Babeş-Bolyai University Cluj-Napoca
Faculty of Chemistry and Chemical Engineering

PhD Thesis

# Computational investigations on weak and noncovalent bonding interactions in biomolecules

Scientific Advisor
Assoc. Prof. Radu Silaghi-Dumitrescu

Ph.D. Student
Juan Francisco Carrascoza Mayén

Cluj-Napoca
2016

Babeş-Bolyai University Cluj-Napoca
Faculty of Chemistry and Chemical Engineering

**PhD Thesis**

# Computational investigations on weak and noncovalent bonding interactions in biomolecules

| Scientific Advisor | Ph.D. Student |
|---|---|
| Assoc. Prof. Radu Silaghi-Dumitrescu | Juan Francisco Carrascoza Mayén |

Jury

| President | |
|---|---|
| Prof. Dr. Vasile-Mircea Diudea | Universitatea Babeş-Bolyai, Cluj-Napoca |
| **Reviewers** | |
| Prof. Dr. Vasile Chiş | Babeş-Bolyai University, Cluj-Napoca |
| CS1. Dr. Attila Bende | National Institute for R&D of Isotopic and Molecular Technologies, Cluj-Napoca |
| Conf. Dr. Ionel Humelnicu | Alexandru Ioan Cuza University, Iaşi |

Cluj-Napoca
February 2016

**Contents**

**Abstract**

In this work the nature of the proteins and peptides is approached from a theoretical point of view, divided into two chapters. The first chapter deals with topics related to the folding of the protein taking into account the mechanics, and the structural properties of the amino acids. More precisely, on the issue of how the amino acids dictate/control the construction of secondary and tertiary structures. Additionally, the first section also includes a section proposing a revision of those structural factors that drives the anomeric effect, but with general implications for the non-covalent interactions of oxygen and sulfur in biomolecules. A range of methods starting from *ab initio* (including post-HF) and DFT and going to molecular mechanics, are employed to these ends.

The second chapter deals with interactions of biomolecules with ligands, as well as with protein-protein interactions, in metalloproteins. Particular interest is placed on globins and DNA, where specific studies of docking protein-ligand and protein-protein are described, trying to provide further insight into selected topics of current debate.

This work also provides novel force field parameters for molecular dynamics of particular metallo-enzymes and drugs.

**Keywords:** metalloprotein; molecular dynamics; QM/MM; docking; hemoglobin; bleomycin; DFT; Amber; force fields; hemerythrin; haptoglobin; protein folding.

Protein folding patterns are generally accepted to rely on non-covalent interactions within the polypeptide chain, as well as on external factors [1,2]. Among the environmental factors that influence the preference for a given folding pattern, are those involving non-amino acidic structures such as the solvent, organic ligands, metals and others [3]. In most cases, probably in all of them, these influences are crucial for the protein's functionality [4].

In this thesis, theoretical studies of protein interactions are reported, *centered on the importance of weak and non-covalent interactions as steps towards building a better understanding of the roles played by metals in biomolecules*. A comprehensive battery of techniques were employed to this extent, illustrating the complexity of the challenges laid out by such systems, and our current ability to deal with them. This includes the comparative use of several force fields for geometry optimizations as well as for dynamics and for docking calculations. Parameterization of organic and inorganic centers was also performed for cobalt and iron centers. The dynamics were performed in several regimens, including the periodic boundary (PBC) conditions. Also applied were semiempirical methods, density functional theory (DFT), Hartree-Fock, and wave function analysis procedures such as the NBO – all for ground state examination as well as for isomer and conformer exploration. Last but not least, *ab initio* dynamics were also applied.

Chapter 1 examines the degree to which weak interactions drive the folding pattern in biomolecules, and primarily in proteins. Examples include the importance of intra – amino acid interactions in folding, as expressed in DFT-derived Ramachandran-like diagrams. Also, a critical analysis of the performance of computational methods (from molecular mechanics to semiempirical, DFT and Hartree-Fock) in accurately describing a quintessential model of biomolecular folding – the peptide α-helix. Applications of the better-performing methods are also given for another canonical model of folding – the β sheet as seen in the protein fibroin, where semiempirical and DFT calculations allow some rationales to be put forth on the particular choice of primary structure in this protein. The final section of Chapter 1 deals with an application on sugars, as an offshoot of the study as well as an illustration of the capability to apply these computational methods beyond the polypeptide case.

Chapter 2 presents a set of cases of protein interactions with external elements. This includes the theoretical analysis of the haptoglobin conformational changes when interacting

with hemoglobin (Hb), as well as on the manner in which hemoglobin and myoglobin interact with small-molecule antioxidants. These interaction is relevant not only as an example of a physiologically-relevant protein-protein interaction, but also for improving the understanding of the manner in which highly reactive species (e.g., the heme group in hemoglobin, which is well known for being a radical generator) are dealt with in our bodies. Also, a description is given of the parameterization and exploratory studies on the dynamics of the non-heme iron-containing protein, hemerythrin, recently proposed as a possible replacement of Hb in blood substitutes precisely based on the distinctly lower reactivity towards oxidative stress agents. Last but not least, Chapter 2 treats yet another case of oxidative stress-related metal-biomolecule interaction. By contrast with the other sections of Chapter 2, this final part deals not with metalloproteins but rather with a DNA-metallopeptide interaction – an arguably more complex matter from a methodological point of view.

Appropriate calculation of molecular conformations is not trivial, since the number of atoms in protein-like structures is often more than 100 atoms (one needs only, to the limit, remind the Levinthal paradox in this respect). High-level theories like post Hartree-Fock, and even Hartree-Fock itself, may be quite expensive from the point of view of computational resources; this problem is sometimes circumvented by carefully choosing models of smaller size - though one may nevertheless note that using fewer atoms offers no warranty of having a simpler problem. More than often, a compromise between accuracy and resources needs to be attempted, which opens room for the implementation of other theories that work faster - like semiempirical and molecular mechanics.

A common challenge when calculating biomolecular folding is in the correction of the energy contributions due to hydrogen bonds [3,5]. Hydrogen bonds can be energetically and structurally underestimated easily even at the DFT level, especially since explicit water solvent is too expensive in computational resources to be taken into account for these types of calculations. Therefore, the first approximation that is needed, is the use of implicit solvent [6], which at times may implicate a potential geometrical error due to the absence of distinct hydrogen bonds with solvent, leading to incorrect assignment of the energetic contributions; we indeed show some such results in our work [7].

Hartree-Fock theory is generally not recommended for calculations that contains more than few atoms (i.e. 25), and it therefore is not the best approach for protein-like structures. Still,

Hartree-Fock has been useful for analyzing small atomic interactions such as those in enzyme active sites (for a review of the topic see [8]).

Probably one of the most popular methodologies for analyzing atomic interactions in amino acids is DFT, which has been shown to yield good results in relatively short time. Here it is worth to mention the Truhlar M06 series of functionals (and subsequent variations thereof), which includes variants with parameters for transition metals, particularly relevant when studying metallo-enzymes [8,9].

Semi-empirical methodologies may in principle be better suited for protein-like structures, since the incorporated approximation of pre-calculated parameters inside the Hartree-Fock equation remarkably accelerates these computations [7].

Yet the main problem remains in the balance between resources (time, computer power) and accuracy (theory level). At the time of writing of this manuscript, theory and hardware has evolved enough that it is possible to obtain accurate calculations at 1.0 Å resolution of polypeptide structures under 80 amino acids using a reasonable amount of computer power [10]. Undoubtedly, the foreseeable future will allow further improvements.

# Chapter 1. Computational estimations of contributions pertaining to weak/non-covalent bonding interactions in the structures of biomolecules

## 1.1 Computational study of protein secondary structure elements: Ramachandran simulations [1]

The secondary structure conformation of a protein is often described as a function of its backbone dihedrals expressed in (φ, ψ) pairs that can be represented in a Ramachandran type graphic for easier interpretation. These plots are typically split in forbidden and allowed regions [1,2]. Around 40% of all the amino acids in a structure are contained in just the 2% of the Ramachandran plot – the so-called "allowed areas" [3,4]. The non-allowed regions are those defined by a minimal contact distance between atoms of the neighboring amino acids (n+1) and (n - 1), 'n' being the amino acid with the central α-carbon of reference [5]. Measuring the changes

[1] Based on Carrascoza, Francisco; Zaric, Snezana; Silaghi-Dumitrescu, Radu. Computational study of protein secondary structure elements: Ramachandran plots revisited. Journal of Molecular Graphics and Modelling, 2014, 50, 125-133

in energy by rotating the ψ and φ dihedral angles around the α-carbon may help understand, in a quantitative way, the conformational preferences of three-dimensional peptide structures - and allow predictions thereon [5-7]. Choices for a given type of secondary structure are dictated by interatomic interactions– both repulsive and attractive (primarily classical and non-classical hydrogen bonds) [8].

In our study, nineteen analog structures of (S) 2-acetamido-N-methylpropanamide were constructed, each similar to one of the nineteen key amino acids. The peptide bonds are each capped with a methyl group. Proline was not examined, because its dihedral angles are constrained due to its internal molecular structure. Potential energy surfaces were then built for the ψ and φ angles in each model using the DFT methods M06-2x/6-311** [1] and PW91/6-311** in vacuum, as implemented in the Gaussian09 software package [2]. The scans were performed in twelve steps of thirty degrees each, giving a total of 169 structures which may be viewed as a matrix of 13 x 13 molecules from -180 degrees, until +180 degrees for ψ and -180 to +180 degrees for φ.

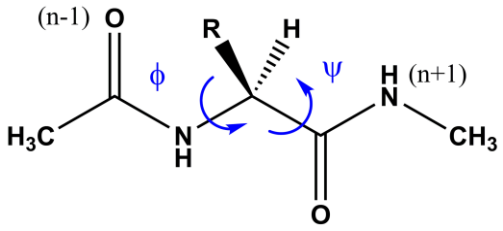



**Figure 1.1.1.** Scheme of 2-acetamido-N-methylpropanamide structures examined in the present study. The φ and ψ angles are marked with blue arrows. O (n-1) refers to the oxygen in the position (n-1) where (n-1) is the amino acid previous to the central amino acid (n). In the same way H (n +1) refers to the hydrogen in position (n +1) with respect to the amino acid n. R is the sidechain of the amino acid.

The level of similarity between plots was measured (Table 1.1.1) indicating that between the 19 common amino acid–like structures, 3 major groups may be defined based on their similar levels of deviation. These groups, derived for single-aminoacid models in vaccum, are a reasonable match pretty to the statistically-derived preferences reported in Malkov's work when examining *experimental* preferences in amino acid secondary structures seen in the x-ray protein structures available in the Protein Data Bank. Table 1.1.1 offers a comparison between the similarities in PES plots here reported and statistical preferences previously reported by Malkov et al.

