

Electronic Structure of Proteins: Exciton Hamiltonian for a Dipeptide

Sarah Badri Jasim, and Mahdi Basim Najim

Wasit University, College of Science, Department of Chemistry

Abstract

Protein folding has attracted a great deal of interest alongside the increased amount of structural information on proteins. Different spectroscopies, for example, X-ray crystallography and Nuclear Magnetic Resonance (NMR), are often used to study protein structure. Optical spectroscopy despite limited spatial resolution, for example, circular dichroism (CD), has considerable interest, arising from the ability to derive information about secondary structures of a protein from its CD spectrum. The aim of the present work is to explore whether fully ab initio Complete Active Space Self-Consistent Field (CASSCF) calculations of the electronic excited states of dipeptides can help to construct the exciton Hamiltonian for a dipeptide. Exciton theory is an approximate approach allowing the calculation of the CD of proteins. We have treated the problem of finding the exciton Hamiltonian matrix elements as an optimization problem in which the ab initio energies represent the “true” solutions and are used to compute the “error” that will be minimized. We use a Monte Carlo algorithm to conduct optimization. We find many possible solutions, all of which upon diagonalizing the exciton Hamiltonian give energies close to the ab initio energies. Some of the off-diagonal elements are also well-defined, while other distributions of some elements are more variable. More work will be needed to reduce the number of solutions.

Key Words: Hamiltonian given energies, Hamiltonian matrix elements.

1. Introduction

1.1 . Protein Structure

Proteins are ubiquitous in all living organisms no matter whether they are simple or complex. An essential building unit to make proteins is an amino acid. There are

twenty commonly and naturally occurring standard amino acids. All of them are α -amino acids with the general structure shown in (figure 1). They differ only in terms of the R group.

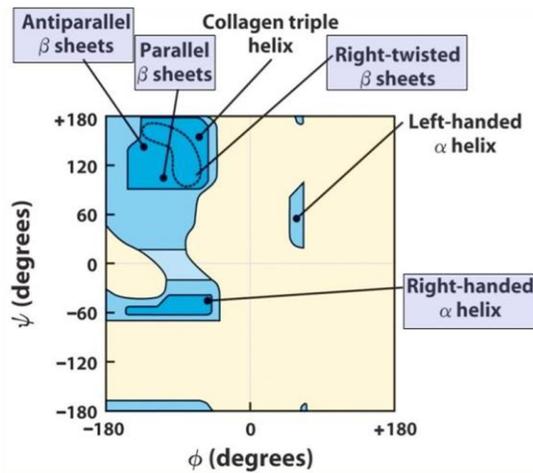


Figure 4: The Ramachandran plot showing the different allowed regions corresponding different secondary structure types: β -sheets, Collagen triple helix, L_{α} and R_{α} [2].

Ramachandran used computer models of small polypeptides to systematically vary structure with the objective of finding stable conformations. For each conformation, the structure was examined for close contact between atoms. Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii. Therefore, values of angles which cause the hard spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone. In (figure 4) the white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their van der Waals radii. These regions are sterically disallowed for all amino acids

except glycine which is unique in that it lacks a side chain. The dark blue regions correspond to conformations where there are no steric clashes. These are the allowed regions namely the α -helical and β -sheet conformations and right-handed α -helix. The light blue areas show the allowed regions if slightly shorter van der Waals radii are used in calculation. Therefore, atoms are allowed to come a little closer together. This brings out an additional region which corresponds to the left-handed α -helix [4, 5]. Protein structure, specifically protein folding, has become the intense focus for many researchers during the last several decades. Different types of important spectroscopies are used to study protein structures, for example, X-ray (crystallography), Nuclear Magnetic Resonance (NMR) and Circular Dichroism (CD) [6, 7].

