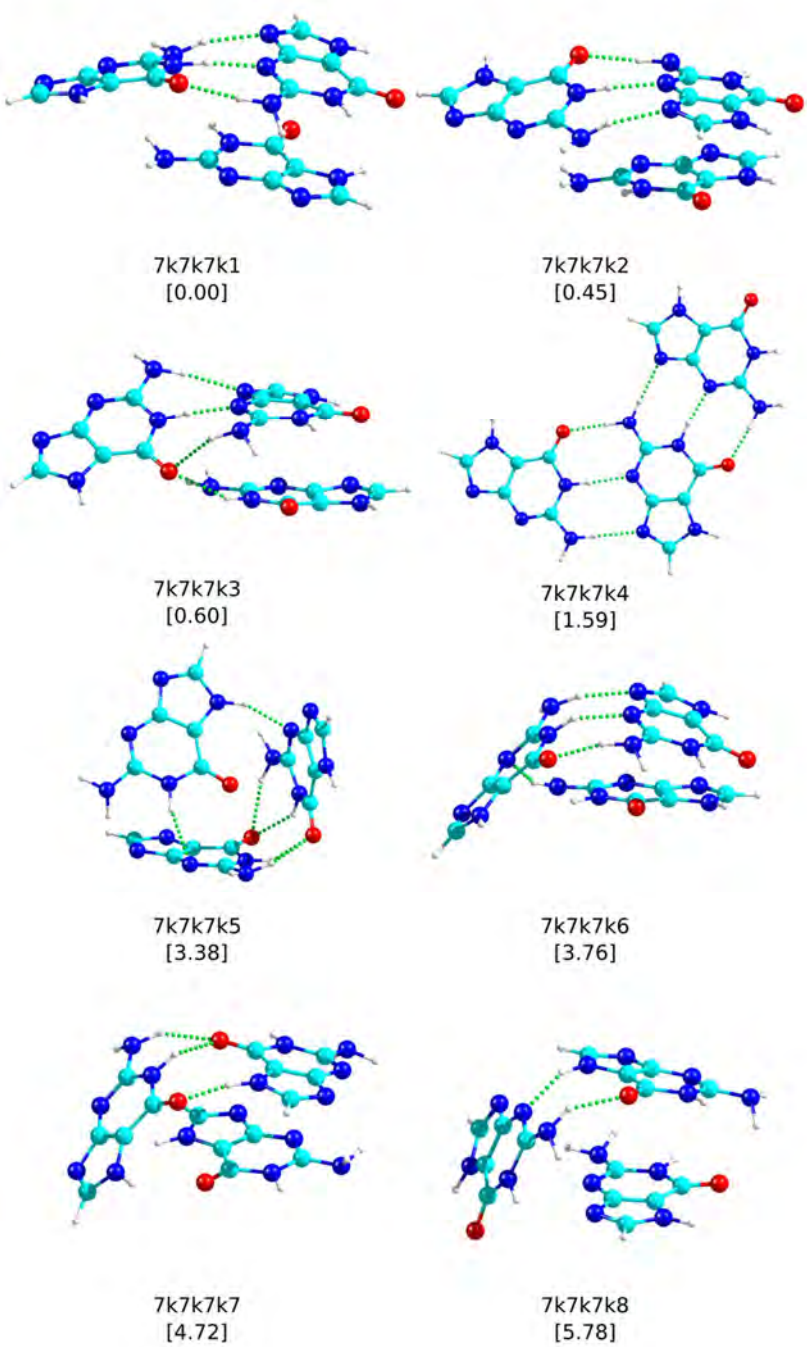


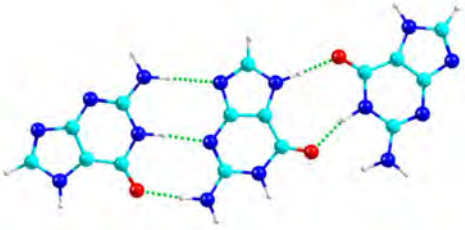
9k7ik7ik14
[15.89]



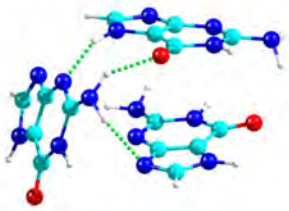
9k7ik7ik15
[15.92]

Figure 8.3.70 *The 15 most stable structures of 7k7k7k guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

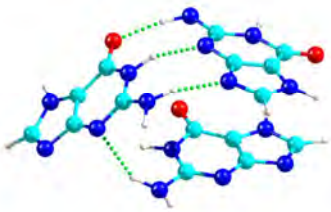




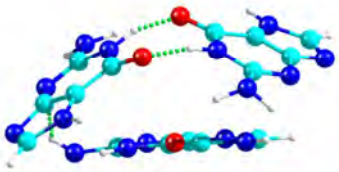
7k7k7k9
[5.92]



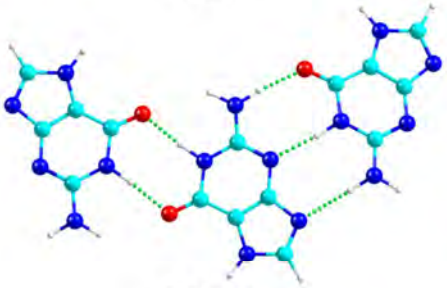
7k7k7k10
[5.99]



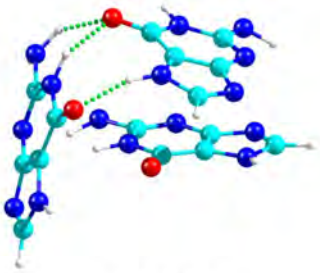
7k7k7k11
[6.98]



7k7k7k12
[7.37]