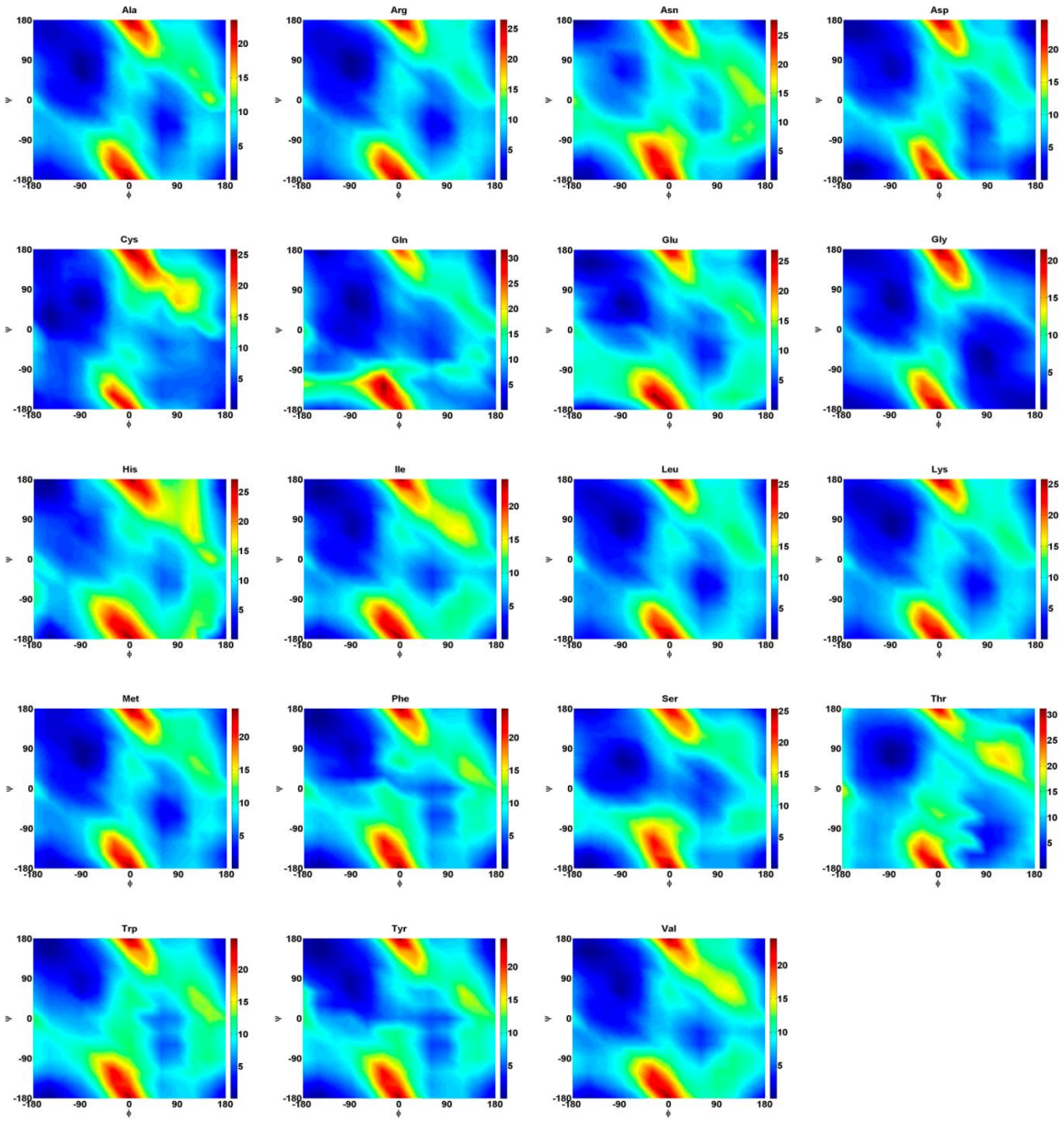



**Figure 1.1.2.** PW91PW91/6-311+G** PES scans for each amino acid model.

Disagreements are of course also found between our DFT-derived data and the experimental/statistics-based classical Ramachandran plots. These differences are explained using supramolecular chemistry considerations: some regions which show local minima in the single-amino acid models are in fact non-profitable at supramolecular level in the sense that interactions with other polypeptide chains or with other amino acids in the same polypeptide chain are hampered. The present data are focused on the intrinsic properties of an amino acid –

**Table 1.1.1.** Deviations calculated for the entire PES scan Plots. ATD = Average Total Deviation. The first column indicates the Groups to which the respective amino acids are assigned. The numbers highlighted in bold point out the members of a group used to calculate the average of this group.

| Group | Amino Acid | ATD | Group 1 | Group 2 | Group 3 |
|---|---|---|---|---|---|
| | Gly | 17.34 | 2.42 | 2.68 | **3.51** |
| | Ala | 15.74 | 0.81 | 1.29 | 2.04 |
| | **Gln** | 15.59 | 1.10 | 1.78 | 2.27 |
| | Lys | 15.11 | **0.42** | 1.27 | 1.86 |
| 1 | Leu | 15.08 | **0.46** | 1.40 | 1.95 |
| | Met | 15.06 | **0.62** | 1.41 | 2.13 |
| | Arg | 14.96 | **0.52** | 1.31 | 1.84 |
| | Ile | 14.95 | **0.80** | 1.64 | 1.98 |
| | Val | 14.88 | **0.85** | 1.70 | 2.03 |
| | Thr | 14.78 | 2.47 | **2.13** | 2.79 |
| | Phe | 14.75 | 1.49 | **1.05** | 1.19 |
| 2 | Tyr | 14.72 | 1.45 | **1.03** | 1.05 |
| | Ser | 14.67 | 1.70 | **1.18** | 1.55 |
| | Trp | 13.90 | 1.82 | 1.40 | **0.81** |
| | **His** | 13.64 | 2.28 | 1.93 | 1.92 |
| | Asp | 13.56 | 2.09 | 1.53 | **0.85** |
| 3 | Glu | 12.87 | 2.62 | 2.20 | **1.44** |
| | **Cys** | 11.93 | 3.30 | 3.27 | 2.86 |
| | Asn | 11.59 | 3.94 | 3.39 | **2.42** |

**Table 1.1.2.** Preferences of amino acids for various types of secondary structure, as derived from statistics (Malkov et al., [3] column 2) and DFT (present work).

| Amino acid | Group by similarity | Group by Statistics | DFT favored minima | DFT-derived second choice |
|---|---|---|---|---|
| Ala | 1 | α-helix | $2.2_7$-ribbon | β-sheet |
| Leu | 1 | α-helix | $2.2_7$-ribbon | β-parallel, bend |
| Glu | 3 | α-helix | β-sheet antiparallel | $2.2_7$-ribbon |
| Gln | 1 | α-helix | $2.2_7$-ribbon | β-sheet antiparallel |
| Arg | 1 | α-helix | $2.2_7$-ribbon | β-sheet, bend |
| Met | 1 | α-helix | $2.2_7$-ribbon | β-sheet, bend |
| Lys | 1 | α-helix | $2.2_7$-ribbon | β-sheet, bend |
| Val | 1 | Strand | β-sheet antiparallel, $2.2_7$-ribbon | β-sheet parallel |
| Ile | 1 | Strand | β-sheet antiparallel, $2.2_7$-ribbon | β-sheet parallel |
| Tyr | 2 | Strand | β-sheet planar, Antiparallel | β-sheet parallel, $2.2_7$-ribbon |
| Phe | 2 | Strand | β-sheet Planar, Antiparallel | β-sheet parallel, $2.2_7$-ribbon |
| Thr | 2 | Strand | Bend type II | 2.27-ribbon, $3_{10}$Turn I |
| Trp | 3 | Strand | β-sheet-Planar, Antiparallel | β-sheet-parallel, $2.2_7$-ribbon |
| Gly | - | Other | $2.2_7$-ribbon | β-sheet, bend |
| Asn | 3 | Other | β-sheet planar | $2.2_7$-ribbon |
| Pro | N.A. | Other | N.A. | N.A. |

an element which to our knowledge is typically ignored, as larger models are always used for the sake of similarity to real biological polypeptides. Nevertheless, for small peptides, including e.g. prebiotic chemistry, such small models hold information not readily available from experimental observations on full protein/polypeptide. Further work is in progress to quantify the effects of intermolecular near-neighboring amino acids, relative to the intra-amino acid ones.

### 1.2 Modelling of Alpha-helical structures [2]

In the present work, geometry optimization results are reported on elements of secondary structure in small proteins and polypeptides. These optimizations were performed either in vacuum or with the CPCM continuum solvent model as implemented in Gaussian09 software package [5] at PM6-D2, HF/3-21G*, M062X/6-31G**, M062/6-311+G**, BP86/6-31G**. MM calculations were done with Amber and UFF force fields as implemented in Gaussian09 [5] and Hyperchem [6]. The semiempirical PM6-D2 method was employed as implemented in the Gaussian09 [5] and MOPAC [7] software packages; the PM3 implementation from Hyperchem [6] was also employed.

With the purpose of providing a critical account of the performance of computational methods, as well as in order to identify the most accurate ones and the limitations thereof in this particular context of polypeptide supramolecular chemistry, α-helical structures in purely theoretical models ($Gly_{10}$, $Gly_{50}$, $Ile_{10}$, $Ile_{50}$) as well as in realistic models (a calmodulin α-helical stretch, and a complete small protein structure, PDB ID's 3CLN [48] and 1ALG respectively) were examined. PM6-D2 has proven to give the most accurate geometrical descriptions for the Gly and Ile models, with M06-2X following suit closely especially when solvation is included. By contrast, most of the other methods fail dramatically in terms of predicting a reasonable canonical helix. Calmodulin and 1ALG structures are discussed below in more detail as representatives of the results here obtained. Particularly for calmodulin, the AMBER method does not properly model the helical structure, predicting a well-defined bending point midway through the structure (cf. Figure 1.2.1) and several unacceptably long (>3 Å) hydrogen bonding distances (amino acids 74-85). A bent helix is also predicted by the HF approach, which in addition presents the inconvenient that the C-terminal carboxylate is eliminated upon geometry

[2] Based on Lupan, Alexandru; Kun, Attila; Carrascoza, Francisco; Silaghi-Dumitrescu, Radu. Performance of computational methods for modeling α helical structures. Journal of Molecular Modeling 2013, **19**(1), 193-203.

optimization as $CO_2$. The MOPAC PM6 method is the only one preserving entirely the α-helical structure; however, in this case the tendency towards a canonical structure is exaggerated: even though there are hydrogen bond distances as long as 3 Å in the experimental structure, the PM6 method predicts all hydrogen bonds in the structure to be in the ~1.9-2.0 Å range.

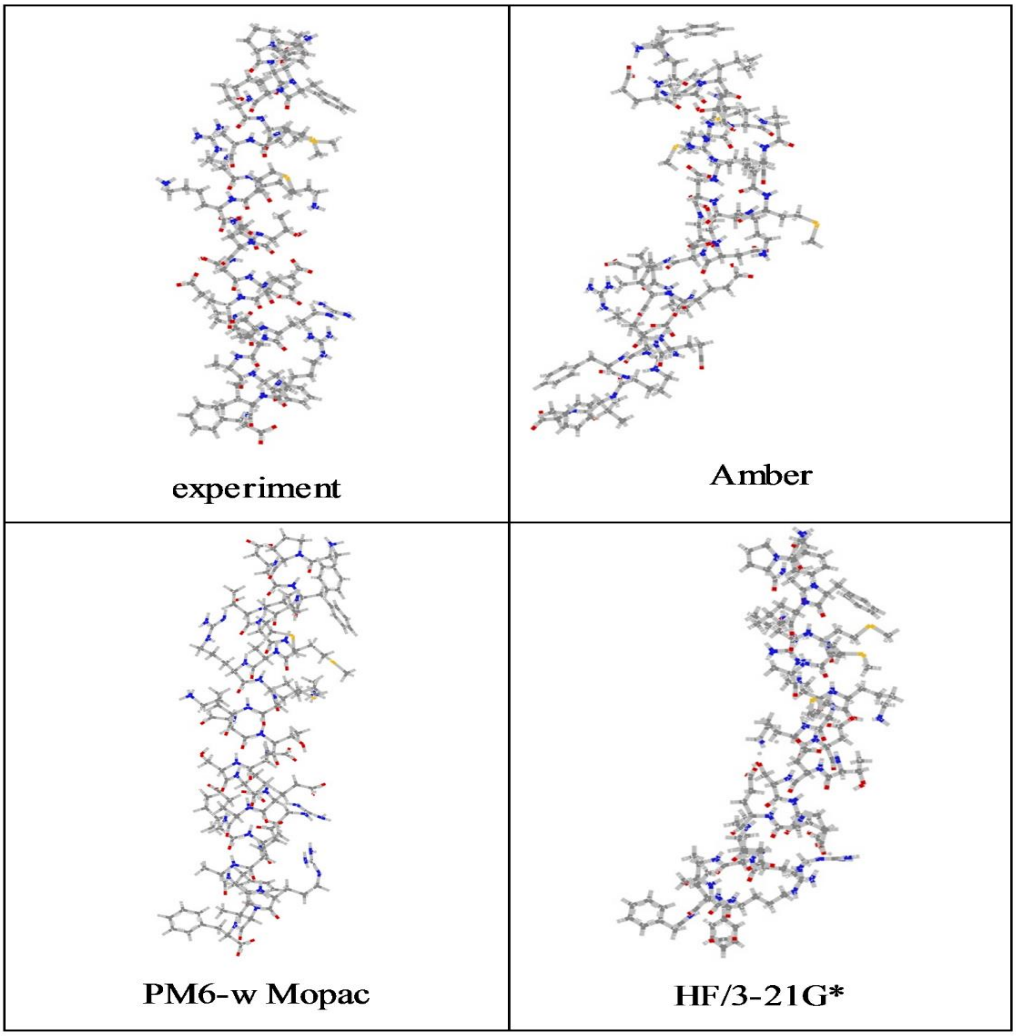



**Figure 1.2.1.** Graphical representation of 3CLN-long loop geometries optimized by various methods.

The PDB structure coded 1ALG is a protein composed of 24 residues with a global α-helix structure. Interestingly, methods which fail to predict canonical helical structures for the smaller/simpler models discussed above (e.g., the Gaussian implementation of PM6, or HF/3-21G* in vacuum), are found to reproduce reasonably well the regularity of the helical structure of the complete, experimentally known, protein 1ALG. In fact, even where the experimental structure shows deviations from α-helix, such as in the first two hydrogen bonds which are