1.2. Circular Dichroism of Protein

Circular dichroism spectroscopy (CD) is a technique widely applied to a large biological system [6]. It is used successfully and specifically to study chiral molecules. Chirality is an important feature of the three-dimensional structures of peptides and proteins. The chirality of the secondary structure is key for protein CD; thus, the application of this spectroscopy is possible

because proteins rotate the plane of the circularly polarized light [2]. Although CD is one of the most helpful spectroscopies for examining protein structure, other spectroscopies are used to achieve similar purposes [6, 7]. However, CD spectroscopy has advantages, such as using lower concentrations and having no size limitations in contrast to NMR. Compared with X-ray crystallography, results have been gained without any of the problems in a structure that may occur during crystallization [7].

CD is the difference in absorption of left and right circularly polarized light. For chiral molecules, the extinction coefficients for the left and right circularly polarized light are different ($\epsilon_{-l} \neq \epsilon_{-r}$). When the differential absorbance, $\Delta\epsilon$, is plotted against wavelength, the CD spectrum will be generated [6]. The CD signal is related to the scalar product of an electric transition dipole moment $\vec{\mu}$ and magnetic transition dipole moment \vec{m} . Therefore, the non-zero values of these two quantities during the excitation process of a chiral molecule leads to the CD signal [7]. The CD strength or rotational strength for an excitation is analogous to the oscillator strength of normal absorption. It is the integral of ΔA over a particular wavelength range for the transition. The rotational strength (R^{0k}) for excitation from

the ground state 0 to an electronic excited state k is the imaginary part of the product of the electric transition dipole moment, $\vec{\mu}$, and the magnetic transition dipole moment, \vec{m} , given by the Rosenfeld equation [8].

$$R^{0k} = \text{Im}(\langle\psi^0|\vec{\mu}|\psi^k\rangle\langle\psi^k|\vec{m}|\psi^0\rangle)$$

(1)

In Equation (1), ψ^0 and ψ^k represent the wave functions of the ground state and excited state respectively. By using the Rosenfeld equation, it is possible to find the rotational strengths of a particular molecule, and the CD spectrum can be calculated. The wave functions for the ground and excited states can only be determined fully ab initio for relatively small compounds. For large systems, such as proteins, there are computational challenges. However, some approximate approaches are well established, for example, matrix method [9]. CD spectroscopy on proteins provides characteristic bands in the near- and far-ultraviolet (UV) region. The far-UV region (190-250 nm) arises in large part from the secondary structure. This spectrum is affected by the backbone structure of a protein as shown in (Figure 5) [10], while the near-UV bands (>250 nm) arise from aromatic side chains.

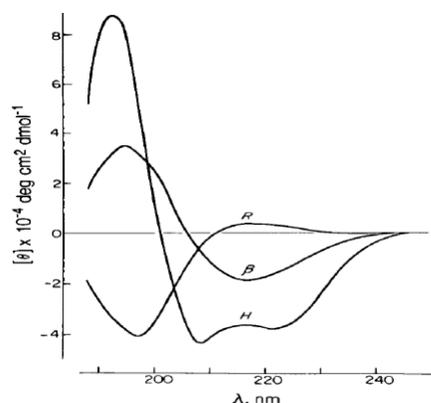


Figure 5: CD spectra of different secondary structures: The helix, β -form, and R unordered form [10].

For an α -helix, there is an intense positive band at 190 nm and a negative band at 208 nm which are from the electronic transition from the amide non-bonding π orbital (π_{nb}) to the anti-bonding π orbital (π^*), as shown in (figure 6) [11]. There is another band arising from the oxygen lone pair orbital n to π^* orbital located at around 222 nm [12-14].

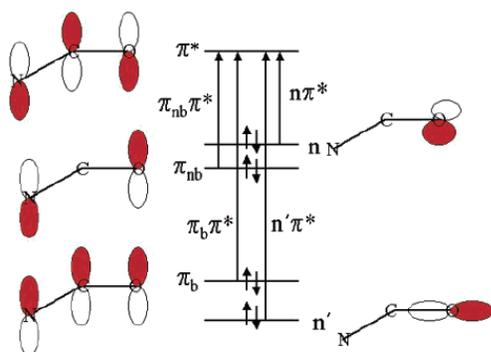


Figure 6: Molecular orbitals relevant to transitions in the far-UV [11].