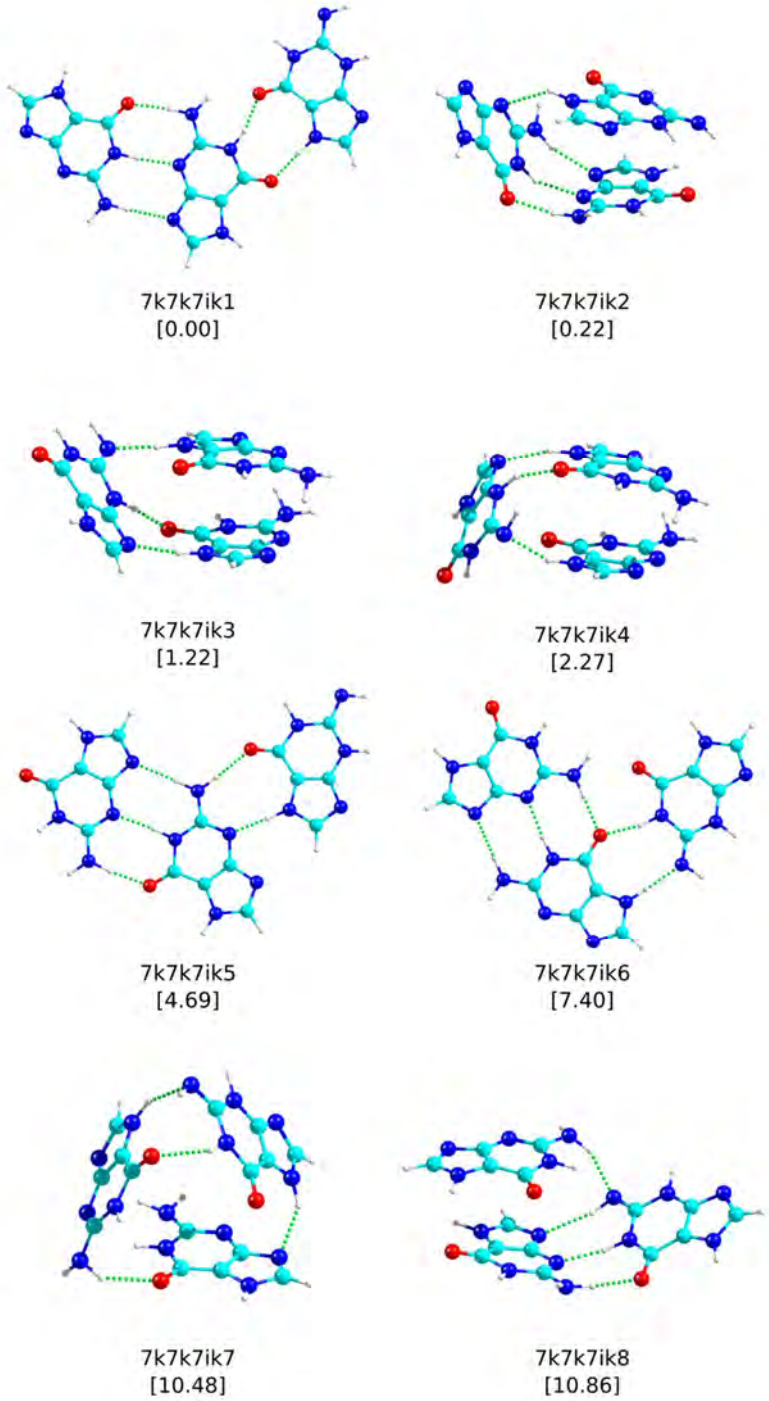
7k7k7k13
[7.84]

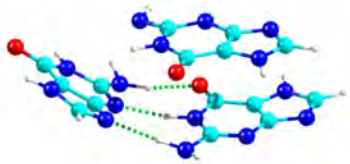


7k7k7k14
[8.33]

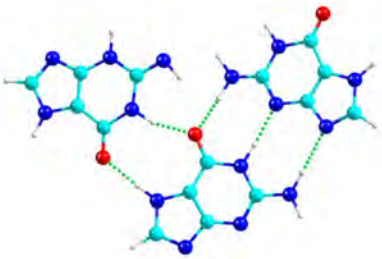
7k7k7k15
[8.63]

Figure 8.3.71 *The 15 most stable structures of 7k7k7ik guanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

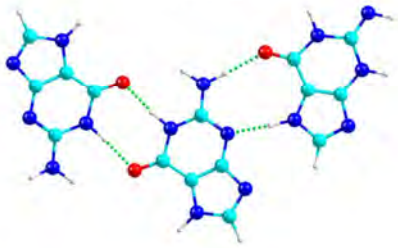




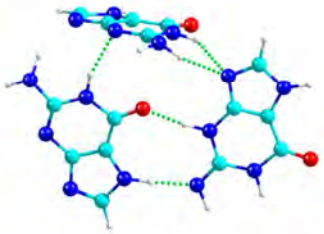
7k7k7ik9
[10.99]



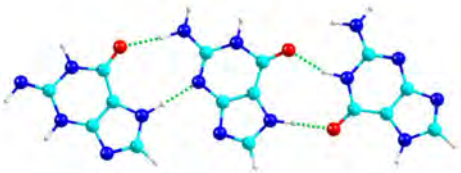
7k7k7ik10
[11.30]



7k7k7ik11
[11.57]

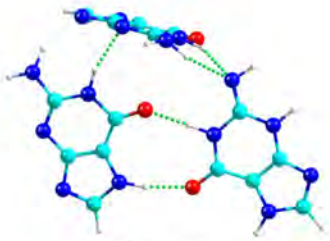


7k7k7ik12
[11.94]



7k7k7ik13
[12.12]

7k7k7ik14
[12.89]



7k7k7ik15
[13.18]

Figure 8.3.72 *Relative Gibbs free energy of the 9k9k9k guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

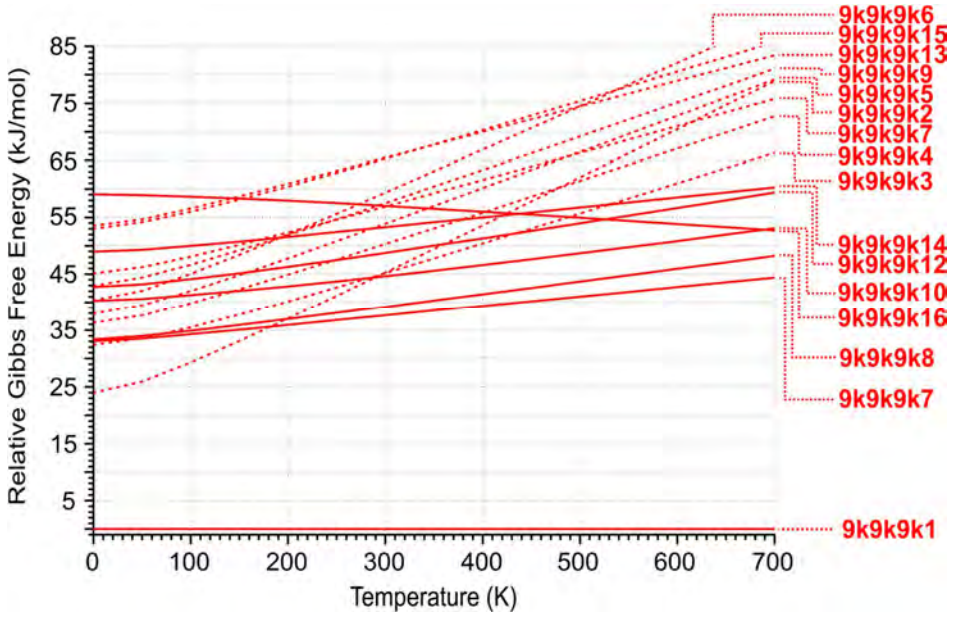


Figure 8.3.73 *Relative Gibbs free energy of the 9k9k7k guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