longer and the last hydrogen bond which is distinctly shorter, the computational methods fail to predict these deviations.

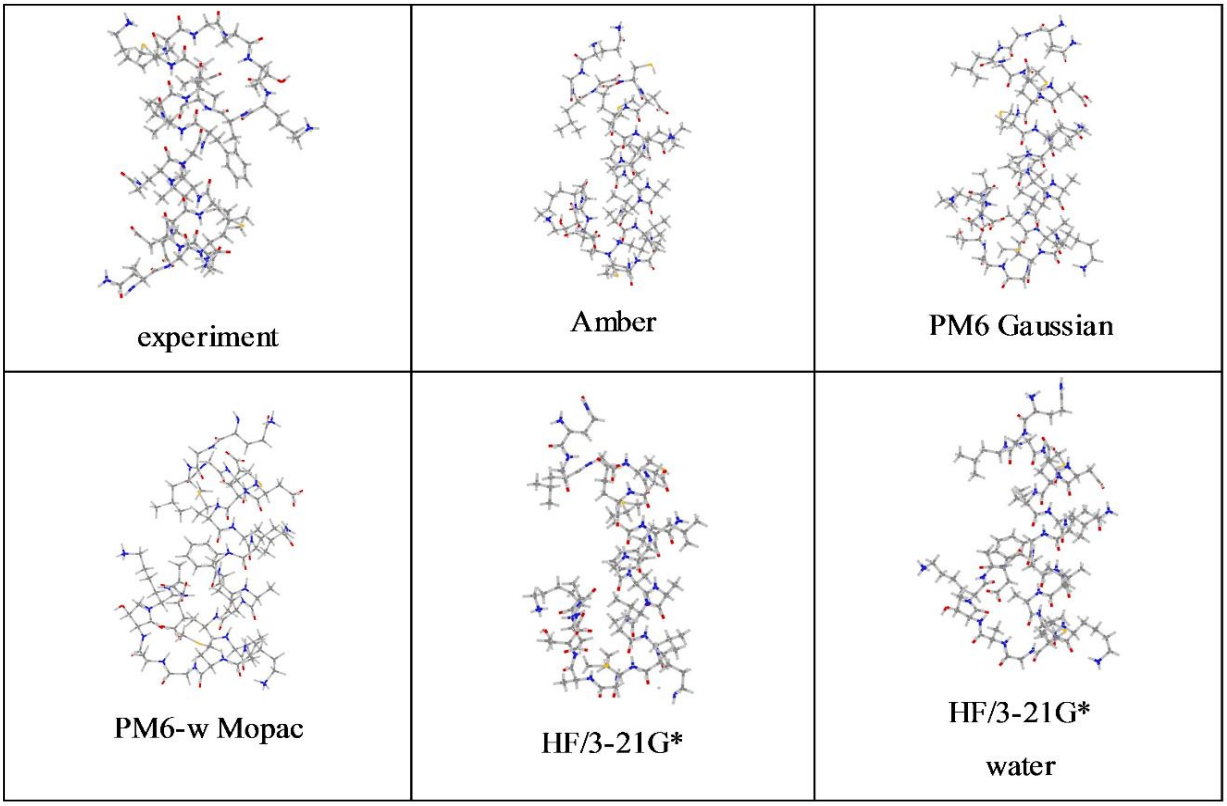



**Figure 1.2.3.** Graphical representation of the 1ALG geometry optimized by various methods.

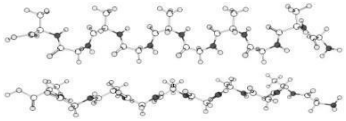
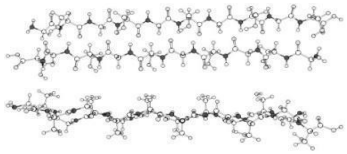
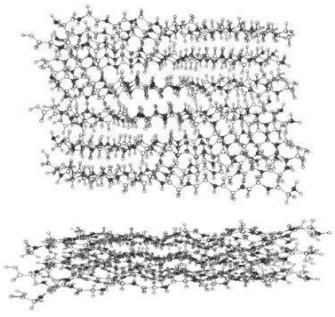
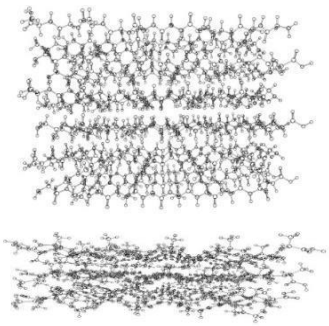
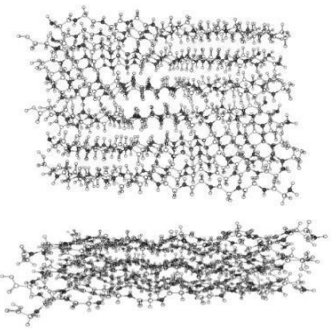
### 1.3 On the roles of alanine and serine in the β-sheet structure of fibroin[3]

In its silk II form, fibroin is almost exclusively formed from layers of β–sheets, rich in glycine, alanine and serine. The rationale for this particular amino acid composition has been speculated upon rather straightforwardly in terms of minimizing sterical repulsions; however, very little numerical evidence has been attached to such arguments so far. Therefore, the present study intends to provide computational data on a set of models designed to delineate the possible individual roles that each of the three aminoacids (Gly, Ala, Ser) play in fibroin. To this end, here we report computational results on fibroin models at semi–empirical PM6-D2 and DFT M062X [13]/6–31G** levels of theory, as well as MM/MD with AMBER all-atom force field

[3] Based on Carrascoza, Francisco; Lupan, Alexandru; Cosar, Ciprian; Kun, Attila; Silaghi-Dumitrescu, Radu. On the roles of the alanine and serine in the β-sheet structure of fibroin Biophysical Chemistry, 2015, 197, 10-17; Cosar, Ciprian. Molecular modeling of fibroin (Master thesis), Babes-Bolyai University, Cluj-Napoca, 1-64 (2012).

calculations which were performed with the AmberToolkit software version 14 [15] simulating physiological conditions in a box of water. In the case of PM6 calculations, electrostatic field was employed within MOPAC, while DFT calculations were performed with both vacuum and CPCM solvent model as implemented in Gaussian09. Models employed include decapeptides $(Gly)_{10}$, $(Gly–Ala)_5$ and $(Gly–Ser)_5$, and 1, 2, 8, 18-mer assembles (Figure 1.3.1 and 1.3.2) .

**Figure 1.3.1.** Obtained geometries after PM6 optimization. N= Number of monomers in the model. 1*= monomer extracted from the center of an 18-mer assemble.

| N | | Gly-Gly | Gly-Ala | Gly-Ser |
|---|---|---|---|---|
| 1* | Top<br>Side |  |  |  |
| 1 | Top<br>Side |  |  |  |
| 18 | Top<br>Side |  |  |  |

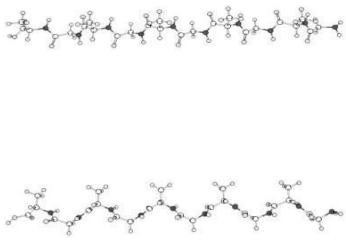
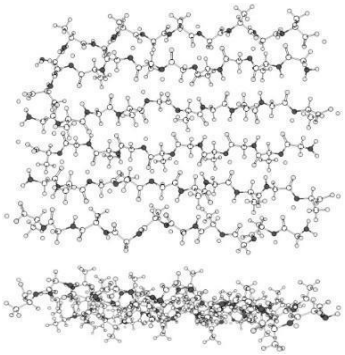
| | | $(GlyAla)_5$ | $[(GlyAla)_5]_{18}$ |
|---|---|---|---|
| Solvent explicit | Top<br>Side |  |  |

**Figure 1.3.2.** PM6 optimized geometries of fibroin–like models. **Left:** hydrogen bonds distance, angle and formation energy. **Center:** lateral view. **Right**: top view**. A.** $(Gly_{10})_{18}$ antiparallel *E***. B.** $((Gly–Ala)_5)_{18}$ antiparallel *E***. C.** $((Gly–Ser)_5)_{18}$ antiparallel *E***.** bonds. **B. Left.** Lengths of hydrogen bonds for a monomer taken from the center of the 18 monomers structure. **D. Right:** tridimensional model of the octadecamer assembles optimized. **Abbreviations:** HB: hydrogen bond average energy.

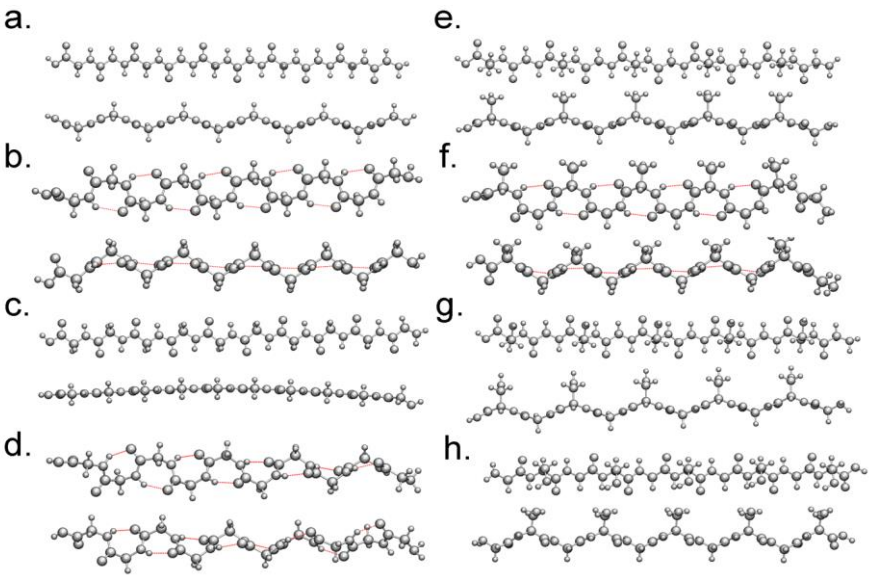
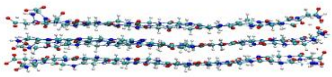
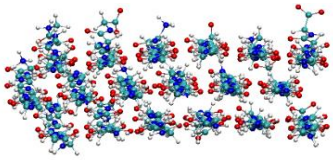
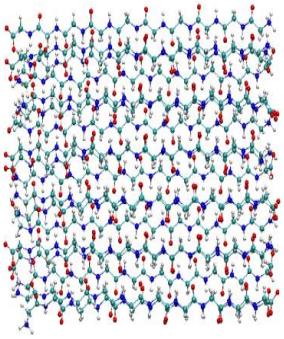