The calculation of the CD spectrum is successful for small molecules. Some

development methods for computing the CD spectrum of a large molecule, such as proteins, have been reported as well. A widely used technique is the matrix method [9]. The matrix method is suitable and successful in protein CD calculations, especially for highly helical proteins [15]. This work forms a part of an ongoing effort to explore whether ab initio Complete Active Space Self-Consistent Field (CASSCF) calculations, of the electronic excited states of dipeptides, which were performed previously by Oakley et al. (2006) [16], can help inform the construction of the required exciton Hamiltonian for calculations of the electronic structure of dipeptides and, ultimately, proteins.

2. Methods

2.1. The matrix method

The matrix method, which is based on the exciton theory [17], is widely used for large molecular systems [18-23], particularly for proteins. Its use as an improved version of a simple approach was introduced by Tinoco in 1962 [24]. One begins by considering the protein as a system with M independent chromophoric groups; each group, i , is solved separately and excitations are only considered into higher electronic states of the same chromophore. The wave functions ψ^k of

excited states are constructed as a linear combination of electronic configurations in which one chromophoric group is excited and the others are in the ground state, Φ_{ia} , is the wave function of the monomer i in the excited state a , as shown in the two equations below [25].

$$\Phi_{ia} = \Phi_{i0} \cdots \Phi_{ia} \cdots \Phi_{j0} \cdots \Phi_{M0} \quad (2)$$

$$\psi^k = \sum_i^M \sum_a^{ni} c_{ia}^k \Phi_{ia} \quad (3)$$

c_{ia}^k are expansion coefficients which represent the interactions of the states. The wave function ψ^k for protein's k^{th} excited state is needed to use the Rosenfeld equation to calculate the rotational strength of a transition in CD spectrum. The wave function ψ^k of electronic excited state k of protein, and its corresponding energy, can be calculated by solving the Schrödinger equation.

$$\hat{H}\psi^k = E^k\psi^k \quad (4)$$

$$\hat{H} = \sum_{i=1}^M \hat{H}_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^M \hat{V}_{ij} \quad (5)$$

The Hamiltonian of the system with M independent chromophores, \hat{H} , is constructed as the sum of the local Hamiltonian \hat{H}_i of the independent group i plus the interactions between these two separated groups, which is denoted as intergroup potential \hat{V}_{ij} , while \hat{V}_{ii} is the interaction between states on the same chromophore. The following matrix is a

general construction of the Hamiltonian matrix for a dipeptide.

$$H = \begin{pmatrix} E_{n\pi^*}^1 & V_{n\pi^*\pi\pi^*}^{11} & V_{n\pi^*n\pi^*}^{12} & V_{n\pi^*\pi\pi^*}^{12} \\ V_{n\pi^*\pi\pi^*}^{11} & E_{\pi\pi^*}^1 & V_{n\pi^*\pi\pi^*}^{12} & V_{\pi\pi^*\pi\pi^*}^{12} \\ V_{n\pi^*n\pi^*}^{21} & V_{n\pi^*\pi\pi^*}^{21} & E_{n\pi^*}^2 & V_{n\pi^*\pi\pi^*}^{22} \\ V_{n\pi^*\pi\pi^*}^{21} & V_{\pi\pi^*\pi\pi^*}^{21} & V_{n\pi^*\pi\pi^*}^{22} & E_{\pi\pi^*}^2 \end{pmatrix}$$

The monopole-monopole approximation is used to assume that the interactions between chromophoric groups are approximated as being solely electrostatic in nature. Exchange is neglected, even though it might be important at short-range [26, 27]. These interactions are calculated as shown in (Equation 6).