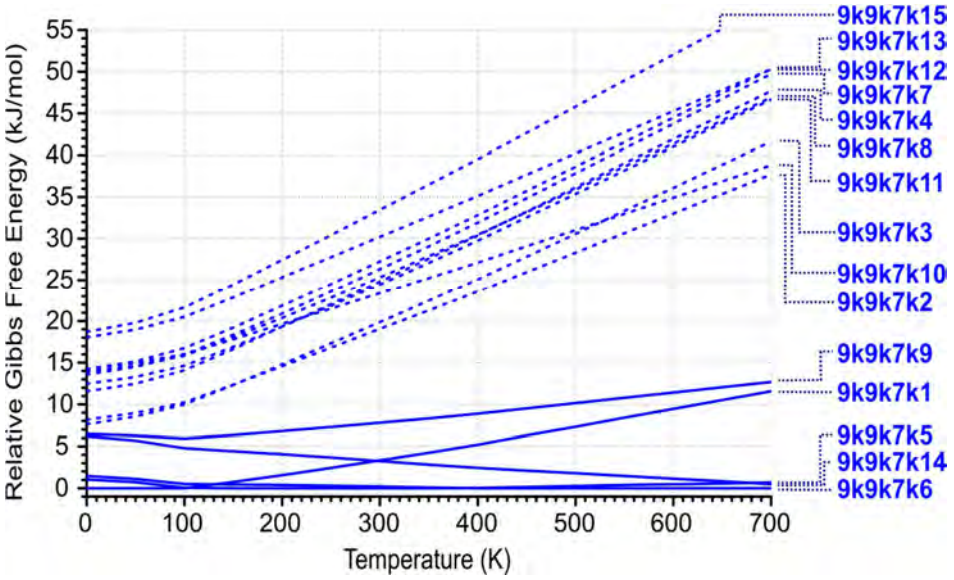


Figure 8.3.74 *Relative Gibbs free energy of the 9k9k7ik guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

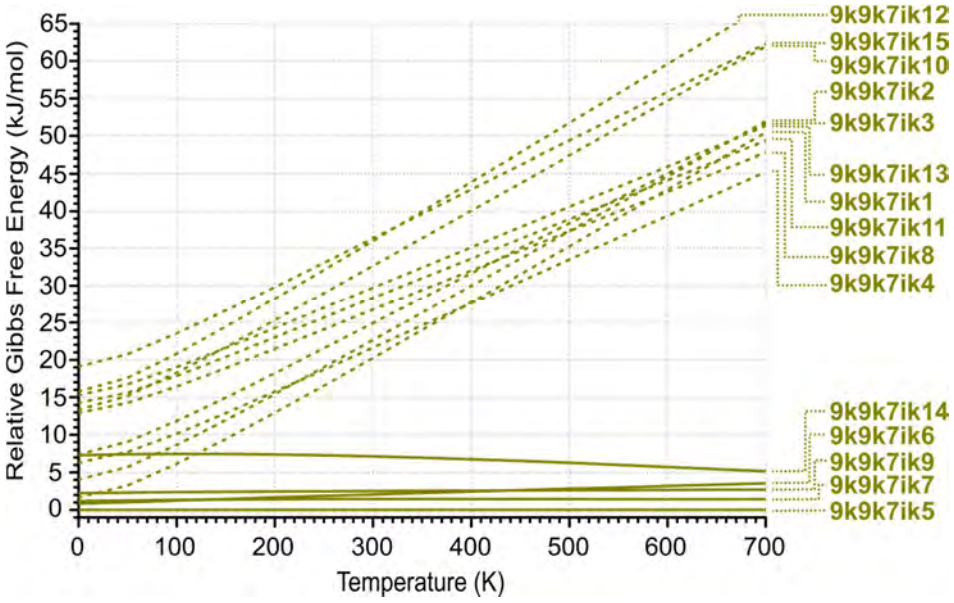


Figure 8.3.75 *Relative Gibbs free energy of the 9k7k7k guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

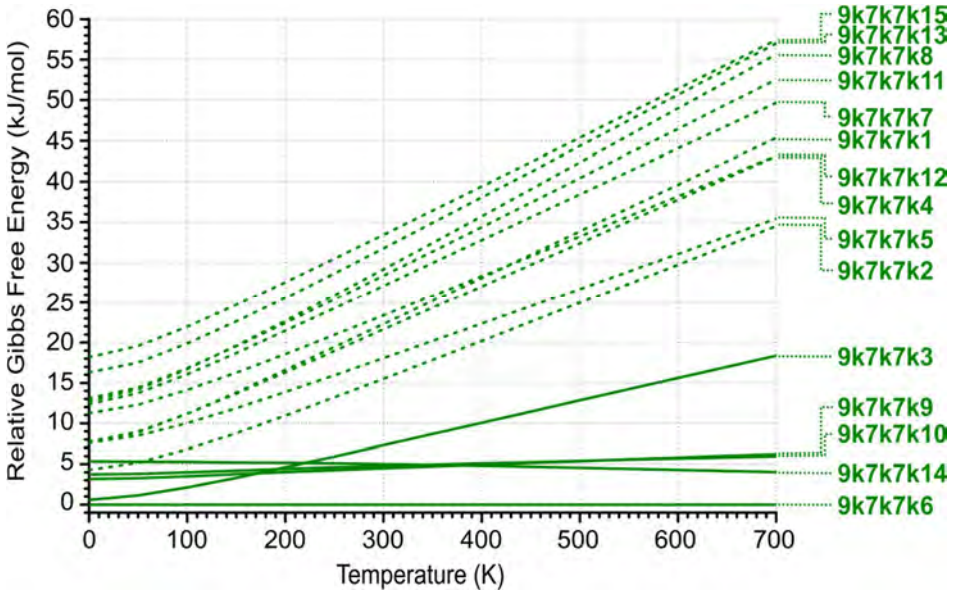


Figure 8.3.76 *Relative Gibbs free energy of the 9k7k7ik guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

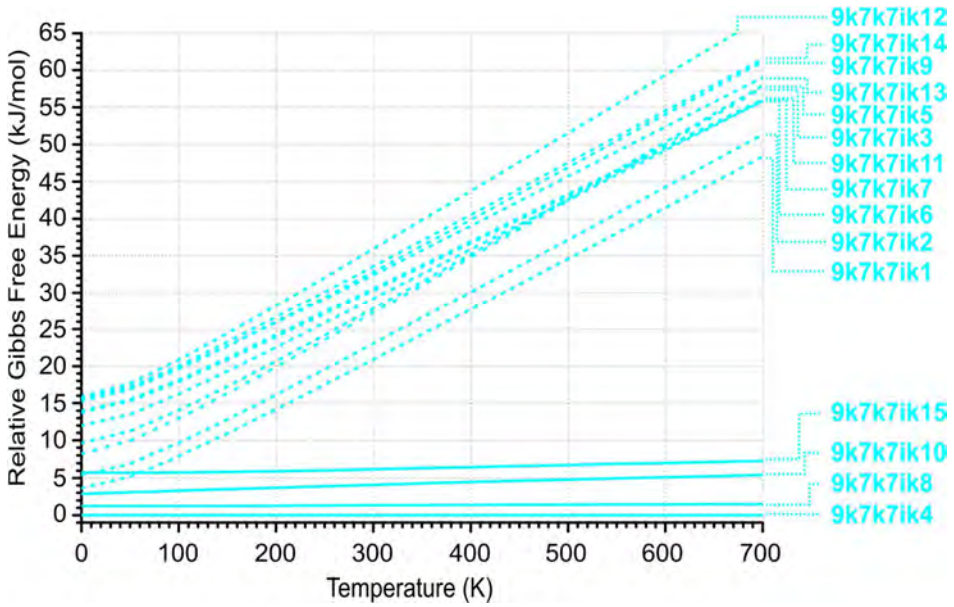


Figure 8.3.77 *Relative Gibbs free energy of the 9k7ik7ik guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