**Figure 1.3.3**. $Gly_{10}$, $(Gly–Ala)_5$ and $(Gly–Ser)_5$ decapeptides organized as single β–sheets: geometries before and after optimization with DFT (M06–2x/6–31G**) and PM6 methods. Internal hydrogen bonds are indicated with dashed red lines. Top and side views are shown for each structure. **A–** $Gly_{10}$ starting geometry. **B–** PM6 –$Gly_{10}$optimized. **C–** Non–pleated $Gly_{10}$ DFT–optimized. **D**– Pleated $Gly_{10}$ DFT–optimized. **E–** $(Gly–Ala)_5$ starting geometry. **F–** $(Gly–Ala)_5$ PM6 optimized. **G–** $(Gly–Ser)_5$ starting geometry. **H–** $(Gly–Ser)_5$ PM6–optimized.

| | Lateral view | Frontal view | Top view |
|---|---|---|---|
| G–G |  |  |  |

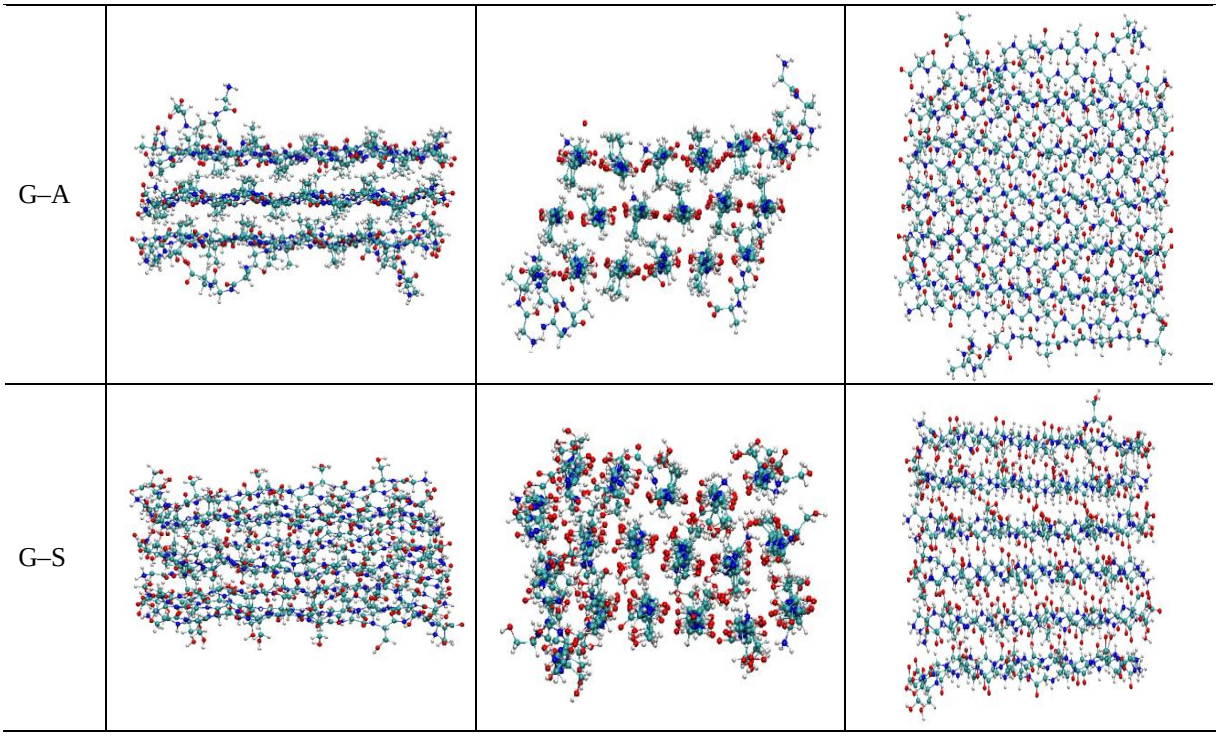



**Figure 1.3.4.** Eighteen–mer structures after 5 nanoseconds of MM/MD.

PM6–derived stabilization energies in the 18 decapeptides on three layers of antiparallel *E* β-sheets suggest that rather than stabilizing the multi–layer pleated β–sheet structure, alanine and serine in fact introduce sterical repulsions (Figure 1.3.3). On the other hand, the side chain of the alanine adds stability to the rigidity of the sheet, allowing it to maintain a properly pleated structure even in a single β–sheet, avoiding two alternative conformations which would interfere with the formation of the multi-layer pleated–sheet structure. These two alternative conformations are a completely planar–sheet, they are less stable than the pleated–sheet, and, more importantly $2.2_7$ ribbon conformation. The role of the serine is proposed to involve modulation of the hydrophobicity, so that the newly synthesized peptide can be amenable to larger scale manipulation upon exiting in the ribosome, in order to construct the supramolecular assembly as opposed to randomly precipitating due to hydrophobicity. The importance of understanding the roll that plays each amino acid in the polymeric structure may be relevant in areas like theoretical model of peptides, materials design or biopolymers manufacture, for instance.

## 1.4 A twist in the anomeric effect [4]

Several explanations have been proposed for the anomeric effect – based mainly on sterical interactions, charge separation, and hyperconjugation. Revisiting this topic with computational methods, we find that the pyranoid oxygen is not $sp^3$-hybridized, and as a result of this situation one of the oxygen lone pairs is found in an eclipsed conformation with respect to an equatorial substituent at a neighboring carbon atom. This sterical conflict by itself appears as an important cause of the anomeric effect. The non-$sp^3$-hybridized nature of the oxygen atom is in fact found not to be limited to carbohydrates, but rather be encountered in basic structural motifs (e.g., water, methanol, formaldehyde).

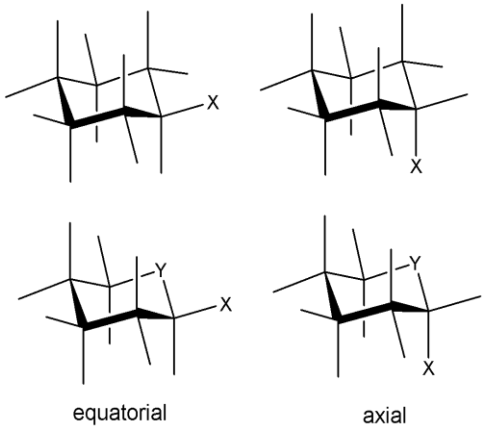



**Figure 1.4.1.** Structures of interest for the present study. X and Y are typically oxygen atoms, in which case the structural motif is relevant for carbohydrate chemistry.

Geometries for all models were optimized using the BP86 functional [17,18], with 6-31G** basis set as implemented in Spartan software package [19]. For the SCF calculations, a fine grid was used, and the convergence criteria were set as standard. NBO analysis were obtained optimizing using DFT BVP86 functional and 6-31G** basis set, at vacuum as implemented in Gaussian09 [2] using default SCF convergence criteria for geometry optimization.

Molecular orbitals were computed at DFT level for the axial and equatorial isomers of tetrahydropyran-2-ol, one of the simplest compounds expected to present an anomeric effect. It may be seen that one of the pyranoid oxygen lone pairs is perpendicular to the ring's C-O-C plane (in the HOMO), while the other lone pair of the same oxygen atom is contained in this

[4] Based on Silaghi-Dumitrescu, Radu; Carrascoza Mayén, Juan Francisco. A twist in the anomeric effect. Studia Universitatis Babes-Bolyai Seria Chemia 2014, LIX(3), 95-101.

plane (HOMO-2 in the equatorial isomer, HOMO-1 in the axial one). Thus, the pyranoid oxygen atom has two p orbitals essentially perpendicular to each other.

This situation is at odds with the $sp^3$ hybridization generally invoked when discussing the anomeric effect and in fact generally expected of oxygen atoms in organic chemistry. Two p orbitals perpendicular with each other, in an atom not involved in multiple bonding, would appear to best be explained as a non-hybridized atom.

This situation, with a non-hybridized oxygen atom, then poses a simple conformational problem – substituents found in equatorial positions at the carbon atoms directly bound to the pyranoid oxygen will be found in an *eclipsed* conformation with one of the lone pairs of the oxygen atom, thus being energetically-disfavored, cf. Figure 1.4.2. *This finding by itself offers a strong explanation for the anomeric effect.*

The *eclipsed* conformation induced by the lack of hybridization at the anomeric oxygen atom is the main reason for the anomeric effect.

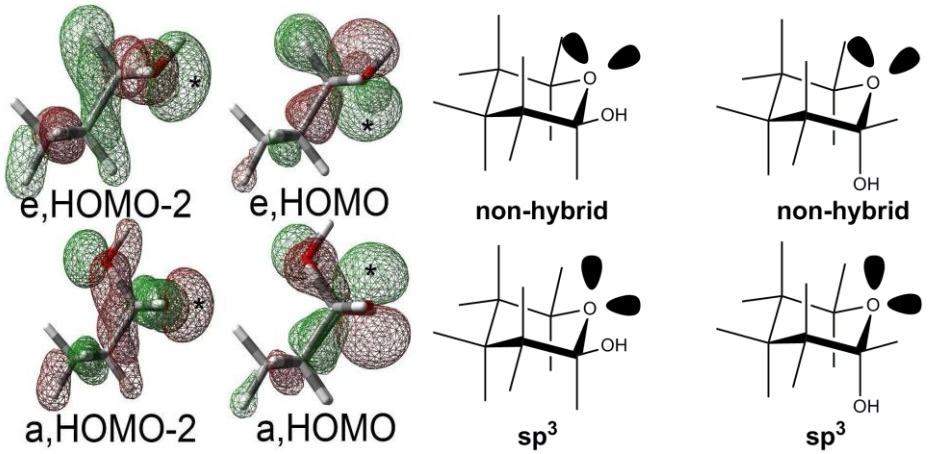



**Figure 1.4.2. Left:** computed frontier orbitals for tetrahydropyran-2-ol, showing the two lone pairs of the pyranoid oxygen atom (each lobe marked with a * symbol), at an angle of 90º to each other. **Right**: comparison of the sterics of a $sp^3$-hybridized vs a non-hybridized pyranoid oxygen atom in the models examined here; lone pairs are shown as hollow wedges at the oxygen.

Information extracted from Natural Bonding Orbital (NBO) analysis indicates that in the equatorial conformer one of the oxygen's lone pairs is 56.4% s and 43.6% p character, while the other lone pair orbital is 0.3% s and 99.7% p, meaning that only one of the lone pairs is not hybrid (Table 1.4.1).

These data then may be taken to support an $sp^2$ description of the pyranoid oxygen. Such a description is in line with observations made on the spatial orientation of the oxygen models in most of the models in.

**Table 1.4.1.** NBO Analysis for pyranoid oxygen in the equatorial conformer. Only lone pair and Rydberg orbitals are shown.

| | Equatorial Conformer | | Axial Conformer | |
|---|---|---|---|---|
| OT | Occ | % Hybridation | Occ | Coefficients/ Hybrids |
| LP(1) | -1.97 | s 56.42% p 43.58% | -1.96 | s 49.61% p 50.39% |
| LP(2) | -1.93 | s 0.36% p 99.64% | -1.90 | s 0.87% p 99.13% |
| RY*1 | -5.97e-3 | s 16.75% p 83.25% | $-5.12e^{-3}$ | s 1.26% p 98.74% |
| RY*2 | -2.40e-3 | s 22.45% p 77.55% | $-2.23e^{-3}$ | s 6.57% p 93.43% |
| RY*3 | -6.80e-3 | s 11.55% p 88.45% | -16.0e-3 | s 44.75% p 55.25% |
| RY*4 | - 4.70e-3 | s 12.73% p 87.27% | - 1.0e-3 | s 10.76% p 89.24% |

Abbreviations. OT: Orbital type; Occ: Occupancy. LP: Lonely pair orbital; RY*: Rydberg orbital.