$$V_{i0a;j0b} = \int \int \frac{\rho_{i0a}(r_i)\rho_{j0b}(r_j)}{4\pi\epsilon_0 r_{ij}} d\tau_i d\tau_j \quad (6)$$

where $\rho_{i0a}(r_i)$ and $\rho_{j0b}(r_j)$ are the transition electron densities of chromophores i and j , ϵ_0 is the vacuum permittivity and r_{ij} is the distance between independent chromophores. The Coulomb interactions between monopoles account for the interaction between groups.

$$V_{i0a;j0b} = \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \frac{q_s q_t}{r_{st}} \quad (7)$$

where q_s and q_t are the charges on two chromophores i and j , the monopoles, N_s and

N_t are the number of charges on the chromophores. This method involves the diagonalization of the Hamiltonian matrix, solving the eigenvalue problem. Diagonalization of the Hamiltonian matrix yields the energy of each transition (eigenvalues) and the expansion coefficients (eigenvectors). The diagonalization process is affected by a unitary matrix U .

$$U^{-1} \cdot \hat{H} \cdot U = H_{diag} \quad (8)$$

The results of the diagonalization are used to calculate the electric and magnetic transition dipole moments of the interacting system from the initial dipole moments of the single groups [28].

$$\vec{\mu}_i = \sum_a U_{ai} \vec{\mu}_a^0 \quad (9)$$

$$\vec{m}_i = \sum_a U_{ai} \vec{m}_a^0 \quad (10)$$

In the matrix method, we must consider the electrostatic potential of each electronic transition. Monopoles are fitted to reproduce the electrostatic potential. For small chromophores, the relevant electrostatic potentials can be calculated *ab initio* using, for example, the Complete Active Space Self-Consistent Field (CASSCF) method [29]. In the present work, the *ab initio*

computed energies are used to explore whether we can improve the construction of the exciton Hamiltonian matrix for a dipeptide. Oakley et al. (2006) used CASPT2 to compute *ab initio* energies for various important dipeptide geometries. For dipeptides, CASPT2 – a second order correction to the multi-configurational SCF, CASSCF energy – is the best compromise between accuracy and tractability [16]. Moreover, (table 1) presents the results of *ab initio* calculations, which were performed by Oakley et al. (2006) [16].

Table 1: CASPT2 energies for various important dipeptide geometries.

$\varphi/^\circ$	$\psi/^\circ$	$n_1 \rightarrow \pi_1^*$ (cm^{-1})	$n_2 \rightarrow \pi_2^*$ (cm^{-1})	$\pi_{nb1} \rightarrow \pi_1^*$ (cm^{-1})	$\pi_{nb2} \rightarrow \pi_2^*$ (cm^{-1})
-120	180	42587	43796	52830	53475
-60	180	42909	44764	50894	54121
-135	135	46539	45410	53959	53717
-120	120	43796	41538	53879	51862
-120	60	44200	43554	49765	52749
-120	0	43796	43796	53459	52023
-60	0	44200	43716	48717	52265
-74	-4	44442	45248	51056	55411
-48	-57	44684	44038	51701	51217
-60	-60	42909	42506	50007	53233

In addition, the *ab initio* study provides computed electric and magnetic transition dipole moments. Table 2 is an example of those transitions' dipole moments for the geometry ($\Phi=-48^\circ$ and $\Psi=-57^\circ$).

Table 2: An example of electric and magnetic transitions dipole moments given as the *xyz* components for the geometry ($\varphi=-48^\circ$ and $\psi=-57^\circ$). Taken from the CASSCF calculations of Oakley et al. (2006) [16].