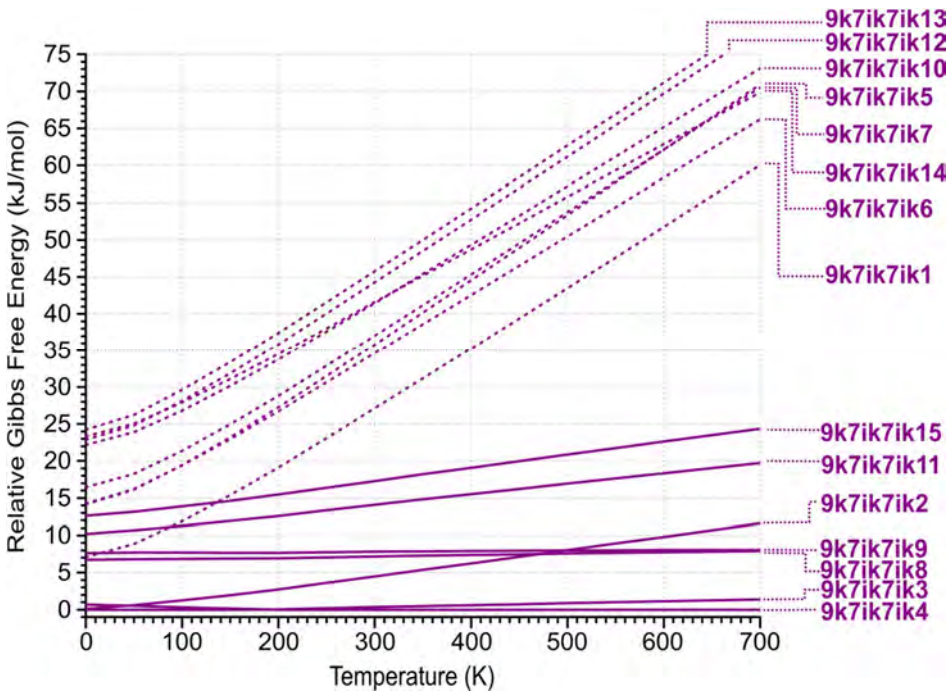


Figure 8.3.78 *Relative Gibbs free energy of the 7k7k7k guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

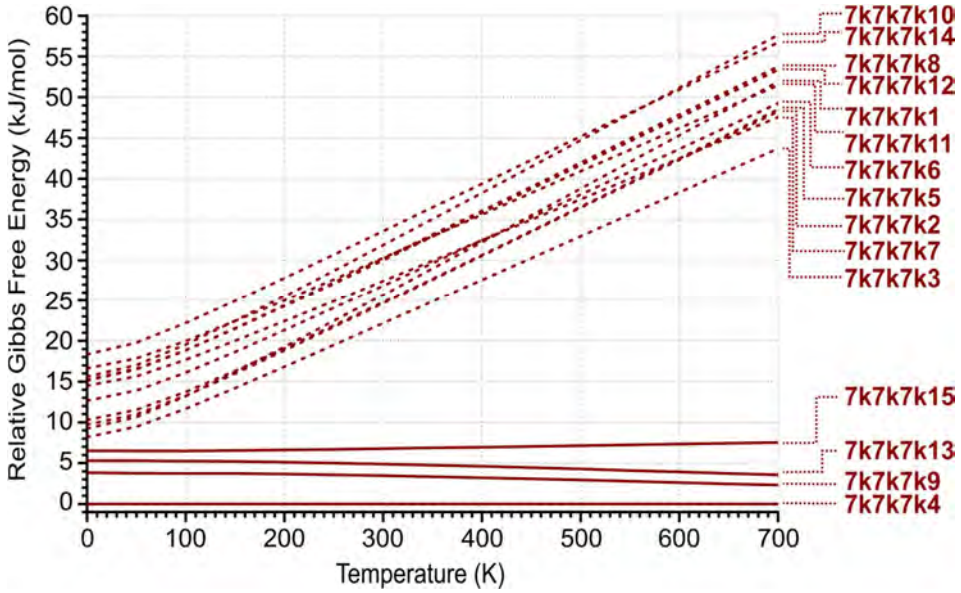


Figure 8.3.79 *Relative Gibbs free energy of the 7k7k7ik guanine trimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

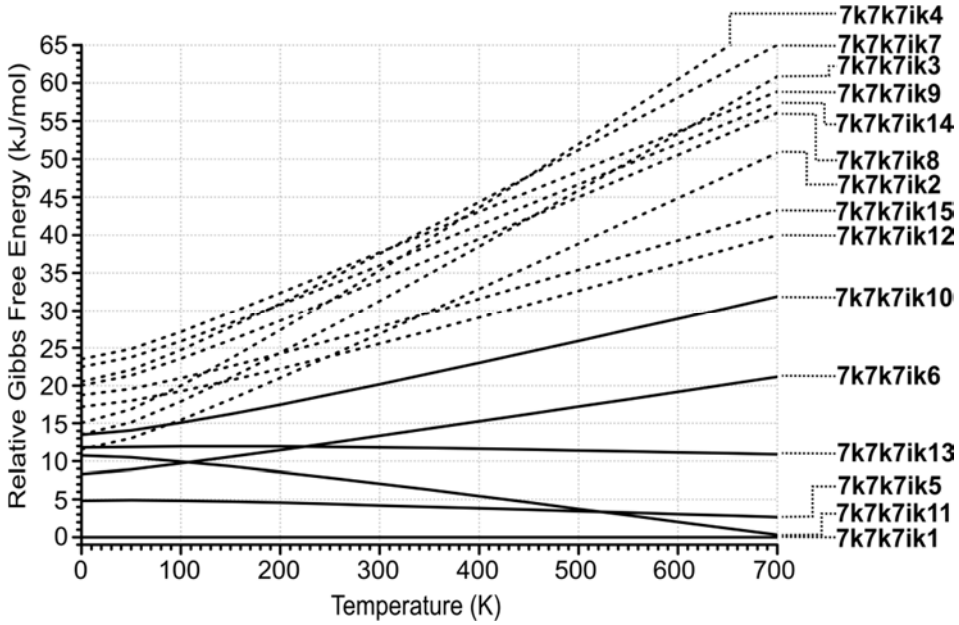


Figure 8.3.80 *2D-NCI plots (s vs. ρ*sign(λ₂)) for the considered trimers of guanine.*