# Chapter 2. Computational investigations on bioinorganic complexes: case studies of interactions dictated by weak/non-covalent bonding interactions

## 2.1. The dynamics of metallo-proteins. Parameterization and general considerations

In this section, we focus on the construction of AMBER [20] force field (FF) parameters for the study of certain transition-metal active sites, for use in efforts described in the following sections of this chapter.

Among the species for which parameter development is described in this chapter is bleomycin, an anticancer drug that reacts splitting DNA with its iron or cobalt metallic center in a widely debated reaction pathway. Also, hemerythrin parameter development is presented. Hemerythrin represents a challenge as a complex double iron active site with paramagnetic properties, whose make it a particular case for theoretical calculations. For the heme group, we complemented the force field parameters within our own development for those bonds near to iron. These results were employed in a series of dynamics to study the haptoglobin-hemoglobin complex, docking, myoglobin and hemoglobin dynamics.

In order to derive such parameters, first a geometrical structure of the metallo-protein/peptide was obtained from protein data bank. Geometry optimization of the truncated

models was performed using the MP2 level of theory [25-27] with the 6-311G** basis set as implemented in the Gaussian 09 package [5]. This energy-minimized model was the structure used for the rest of the calculations.

Charge derivation was performed using the RESP [28] approach. Antechamber was used for atom type assignment. Afterwards, the Leap package included in AmberTools was employed to deduce missing parameters from AMBER FF 99 SB; some parameters were copied from AMBER03 FF or GAFF. The list of missing parameters was kept for the following step. Then, van der Waals-Lennard Jones parameters were assigned by analogy. Angle and bonded parameters were calculated using a Hessian Matrix computed at the DFT level with the same functional and basis set as described above. Dihedral parameters were deduced by creating PES plots for each dihedral, optimizing again at the DFT level as described before.

RMSE obtained for bleomycin (Figure 2.1.1) is until now far from acceptable for torsional parameters (Figure 2.1.2), where a normal threshold should be under 1 Å RMSE; however, it is possible to observe the set of parameters can nevertheless locate conformations within local minimums correctly upon QM data – and, more importantly, the combination of the present parameters with QM data has led to QM/MM results that show very good RMSD values, under 0.5 Å.

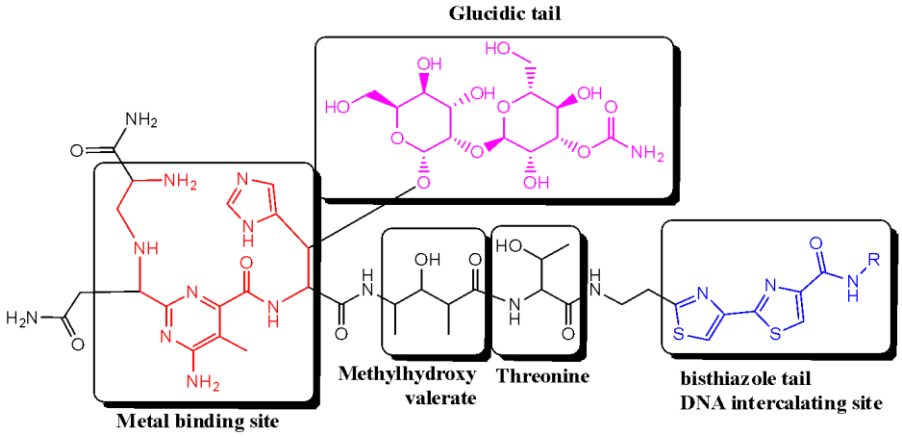



**Figure 2.1.1.** Bleomycin (BLM) structure. R = $CH_2S^+(CH_3)_2$

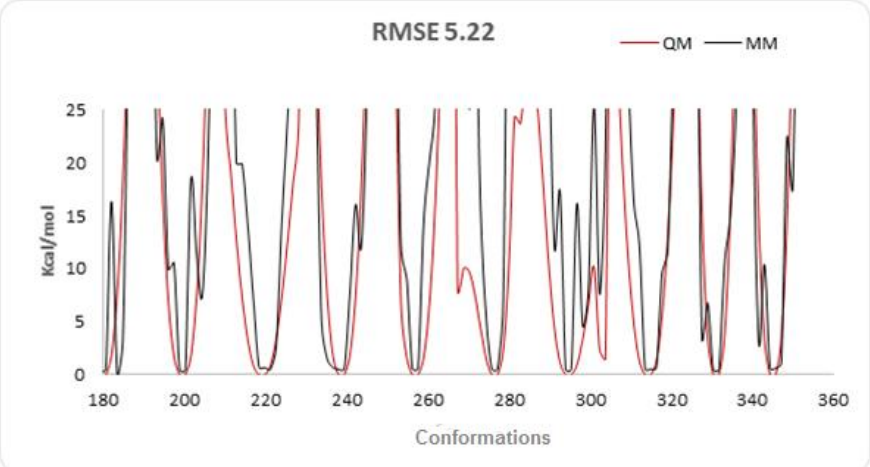



**Figure 2.1.2.** Bleomycin. Validation of dihedral parameters of bleomycin, only 360 conformers are shown in order to preserve clarity.

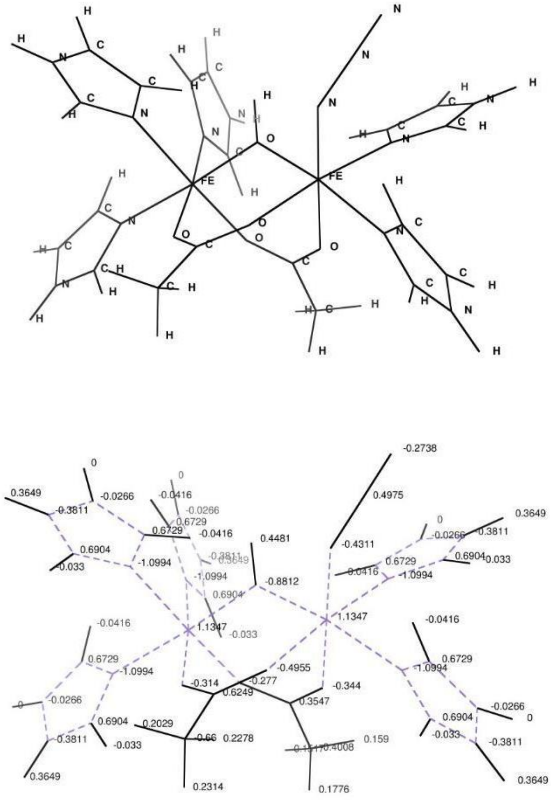


**Figure 2.1.3.** Reduced models of Hr active site. **Up-** Atomic construction. **Down-** RESP charges assigned per atom.

The following sub-chapters will illustrate applications of these parameters in MM simulations relevant for the mechanisms of action of the respective proteins/peptides/DNA.

## 2.2. The dynamics of hemoglobin-haptoglobin complexes. Relevance for oxidative stress

Haptoglobin (Hp) is known as a protein that will attach to hemoglobin (Hb) when Hb is released to the cytoplasm of the red blood cell (RBC). The Hp-Hb interaction is important since Hb contains in its active site the highly-reactive iron-containing heme group, which is a potential generator of free radicals by redox reactions, hence one which needs to be controlled [30]**.**

A theoretical model of the complex of human haptoglobin α and β hemoglobin chains [Hp- αHb; βHb] was constructed by homology and MD techniques. Initial side chains were borrowed from the pig complex, and the missing loops were minimized with the same molecular mechanics protocol as described above.

For the homology, the entry **>gi|386783|gb|AAA88080.1| haptoglobin [Homo sapiens]** was used as starting sequence. As template of reference, PDB ID structure **4F40** from Protein Data Bank, was employed.

MM/MD calculations were performed later using YASARA [32] program, with AMBER [20] 03FF with PBC, at physiological conditions in a cubic box of water was modeled 5 Å far away from all protein atoms. The system was taken to energy minima, equilibrated at NVT and NPT onditions and 5 ns time were sampled at NPT condtions, finally stopped and minimized again to obtain the final conformation.

*Sus scrofa* and *Homo sapiens* Hp sequences are 74.11 % identical, as estimated with the UCSF Chimera software package [14] using only a truncated section of the protein.

A key driving force to our study, as well as to other experimental ones, is the correct assignment of a free-radical signal seen in EPR spectra of peroxide-interacted hemoglobin in vitro but also in vivo [6].

SAS data appears to support Y145, Y35 and Y24 as likely locations of the Hb radical – arguing strongly against Y42, Y130 and Y140 (cf. Table 2.2.1).

**Table 2.2.1** Contribution to Solvent Accessible Surface area (SAS) for chosen tyrosines in hemoglobin α and β-subunits. Numbers expressed are the contribution to the total SAS in $Å^{2.}$

| Structure | Y42.α | Y140.α | Y35.β | Y145.β | Y130.β | Y24.α |
|---|---|---|---|---|---|---|
| Hb X-Ray | 23 | 17 | 21 | 59 | 1 | 9 |
| Hb Ave | 26 | 50 | 35 | 85 | 4 | 63 |
| Hp X-Ray complex | 59 | 16 | 17 | 4 | 2 | 4 |
| Hp Ave complex | 59 | 49 | 21 | 72 | 4 | 45 |

For Tyr42 the phenol-associated hydrogen bonds remain essentially unchanged upon binding to Hp, or are at least not strenthened Overall, our data would provide at most a weak support for the interpretation that Tyr42 is the location of the experimentally-observed Hb and Hb/Hp radical. None of the other Tyr residues was found of interest in changes related to hydrogen bonding.

Another argument is the change in the dihedral angle of the planarity of the Tyr ring upon going from the Hb tetramer to the Hp:Hb complex in the crystal structure (Table 2.2.2). The X-ray diffraction and MD θ values appear to favor Tyr42. However, the MD data suggests that the Tyr35 is also a good candidate, and also that taking into account the limits of the method, Tyr145 and Tyr140 cannot be ruled out. Tyr130 is supported by the crystal structure data, but definitely not by MD. Tyr42 appears a very unlikely candidate, by all accounts.

**Table 2.2.2.** Values of θ dihedral angles in tyrosine (expressed in in degrees). Only the lowest of the two complementary angles is shown in each case (one may obtain the complement by subtracting from 120°).

| Element | Y42 | Y145 | Y35 | Y140 | Y130.β | Y24.α |
|---|---|---|---|---|---|---|
| Hb X-Diff | 47 | 10 | 35 | 37 | 46 | 11 |
| Hb ave | 57 | 47 | 48 | 59 | 37 | 42 |
| Hp X-Diff | 54 | 39 | 51 | 53 | 54 | 23 |
| Hp ave | 47 | 30 | 45 | 46 | 7 | 3 |

In brief, the hydrophobicity-related parameters appear to support Y145, Y35 and Y24 as likely locations of the Hb radical – arguing strongly against Y42, Y130 and Y140. The θ values appear to favor Y42, with the MD data nevertheless also arguing in favor of Y35 and, to a lesser extent, Y145 and Y140. Y130 and Y24 appear the least likely candidates according to the θ values, especially in view of the MD data. Overall, Y35 and Y145 appear to be the two candidates with least arguments against them for satisfying the two criteria (θ values and hydrophobicity) for matching the experimentally-observed EPR radical in Hb and Hb:Hp.