Transition	Magnetic transition dipole moment (Bohr Magnetron)			Electric transition dipole moment (Debye)		
	X	Y	Z	X	Y	Z
$n\pi_1^*$	-0.018	-0.152	-0.012	0.366	-0.182	-0.767
$\pi\pi_1^*$	0.145	0.423	-3.920	0.120	0.514	-0.754
$n\pi_2^*$	0.026	0.258	0.061	-0.378	-0.318	0.739
$\pi\pi_2^*$	-0.009	2.932	0.025	0.415	-1.204	0.040

We consider for a dipeptide a 4x4 symmetric matrix, corresponding to 10 unknown variables, comprising the lower triangle of the Hamiltonian matrix. This is too large for systematic search; therefore, we use an optimization approach.

2.2. Monte Carlo optimization

The Monte Carlo optimization seeks to reduce the difference between the *ab initio* energies and our computed energies. This difference is the cost function or error which is being minimized. At the starting point for

each specific geometry, an initial guess of the exciton Hamiltonian has been generated; then it has been diagonalized. We used a standard Monte Carlo algorithm with the Metropolis criterion, Boltzmann-weighted acceptance [30].

$$p = e^{-\Delta E/kT} \quad (11)$$

where p is the probability of the acceptance, ΔE is the energy difference between the energy from the *ab* initio calculations and the energy from exciton Hamiltonian. kT is an effective temperature, set in our code so we can know the rejection ratio, for example, for the geometry ($\phi = -48^\circ$ and $\psi = -57^\circ$) the acceptance ratio is about 8 % percentage. The algorithm is iterated, with better solutions always accepted and worse ones accepted if they meet the Metropolis criterion. This will push the possible solution towards lower errors. Here is an example Monte Carlo optimized Hamiltonian for the geometry ($\phi = -48^\circ$ and $\psi = -57^\circ$) for all matrix elements.

44900	-410	-240	-590
-410	52000	500	-49
-240	500	45100	815
-590	-49	815	52400

2.3. Post-processing

To reduce the number of the solutions, an additional post-processing

criterion was computed which assesses the similarity between the *ab* initio transition dipole moments and those from the exciton method, by calculating the dot product between the *ab* initio electric transition dipole moments and the transformed electric transition dipole moments of the dimer. The transformed electric transition (μ') for a dipeptide which is shown in Equation (12) is computed from multiplication of the eigenvector by the electric transition dipole moments for the monomer.

$$\mu' = U \cdot \mu \quad (12)$$

where U is the matrix of eigenvectors, μ is the monomer transition dipole moment for each monomer from the *ab* initio dimer calculations. An example of the eigenvector matrix is shown below for the geometry ($\phi = -48^\circ$ and $\psi = -57^\circ$). Only two decimal places are shown, although it is important in calculations to use more.

0.70	0.00	-0.71	0.00
0.01	0.70	0.01	0.70
0.71	-0.01	0.70	0.01
0.00	-0.70	0.00	0.70

The mathematical formulation (or construction) of the exciton matrix method means that when the transition dipole moments are transformed (from the isolated) to the 'mixed' (or interacting) system that the

sum of dipole strengths is conserved. Possible cases are: (i) there is no interaction, and each dipeptide transition dipole is the same as in the monomer; (ii) there is an interaction such that all the intensity appears in one transition and no intensity in the other transition; (iii) intermediate cases, where both transitions have non-zero transition dipole, but one is stronger than the other.

Thus, if we aim to ‘reproduce’ the *ab initio* transition dipole moments, we need to ensure that the problem is well-formulated, in that it recognises the implicit conservation of the overall dipole strength. It is impossible, for example, to generate through mixing (as embodied in the exciton method) two dipeptide transitions whose summed dipole strength is greater than twice the dipole strength of the monomer. To ensure that this is the case, we use the *ab initio* transition dipole moments to compute a monomer transition dipole moment, following Equations 13 to 15.

$$\mu_o^2 = D_o \quad (13)$$

$$\mu_1^2 + \mu_2^2 = 2D_o \quad (14)$$

$$\mu_o^2 = \frac{\mu_1^2 + \mu_2^2}{2} \quad (15)$$

D_o represents the dipole strength and it reflects the intensity of a band in the absorption spectrum. μ_o is an electric

transition dipole moment for a monomer. Similarly, we generated magnetic transition dipole moments for a monomer.