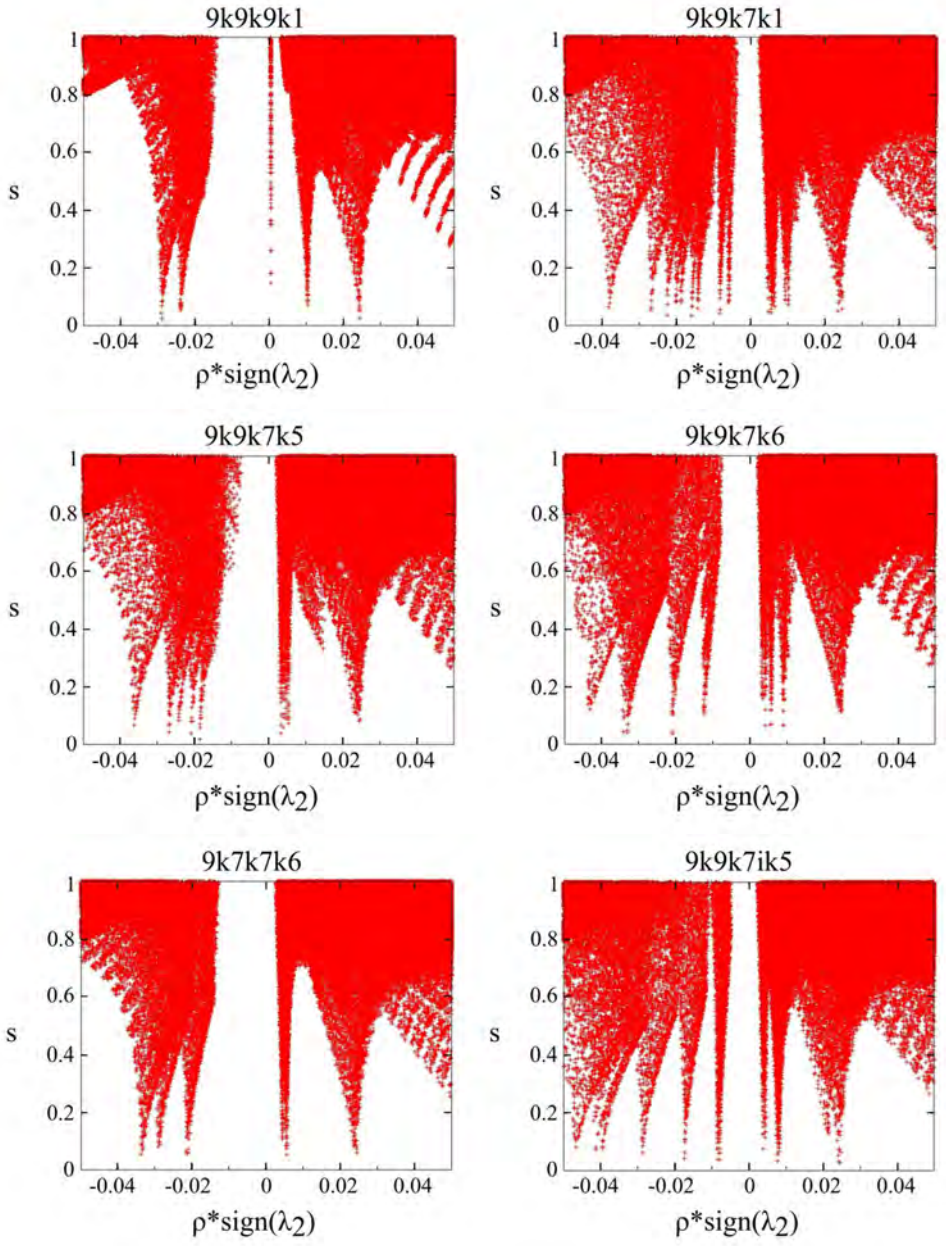


Figure 8.3.81 *IR spectra of dimer of guanine, scanned with the UV wavelenght fixed at 33108 cm^{-1}(blue trace), 33734 cm^{-1} and 34063 cm^{-1}(blak trace). The positions of the band has been jutted out with different color bars.*

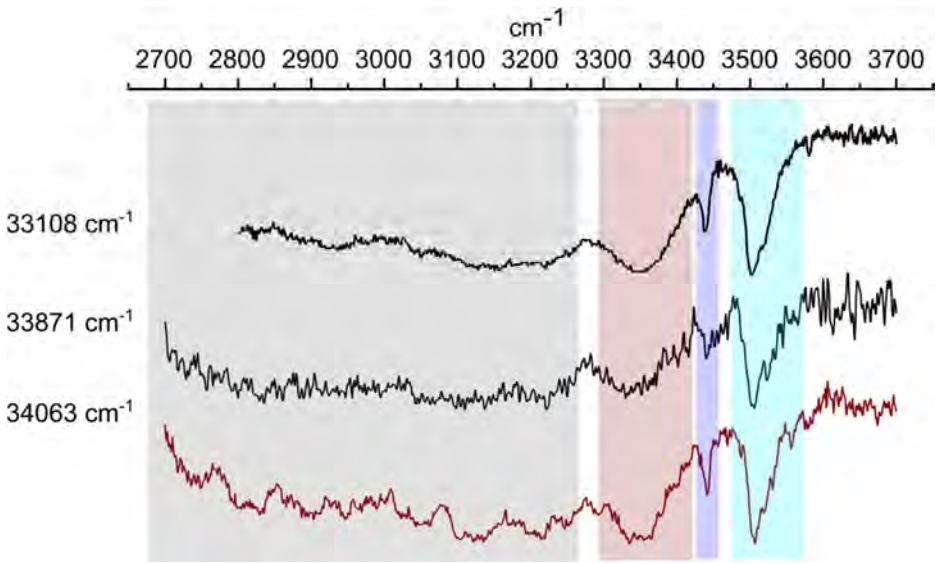


Figure 8.3.82 *The 8 most stable structures of 9m-kk methylguanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

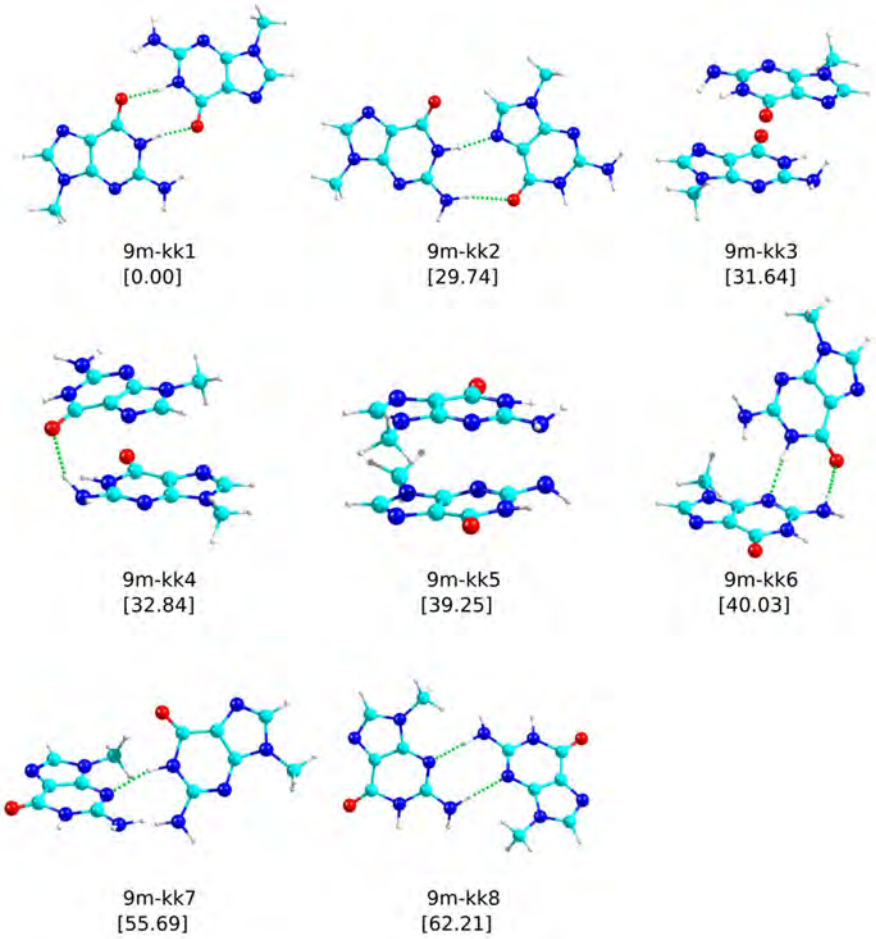


Figure 8.3.83 *The 15 most stable structures of 9m-ke methylguanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