### 2.3. The dynamics of hemerythrin and hemerythrin derivatives

Hemerythrin (Hr) is a homo-octameric α-helical protein with a non-heme diiron active site found in marine worms or bacteria, and has been studied as a potential alternative to hemoglobin as starting material for artificial oxygen carriers ("blood substitutes") either in PEGylated form or in glutaraldehyde-polymerized form [33-35]. The stability of these materials under physiologically relevant conditions has received little attention so far. Recently, (1-oxyl-

2,2,5,5-tetramethylpyrroline-3- methyl)methanethiosulfonate (MTSSL) has been attached at the C51 position of each monomer of Hr in order to analyze EPR signals that can render information about the stability of the Hr octameric structure [36]. Here, we present a theoretical attempt to explore the behavior of MTSSL-labeled Hr.

MM/MD simulations were performed using YASARA [32]. Monomeric and octameric Hr structures were modeled using the X-ray diffraction based structure PDB ID 1I4Y [37] from the Protein Data Bank [38] website; its geometry was energetically taken to a minimum at the semiempirical PM6 level of theory. To each one of the monomers, it was attached MTSSL in Cys51 position yielding the structures illustrated in Figure 2.3.1.

Each system was solvated with TIP3P [42] water molecules simulating physiological conditions in a 10 Å cubic box. Each system was subjected first to steepest descent minimization, equilibration at NVT and later at NPT conditions and sampling was carried out with a time step of 5 ns and with snapshots saved in the trajectory every 3 ps.

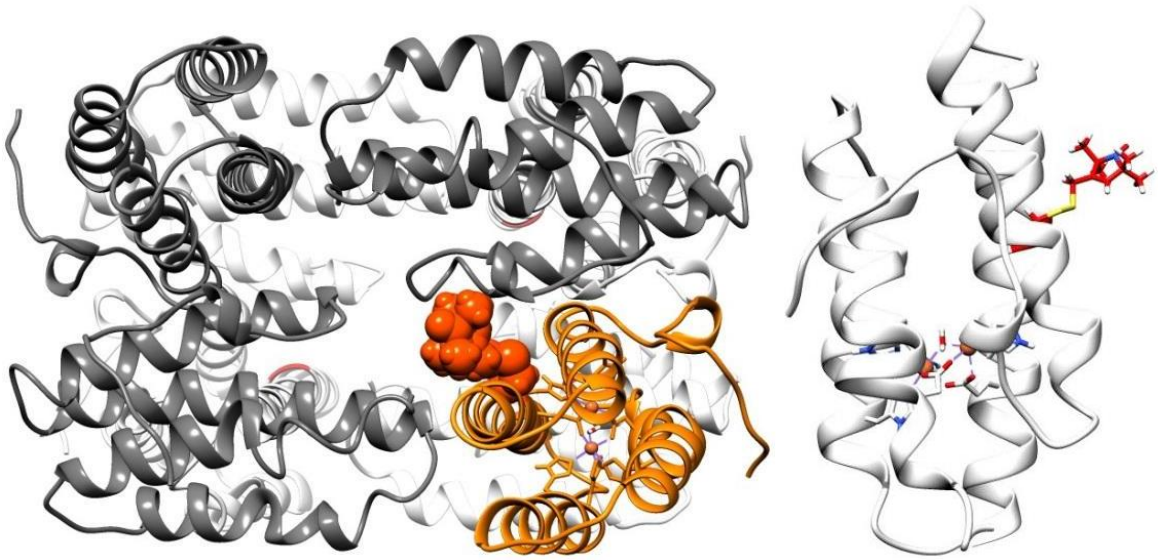


**Figure 2.3.1. Left-** Hr monomer-MTSSL (orange ribbons with MTSSL label in orange VDW surfaces) labeled at C51 position interacting with neighbor monomers (gray ribbons). Hr back monomers (not neighbors, not interacting with MTSSL label) are represented in withe ribbons for clarity. **Right-** Hr monomer (in white ribbons) MTSSL label in red ball and sticks, metallic center, Fe(Cl )-OH-Fe, represented in balls and sticks.

Thus, Figure 2.3.2 reveals that the spin-labeled cysteine residue lies sufficiently close to the inter-monomer interface in order to disrupt monomer-monomer interactions. Indeed, the distance between the α carbons of labeled Cys51 (Cyx51) in one chain and Asp23 in the neighboring chain, is on average 5 Å longer in the spin-labelled structure – an increase sufficient

to disrupt most of the typical intermolecular forces (hydrogen bonds and hydrophobic interactions especially). This difference is already seen in the geometry-optimized structures, and is maintained throughout the 5 nanoseconds of MD simulation – albeit with a slight decrease in time.

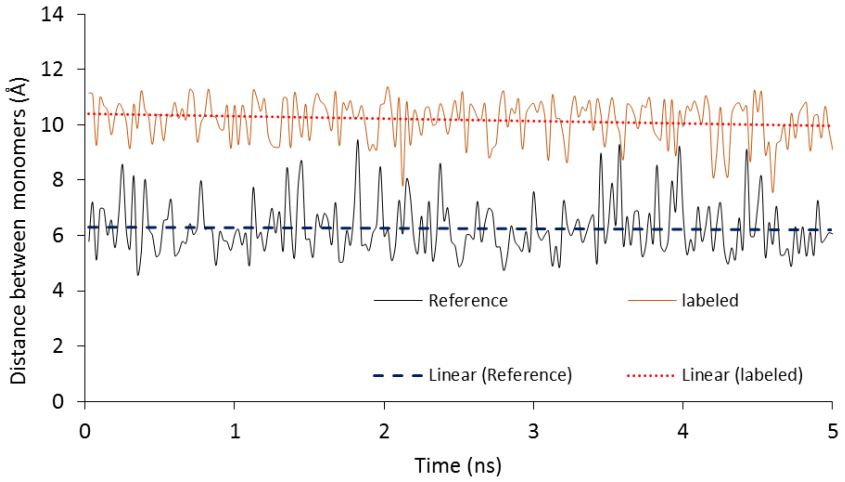



**Figure 2.3.2.** Distance between two monomers in Hr octamer; native and MTSSL-labeled Hr are shown. The distance is measured between CYX50 Cα and Asp23 Cα Chains A and B, respectively, in Hr octameric model. Black/Blue lines shows data and trend for Hr octamer model of reference. Brown/Red lines shows data and trend for Hr octamer model with MTSSL label.

The data provide support for interpreting experiments where Hr is a template or starting material for blood substitutes – hence, in multimeric form and chemically derivatized with polyethylene glycols, spin labels, or other reagents. Derivatization with a spin label at the native Cys51 position in Hr, is found to destabilize the octameric structure and to lead to a slight increase in solvent accessibility of the diiron center. Both of these consequences of the chemical derivatizations would be functional disadvantages for an oxygen-carrying protein. This puts some limitations on the relevance of the EPR spin labeling studies with Hr, and also suggests the need for a careful investigations of the effects of the PEGylation procedures in the context of blood substitute candidates.

## 2.4. Antioxidant binding to hemoglobin and myoglobin

Despite their limited structural similarity, ascorbate and urate, as the main antioxidants in plasma, appear to both display high $K_m$ values for Hb. One may then inquire on the width of the range of compounds with potential biological activity which may also display such high kinetic-derived affinities, and whether stable complexes with Hb are detectable for all of them.

Our own unpublished data on the peroxidase activity of Hb and myoglobin (Mb) towards gallic acid, rutin, caffeic acid, and 3-hydroxyflavone – viewed in parallel with ascorbate and urate reveals $K_m$ values that are again very low, even though substances of no obvious direct physiological relevance are involved. $^{1}H$ and $^{13}C$-NMR spectra reveal a selective, but multiple-site interaction between Hb and antioxidant molecules.

Globin-ascorbate docking calculation are reported in this chapter that indeed confirm multiple-site interactions, of which several have low-micromolar affinities.

PDB ID 2DXM [46] and 1MBO [47] entry codes were used respectively. Each system was subjected to MM/MD and after steepest descent energy minimization using AMBER [48] 99SB FF

Afterwards, the AutoDock Tools software package [44] was employed to assign rotatable bonds with root automatically detected as well Gasteiger charges [45] assigned for each ligand, and for the heme group, they where derived at DFT B3LYP/6-311G** level of theory fitted manually using the Resp program as implemented in AmberToolkit software.

For all docking computations presented here, the autodock4 and autogrid4 software packages [49] were used. Generalized docking was performed with a grid around the whole protein.

First, docking taking into account the full protein was performed and after finding the best docking positions, each one of the best 3 positions found were refined with a localized docking using GA algorithm.

Gallate and caffeate are perhaps surprisingly seen to have higher affinities for Hb than two endogenous antioxidants from the blood, ascorbate and urate (cf. Table 2.4.1); nevertheless, this predicted order *is* matched by experiment.

**Table 2.4.1.** Affinities of Mb and Hb to caffeate, gallate, urate and ascorbate.

| Affinity to: | Trend |
|---|---|
| Mb | Caffeate > gallate >urate>ascorbate |
| Hb | Gallate>caffeate>ascorbate>urate |
| Hb experimental | Gallic acid > caffeic acid >ascorbate>urate |

The amino acids involved in defining the ascorbate binding sites in Hb include Tyr42 at the best docking position. In fact, this tyrosine is the only one present in either of the three best binding sites in Hb, according to these calculations (cf. Figure 2.4.1, Table 2.4.2)

The caffeate binding sites in Hb contain only Tyr145, which has indeed been invoked as important in the biologically-relevant redox reactivity of Hb. In fact, our data in Chapter 2.2, on

the predicted location of an experimentally-detected tyrosyl radical in Hb, suggest that Tyr145 together with Tyr35 are the two most likely locations based on the EPR lineshape features – at odds with most previous interpretations focusing on Tyr42 (cf. Figure 2.4.1, Table 2.4.2).

As in the case of ascorbate, gallate and urate sees Tyr42 at one of the top three binding sites.

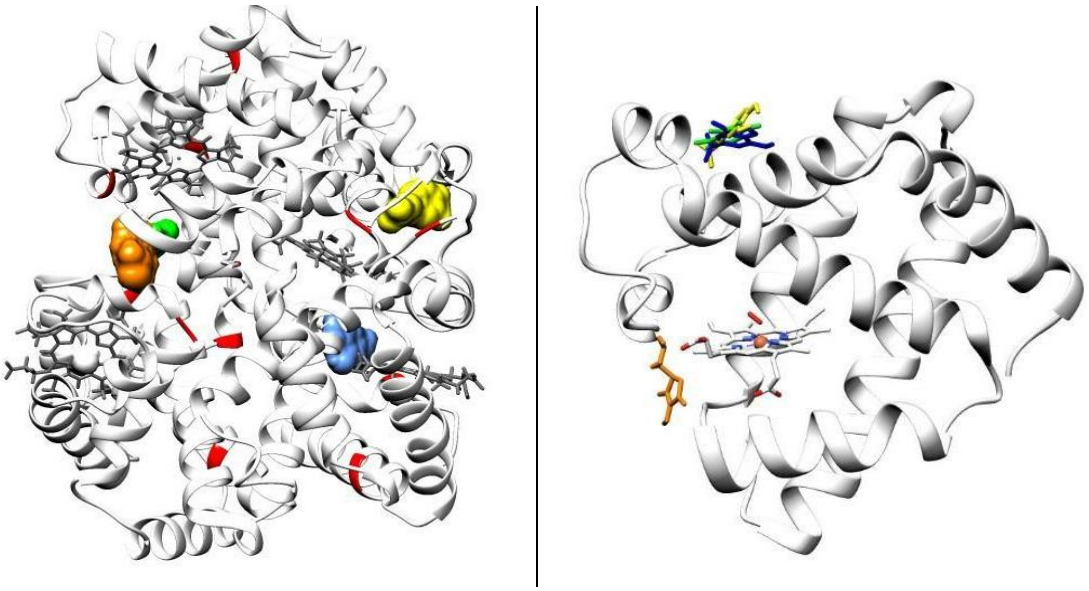


**Figure 2.4.1.** Highest-affinity binding positions for all 4 ligands. Myoglobin and Hemoglobin are represented in white ribbons, heme group in with sticks. Ligands represented in Van der Waals surfaces: ascorbate, orange. Caffeate, yellow. Gallate blue and Urate, green.