Table 3: An example of generated electric μ and magnetic m transition dipole moments for the geometry ($\varphi = -48^\circ$ and $\psi = -57^\circ$).

Magnetic transition dipole moment (Bohr Magnetron)				Electric transition dipole moment (Debye)		
Transition	X	Y	Z	X	Y	Z
$\pi\pi_1^*$	-0.031	-0.301	-0.052	0.524	0.103	-1.065
$\pi\pi_1^*$	0.096	2.368	-2.759	0.379	-0.491	-0.509
$\pi\pi_2^*$	0.003	0.032	0.102	-0.012	-0.348	-0.013
$\pi\pi_2^*$	-0.108	1.778	2.782	0.211	-1.214	0.557

We used these generated electric and magnetic transitions dipoles moments to calculate $\cos \theta$ for each geometry. θ is the angle between the two electric transition dipole moments vectors. The first is the *ab initio* dipoles and the second is the transformed dipoles. There is a particular focus on the two transitions $\pi\pi_1^*$ and $\pi\pi_2^*$. We calculated the total \cos from the sum of $\cos \theta$ for each transition; we are considering two pairs of vectors. Therefore, if both pairs of vectors are similar then both cosines will be 1.0 and the sum of them will be close to 2.0. This is considered as the quality index. We used Equation (16) to calculate the angle (θ):

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} \quad (16)$$

2.4. DichroCalc

The matrix method is implemented in our in-house software DichroCalc [31]. The first step is to convert a PDB (protein data bank) file to a DichroCalc input file, which contains all the parameters for the calculation as well as the structural information of the protein, which atoms make up the chromophores, including all atom xyz coordinates in Ångstrom.

DichroCalc constructs the exciton Hamiltonian matrix based on the chromophores and their location. Diagonalization of the matrix is the next step. The parameters to construct the local Hamiltonian are from ab initio calculations of small molecules, which represent chromophores in protein. *N*-methyl acetamide (NMA) is used for studying the peptide bond [32, 33]. We used the data in the parameter file of the DiChrocalc, which gives reference coordinates of the peptide chromophore, and the electric transition dipole moment (μ) related to that orientation.

3. Results and discussion

3.1. Convergence of Monte Carlo

Monte Carlo optimization uses the Metropolis criterion Boltzmann-weighted acceptance [30]. The typical acceptance ratio is related to kT set in our code in order to be

appropriate. If it is too low, then the algorithm will tend to stay stuck in local minima; if it is too high, then the algorithm will not converge effectively.

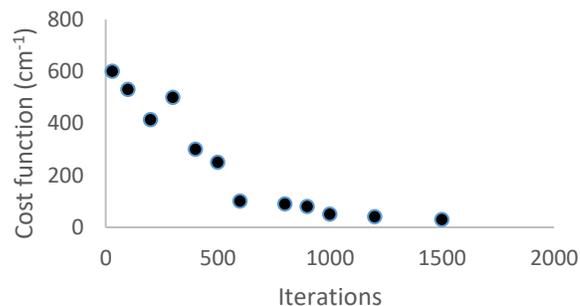


Figure 7: Convergence of the Monte Carlo search.

Figure 7 shows an effective optimization. At the earlier iterations of the algorithm the cost function was high, but by repeating the steps, (several hundred usually), it became lower and sufficient.

3.2. Distribution of Hamiltonian matrix elements

The following three figures 8, 9 and 10, illustrate the distributions for specific elements in the matrix, corresponding to a particular geometry of a peptide. In this case the geometry is ($\varphi = -135^\circ$, $\psi = 135^\circ$). All these distributions are computed from the final solutions of multiple independent Monte Carlo optimizations starting with different random numbers.