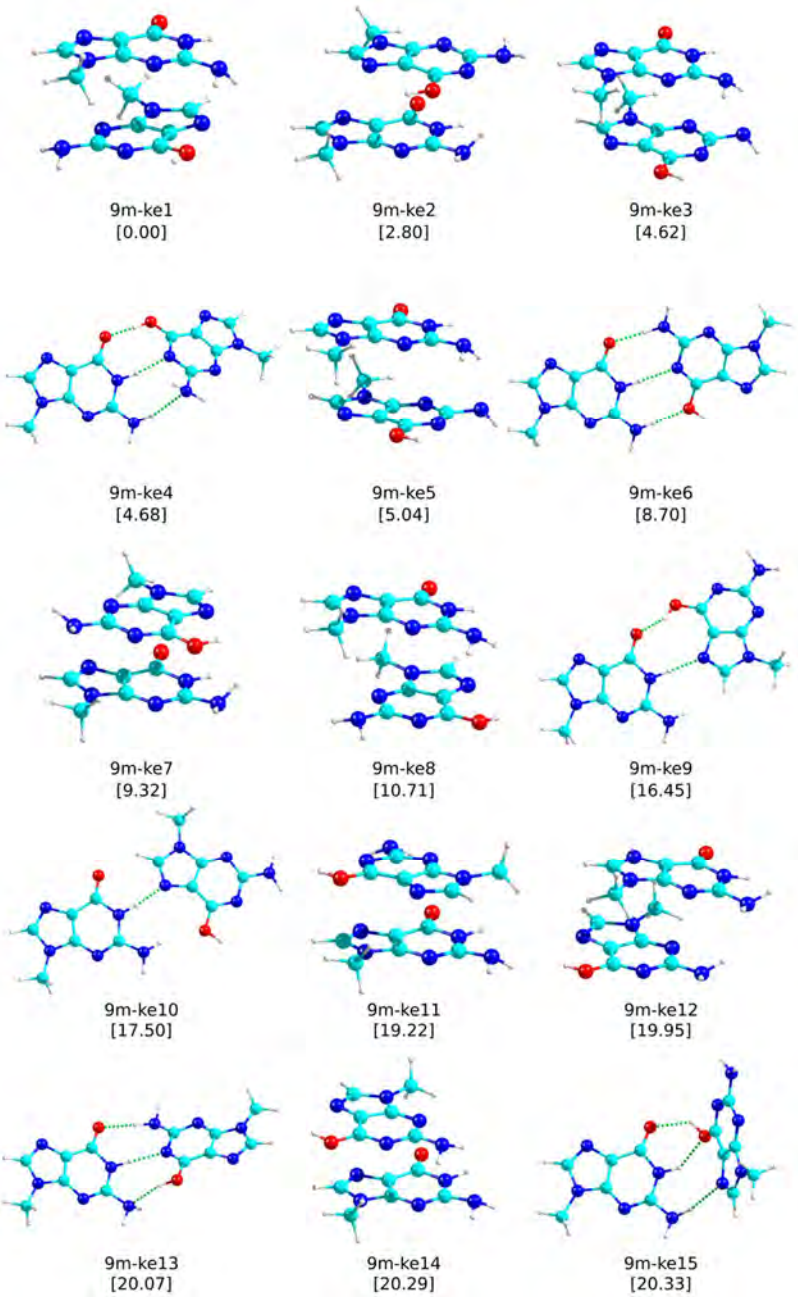


Figure 8.3.84 *The 15 most stable structures of 9m-ee methylguanine dimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

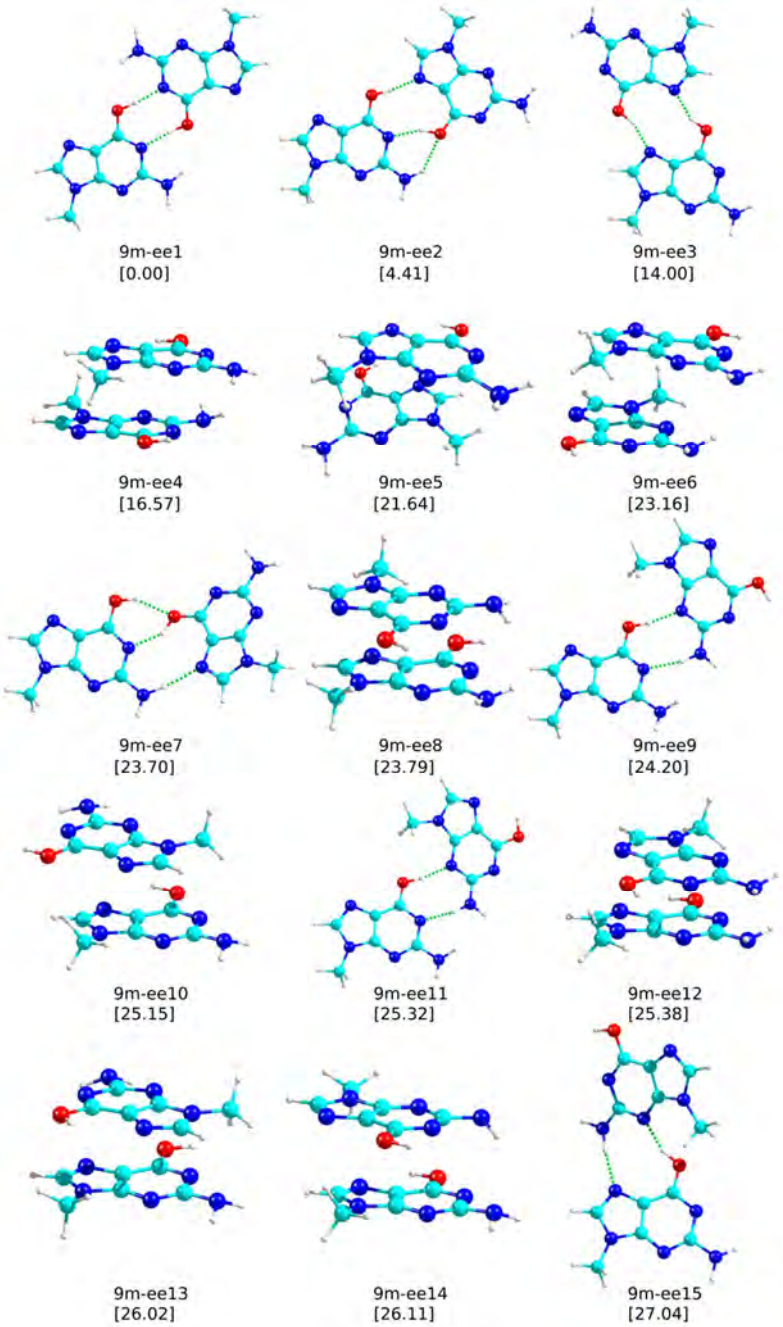


Figure 8.3.85 *Relative Gibbs free energy of the 9m-kk methylguanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