**Table 2.4.2.** Site directed docking results: 3 best positions scored based on binding energy (kcal/mol) and inhibition constant (mM).

| Myoglobin | * Conformations | | |
|---|---|---|---|
| Ascorbate | 1 | 2 | 3 |
| Binding Energy | -8.3 | -7.9 | -6.3 |
| Binding Constant | 0.79 | 2.1 | 22 |
| Caffeate | | | |
| Binding Energy | -10.8 | -10.3 | -9.2 |
| Binding Constant | 0.01 | 0.03 | 0.19 |
| Gallate | | | |
| Binding Energy | -10.6 | -8.8 | -7.5 |
| Binding Constant | 0.016 | 0.37 | 3.1 |
| Urate | | | |
| Binding Energy | -9.1 | -8.2 | -6.5 |
| Binding Constant | 0.23 | 0.99 | 17 |

| Hemoglobin | *Conformations | | |
|---|---|---|---|
| Ascorbate | 1 | 2 | 3 |
| Binding Energy | -8.6 | -6.6 | -3.2 |
| Binding Constant | 0.54 | 14 | 2000 |
| Caffeate | | | |
| Binding Energy | -10.9 | -10.4 | -6.9 |
| Binding Constant | 0.01 | 0.03 | 8.4 |
| Gallate | | | |
| Binding Energy | -11.4 | -10.1 | -9.0 |
| Binding Constant | 0.004 | 0.420 | 0.24 |
| Urate | | | |
| Binding Energy | -7.9 | -7.8 | -7.1 |
| Binding Constant | 1.8 | 1.9 | 6.4 |

Some of these binding sites match previous proposals based on experiments, with involvement of Tyr residues for electron shuttling. Despite experimental data suggesting very similar binding patterns for all small-molecule antioxidants (high affinity, multiple binding sites), differences are revealed by the docking calculations both in terms of the location of the sites and in terms of the energetics of interaction. Support is found for involvement of Tyr42 in binding of three out of the four substrates investigated in Hb (ascorbate and urate, as blood-contained relevant substrates, are included in this list), but also for Tyr145 (with urate and caffeate) and Tyr35 (with gallate).

**2.5. DNA - ligand interactions. Bleomycin.**

Bleomycin (BLM, Figure 2.1.2) is an antitumor drug used clinically. Nevertheless the mechanism of reaction, in which double strand (ds) DNA breaks down and generates two oxidized products [1-3], is not completely understood and it is a current topic of debate (2.5.1).

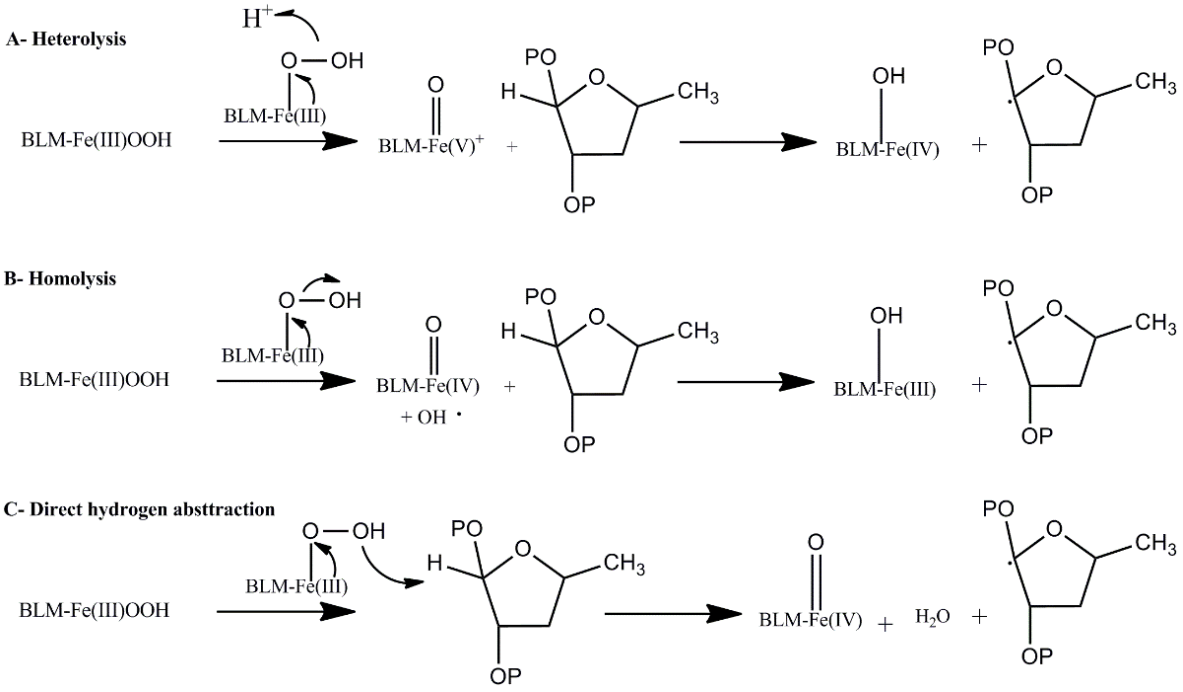



**Figure 2.5.1.** Theories for the reaction pathway between ABLM and C4' cytosine.

All MM/MD calculations were performed simulating physiological conditions into a cubic box of water 10 Å distance from the protein and implementing AMBER [48] 03 FF. YASARA calculations using ABLM with Fe(III) and Co(III) were developed with, calculations using GROMACS [53] were performed for ABLM with Fe(II), Fe(III), Co(II) and Co(III) and for semi-empirical/MM/MD calculations only Fe(III) was worked out. Parameters for ABLM were obtained as described in Section 2.1 and PM6 [11] and AMBER [48] 03 FF level of theories were employed as implemented in AmberTools software version 14 [48]. The atoms defined for the semi-empirical region are described in Figure 2.5.2; Iron charge was set at 3+. The rest of the atoms were handled at classical MM/MD.

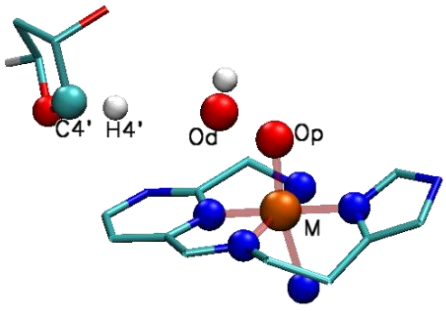



**Figure 2.5.2.** Model of the complex Bleomycin-DNA used for defining the QM region for the QM/MM calculations. Only relevant hydrogen atoms are shown. Graphic done with VMD software [58].

Results here obtained shown that the average structures allow little solvent accessibility for the peroxide moiety. However, the GROMACS and QM/MM data (but not the YASARA) also predict that the peroxide pocket does open up transiently, exposing between 3.5 $Å^2$ and 21 $Å^2$ of the peroxide moiety (with the exact values depending on methodology, metal, and oxidation state). For Fe(III) ABLM, which is the most relevant case for the issue at hand, ~6 $Å^2$ are predicted by GROMACS (Table 2.5.1). Figure 2.5.3 gives a graphical illustration of this latter case.

**Table 2.5.1.** SAS per atom over the trajectory. All values are presented in $Å^2$.

| Software | ABLM | Atom | Average Area | Standard Deviation | Maximum SAS |
|---|---|---|---|---|---|
| GROMACS | Co(II) | Op | 0.00 | 0.00 | 0.00 |
| | | Od | 0.01 | 0.07 | 1.00 |
| | Co(III) | Op | 0.00 | 0.00 | 0.00 |
| | | Od | 0.00 | 0.02 | 0.33 |
| | Fe(II) | Op | 1.85 | 0.31 | 0.33 |
| | | Od | 2.52 | 0.40 | 0.33 |
| | Fe(III) | Op | 1.81 | 0.31 | 0.00 |
| | | Od | 2.55 | 0.41 | 0.33 |
| YASARA | Co(III) | Op | 0.00 | 0.00 | 0.00 |
| | | Od | 0.00 | 0.00 | 0.00 |
| | Fe(III) | Op | 0.00 | 0.00 | 0.00 |
| | | Od | 0.00 | 0.00 | 0.00 |
| AmberTools | Fe(III) | Op | 0.00* | 0.85 | 0.00* |
| | | Od | 0.00* | 1.60 | 3.48 |

*Computed values are negative. The value shown is qualitative.

The distal oxygen of the metal-peroxo moiety is predicted to spend most of its time within 2.5 to 5 Å from the closest DNA hydrogen atom. The NMR-averaged distances for the Co structures are at the lower end of this range, and suggest close contact. Our unpublished QM calculations, as well as previously-reported ones [21,22] suggest that at such a close distance it would be unavoidable to see facile O-OH bond cleavage, concerted with insertion of the peroxide-derived OH group into the ribose.

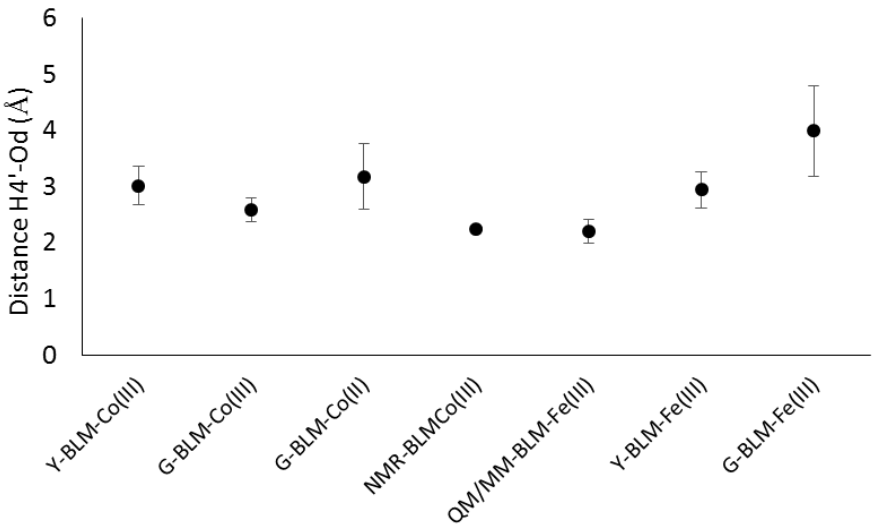



**Figure 2.5.3.** Comparison of results among different software used for classical MD. Label's notation is as follows: First letter: Y = YASARA, G= GROMACS, NMR= MD with NMR constrains as shown by Stubbe teamwork in 2001 [59] (STD is +/-0.003 Å). QMMM= QMMM dynamic calculations run with AmberTools. BLM-Fe/Co(II/III) stands for ABLM each variant.