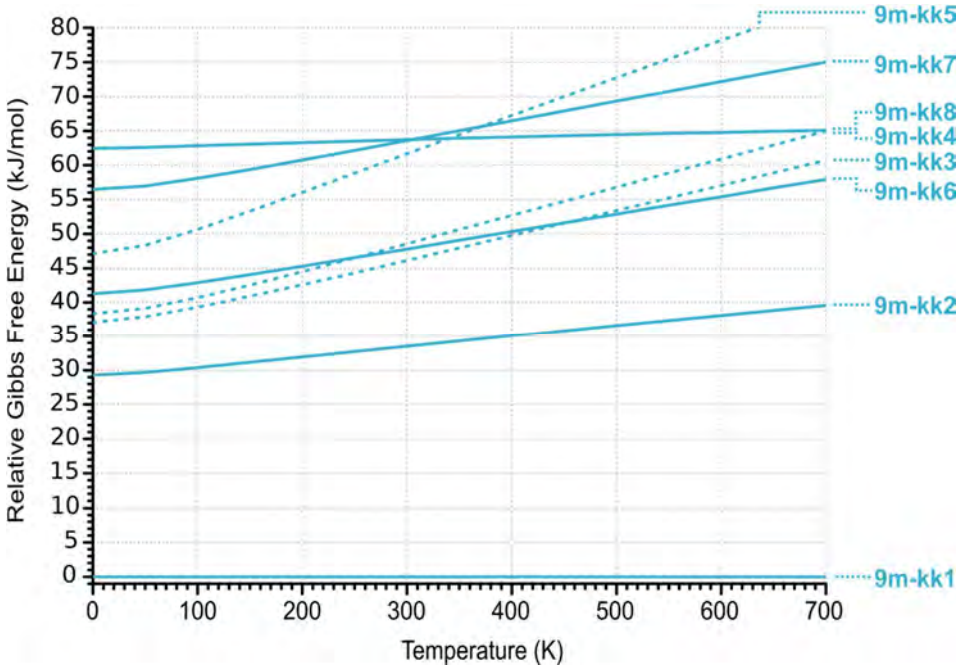


Figure 8.3.86 *Relative Gibbs free energy of the 9m-ke methylguanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

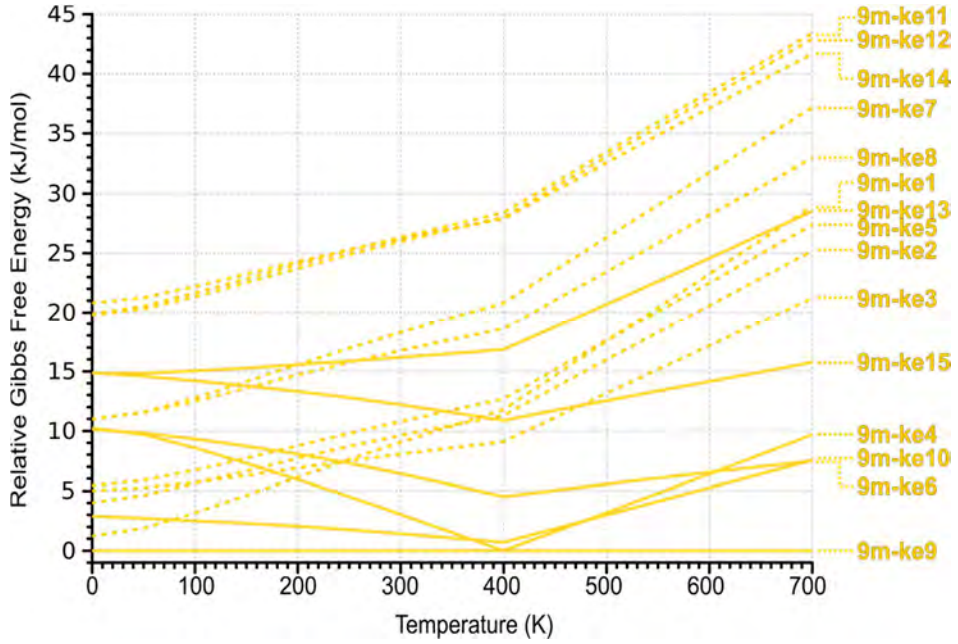


Figure 8.3.87 *Relative Gibbs free energy of the 9m-ee methylguanine dimer in the 0-700 K interval. Continous line: planar structure; Dot line: stacking structure.*

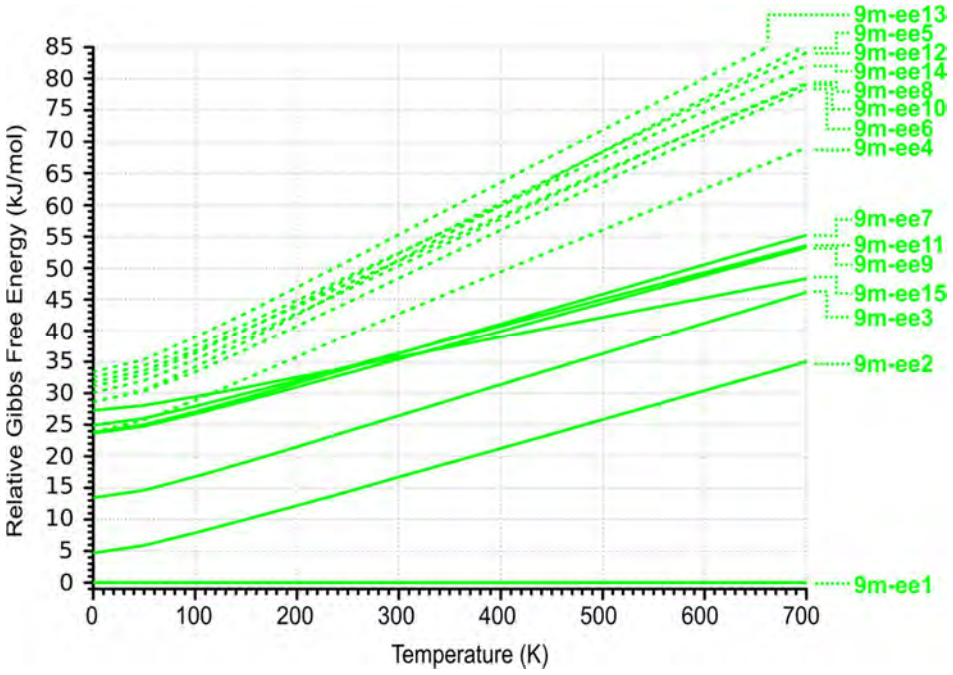


Figure 8.3.88 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered dimer of 9-methylguanine.*

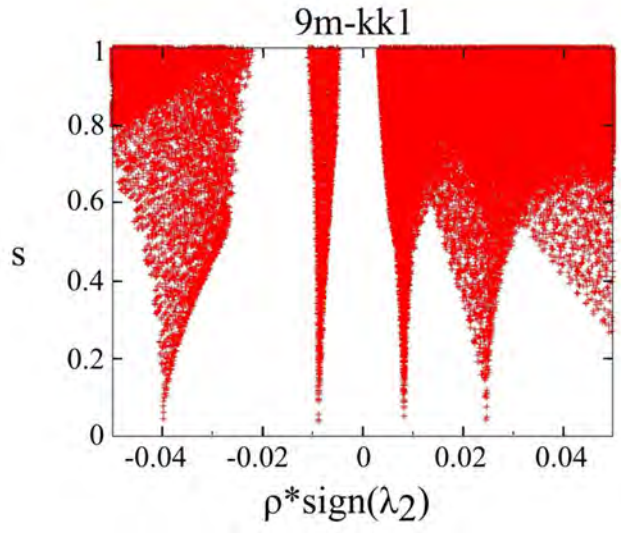
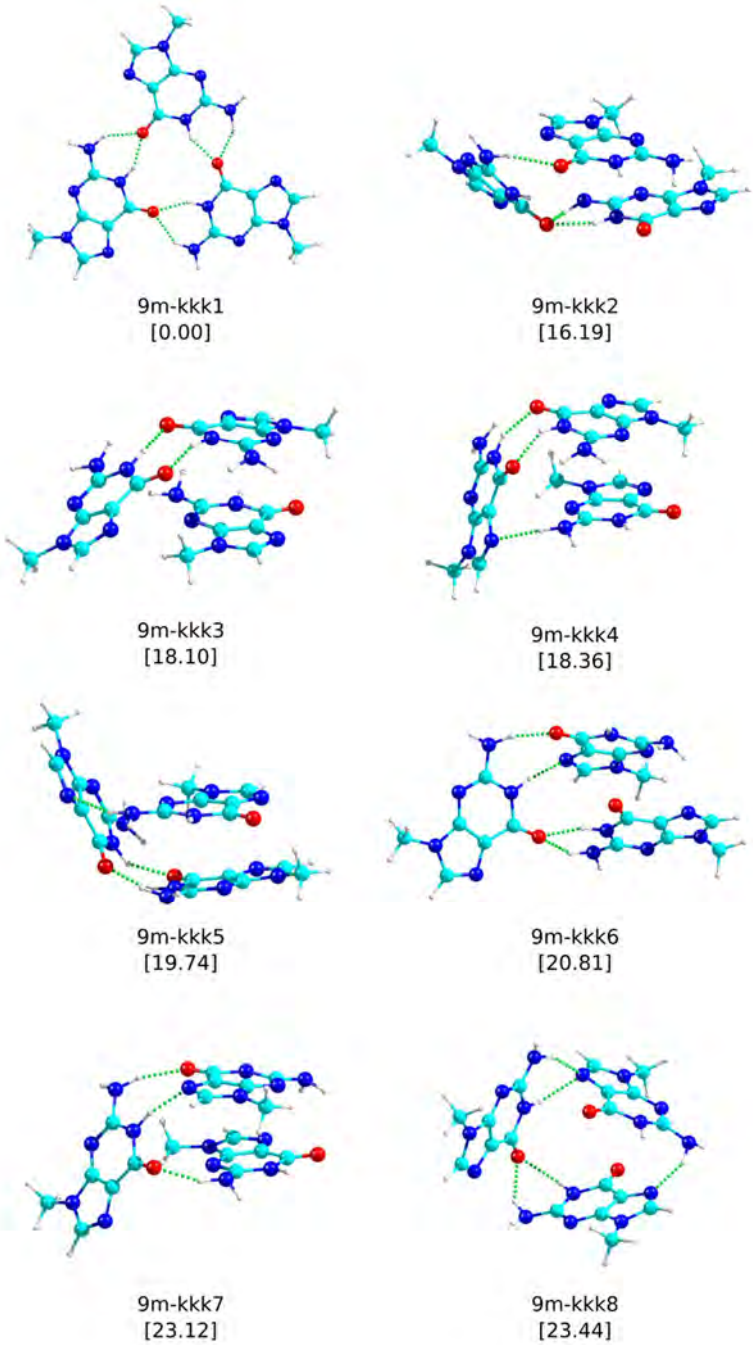
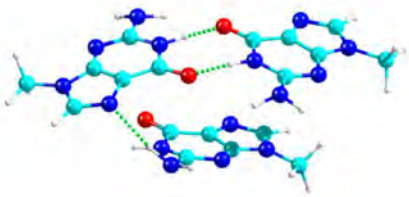
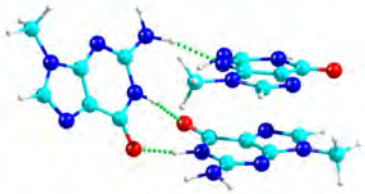


Figure 8.3.89 *The 15 most stable structures of 9m-kkk methylguanine trimer at M06-2X/6-311++G(d,p) level, together with their relative stability (kJ mol^{-1}) in brackets. ZPE correction was applied to the energy values.*

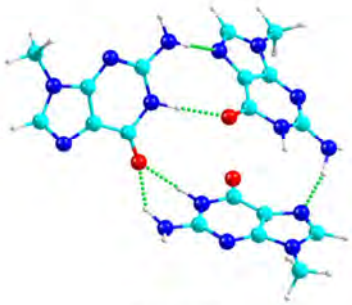




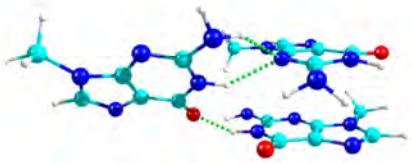
9m-kkk9
[24.04]



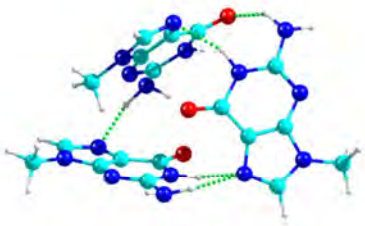
9m-kkk10
[24.08]



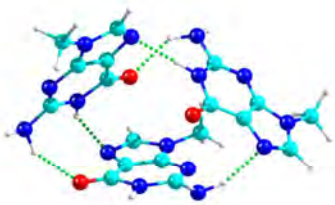
9m-kkk11
[24.80]



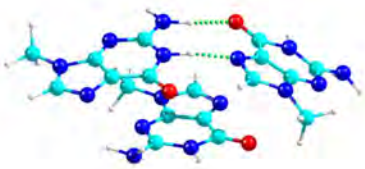
9m-kkk12
[25.49]



9m-kkk13
[25.78]



9m-kkk14
[29.20]



9m-kkk15
[30.96]

Figure 8.3.90 *2D-NCI plots (s vs. ρ*sign(λ_2)) for the considered trimer of 9-methylguanine.*