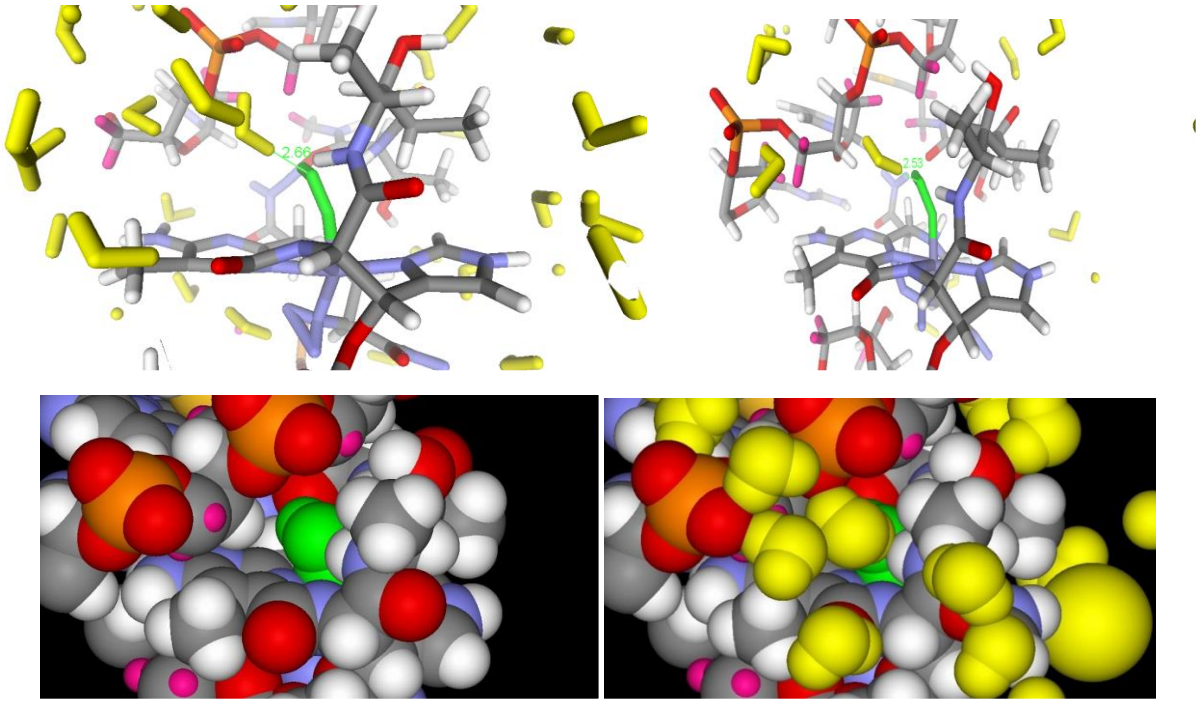


**Figure 2.5.4.** Snapshot illustrating peroxide-solvent contacts during ABLM-DNA molecular dynamics simulations for the Fe(III) and Fe(II) system in GROMACS. **Top:** Stick-representation (peroxide-water oxygen-to-oxygen distances are marked in green; Fe(III) on the left, Fe(II) on the right); **bottom:** space-fill representation *with* (right panel) and *without* (left panel) solvent depicted. Colors: solvent water – yellow; peroxide – green; all others – standard CPK.

Figure 2.5.4 illustrates that water molecules are indeed predicted by the dynamics calculations to come into direct contact with the hydroperoxo ligand. If these data are correct, then the general picture provided by computations on the ABLM-DNA interaction would be one where the oxygen-oxygen bond is initially cleaved to generate a high-valent intermediate, and this high-valent intermediate oxidizes the deoxyribose in outer-sphere manner. Whether this would occur directly, or mediated by solvent-derived hydroxyl radicals generated within the active site upon FeO-OH homolysis, remains to be explored by more accurate DFT-based QM/MM calculations.

**Selected References**


1 Zhao, Y., Yan, Z. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06 functionals and 12 other functionals. *Theor. Chem. Acc.* **119**, 525-525, doi:10.1007/s00214-007-0401-8 (2008).

2 Gaussian 09 (Gaussian, Inc., Wallingford, CT, USA, 2009).

3 Malkov, S. N., Živković, M. V., Beljanski, M. V., Hall, M. B. & Zarić, S. D. A reexamination of the propensities of amino acids towards a particular secondary structure: classification of amino acids based on their chemical structure. *J. Mol. Model.* **14**, 769-775, doi:10.1007/s00894-008-0313-0 (2008).

4 Spartan v. Spartan 5.0 (, Inc., 18401 Von Karman Avenue Suite 370, Irvine, CA 92612 U.S.A.).

5 Gaussian v. 09a (Wallingford CT, 2009).

6 Hypercube, I. *HyperChem Release 7 for Windows: Getting Started*. (2002).

7 Stewart, J. J. P. *et al. MOPAC Manual (version 3.10): A General Molecular Orbital Package -- Cray-XMP Version*. (1986).

8 Wavefunction, I. *Spartan Physical Chemistry Edition - Tutorial and Activities*. (2006).

9 Korth, M., Martin, K., Michal, P., Jan, Ř. & Pavel, H. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J. Chem. Theory Comput.* **6**, 344-352, doi:10.1021/ct900541n (2010).

10 Klamt, A. & Schuurmann, G. Cosmo - a New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J Chem Soc Perk T 2*, 799-805, doi:Doi 10.1039/P29930000799 (1993).

11 Stewart, J. J. P. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **13**, 1173-1213, doi:10.1007/s00894-007-0233-4 (2007).

12 Stewart, J. J. MOPAC: a semiempirical molecular orbital program. *J. Comput. Aided Mol. Des.* **4**, 1-105 (1990).

13 Zhao, Y. & Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **41**, 157-167, doi:10.1021/ar700111a (2008).

14 Gaussian v. 09A (2009).

15 AMBER v. 2014 (University of California, San Francisco.).

16 Ryckaert, J.-P., Jean-Paul, R., Giovanni, C. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341, doi:10.1016/0021-9991(77)90098-5 (1977).

17 Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098-3100 (1988).
18 Perdew, J. P. Erratum: Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **34**, 7406-7406, doi:10.1103/physrevb.34.7406 (1986).
19 Wavefunction, I. *Getting Started with Spartan (Second Edition)*. (2003).
20 AMBER v. 99SB (University of California, San Francisco., 2015).
21 Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688, doi:10.1002/jcc.20290 (2005).
22 Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E. & Gumbart, J. C. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comput. Chem.* **34**, 2757-2770, doi:10.1002/jcc.23422 (2013).
23 Mayne, C. G., Gumbart, J. C. & Emad, T. The Force Field Toolkit: Software for the Parameterization of Small Molecules from First Principles. *Biophys. J.* **104**, 31a, doi:10.1016/j.bpj.2012.11.209 (2013).
24 Oda, A., Yamaotsu, N. & Hirono, S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.* **26**, 818-826, doi:10.1002/jcc.20221 (2005).
25 Head-Gordon, M., Martin, H.-G., Pople, J. A. & Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **153**, 503-506, doi:10.1016/0009-2614(88)85250-3 (1988).
26 Sæbø, S., Svein, S. & Jan, A. Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Lett.* **154**, 83-89, doi:10.1016/0009-2614(89)87442-1 (1989).
27 Frisch, M. J., Martin, H.-G. & Pople, J. A. A direct MP2 gradient method. *Chem. Phys. Lett.* **166**, 275-280, doi:10.1016/0009-2614(90)80029-d (1990).
28 Cornell, W. D., Piotr, C., Bayly, C. I. & Kollmann, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620-9631, doi:10.1021/ja00074a030 (1993).
29 Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648, doi:10.1063/1.464913 (1993).
30 Chiabrando, D., Deborah, C., Francesca, V., Veronica, F. & Emanuela, T. in *Acute Phase Proteins - Regulation and Functions of Acute Phase Proteins* (2011).
31 Stoedkilde, K., Torvund-Jensen, M., Moestrup, S. K. & Andersen, C. B. F. Structure of T. brucei haptoglobin-hemoglobin receptor binding to human haptoglobin-hemoglobin. doi:10.2210/pdb4wjg/pdb (2014).
32 Krieger, E. & Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30**, 2981-2982, doi:10.1093/bioinformatics/btu426 (2014).
33 Hathazi, D. *et al.* Oxidative protection of hemoglobin and hemerythrin by cross-linking with a nonheme iron peroxidase: potentially improved oxygen carriers for use in blood substitutes. *Biomacromolecules* **15**, 1920-1927, doi:10.1021/bm5004256 (2014).
34 Fischer-Fodor, E., Mot, A., Deac, F., Arkosi, M. & Silaghi-Dumitrescu, R. Towards hemerythrin-based blood substitutes: comparative performance to hemoglobin on human leukocytes and umbilical vein endothelial cells. *J. Biosci.* **36**, 215-221 (2011).
35 Kryatov, S. V., Rybak-Akimova, E. V. & Schindler, S. Kinetics and mechanisms of formation and reactivity of non-heme iron oxygen intermediates. *Chem. Rev.* **105**, 2175-2226, doi:10.1021/cr030709z (2005).
36 Takacs, I. M., Augustin, M., Radu, S.-D. & Grigore, D. EPR investigation of libration motion of spin labeled hemerythrin. *J. Mol. Struct.* **1073**, 18-23, doi:10.1016/j.molstruc.2014.01.074 (2014).
37 Farmer, C. S. *et al.* The crystal structures of Phascolopsis gouldii wild type and L98Y methemerythrins: structural and functional alterations of the O2 binding pocket. *J. Biol. Inorg. Chem.* **6**, 418-429, doi:10.1007/s007750100218 (2001).
38 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic acids research* **28**, 235-242 (2000).

39 Frisch, Æ., Roy Dennington, I. I., Keith, T. A. & Millam, J. *GaussView Reference: Version 4*. (2007).

40 AMBER v. 2003 (2003).

41 Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577, doi:10.1063/1.470117 (1995).

42 Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **103**, 335-340, doi:10.1021/ja00392a016 (1981).

43 Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37**, 785-789 (1988).

44 Morris, G. M., Goodsell, D. S., Huey, R. & Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput. Aided Mol. Des.* **10**, 293-304 (1996).

45 Gasteiger, J., Johann, G. & Mario, M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* **19**, 3181-3184, doi:10.1016/s0040-4039(01)94977-9 (1978).

46 Chatake, T. *et al.* Protonation states of buried histidine residues in human deoxyhemoglobin revealed by neutron crystallography. *J. Am. Chem. Soc.* **129**, 14840-14841, doi:10.1021/ja0749441 (2007).

47 Phillips, S. E. V. Structure and refinement of oxymyoglobin at 1·6 Å resolution. *J. Mol. Biol.* **142**, 531-554, doi:10.1016/0022-2836(80)90262-4 (1980).

48 AMBER v. 14 (2014).

49 Morris, G. M., Ruth, H. & Olson, A. J. in *Current Protocols in Bioinformatics* (2002).

50 Li, C.-L., Chun-lian, L., Yu, S., Dong-yun, L. & Xi-cheng, W. in *Lecture Notes in Computer Science* 1159-1163 (2005).

51 Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902-3909, doi:10.1021/ja00299a024 (1985).

52 Cerutti, D. S., Duke, R. E., Darden, T. A. & Lybrand, T. P. Staggered Mesh Ewald: An extension of the Smooth Particle-Mesh Ewald method adding great versatility. *Journal of chemical theory and computation* **5**, 2322, doi:10.1021/ct9001015 (2009).

53 Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854, doi:10.1093/bioinformatics/btt055 (2013).

54 Cheatham, T. E. & Kollman, P. A. Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]2 in the presence of hexaamminecobalt(III). *Structure* **5**, 1297-1311, doi:10.1016/s0969-2126(97)00282-7 (1997).

55 Bryce, R. Manchester's Amber force field data base.

56 Shen, T., Wong, C. F. & McCammon, J. A. Atomistic Brownian dynamics simulation of peptide phosphorylation. *Journal of the American Chemical Society* **123**, 9107-9111 (2001).

57 Chaudhri, A. & Lukes, J. R. Velocity and stress autocorrelation decay in isothermal dissipative particle dynamics. *Physical review. E, Statistical, nonlinear, and soft matter physics* **81**, 026707 (2010).

58 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38, 27-38 (1996).

59 Hoehn, S. T., Junker, H. D., Bunt, R. C., Turner, C. J. & Stubbe, J. Solution structure of Co(III)-bleomycin-OOH bound to a phosphoglycolate lesion containing oligonucleotide: implications for bleomycin-induced double-strand DNA cleavage. *Biochemistry* **40**, 5894-5905 (2001).