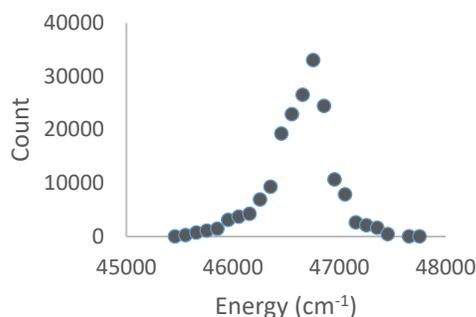


Figure 8: The distribution of the diagonal elements $E_{n\pi^*}^1$ for the dipeptide geometry ($\varphi = -135^\circ$, $\psi = 135^\circ$).

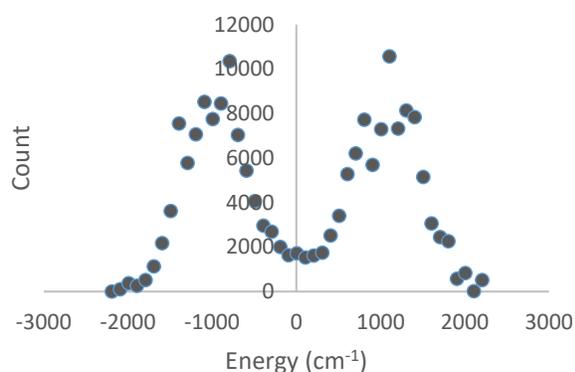


Figure 9: Bi-modal distribution of the off-diagonal elements $V_{n\pi^*\pi\pi^*}^{11}$ for the dipeptide geometry ($\varphi = -135^\circ$, $\psi = 135^\circ$).

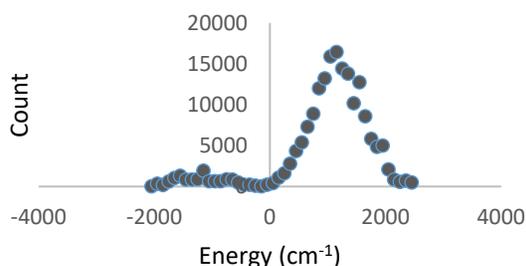


Figure 10: Mono-modal distribution of the off-diagonal elements $V_{n\pi^*\pi\pi^*}^{22}$ for the dipeptide geometry ($\varphi = -135^\circ$, $\psi = 135^\circ$).

In (figure 8) which represents one of the diagonal elements we found the distribution

is well-defined. In general, this is the case for all other diagonal elements of different matrices for different dipeptide geometries. Figure 9 exhibits a more variable distribution for one of the off-diagonal elements of the exciton Hamiltonian matrix for a dipeptide. There are many other similar distributions which are bi-modal as shown in (figure 9) or have one broad peak. In such a case, we cannot predict a specific value for the matrix element. Therefore, these elements are still undefined, and they need more optimization or maybe a new criterion to help us narrow the number of solutions.

On the other hand, in (figure 10) which shows the distribution of a different off-diagonal elements, we found a well-defined distribution with one clear peak similar, to some extent, to the diagonal elements' distributions. As a result, the generated off-diagonal elements divided into two different parts. One is consistent and similar solutions for multiple Monte Carlo runs, the other is more variable. After all, these figures show that we can find many different exciton matrices, using the Monte Carlo search, that diagonalise transition energies close to the ab initio energies. A summary of our matrix elements for all different peptides geometries is illustrated in (table 4).

Table 4: A summary of the mean (\bar{E}) of the matrix elements energies (cm^{-1}) with the standard deviation (σ) for different dipeptides geometries.

Matrix elements		$E_{n\pi^*}^1$	$V_{n\pi^*\pi\pi^*}^{11}$	$E_{\pi\pi^*}^1$	$V_{n\pi^*\pi\pi^*}^{21}$	$V_{n\pi^*\pi\pi^*}^{21}$	$E_{n\pi^*}^2$	$V_{n\pi^*\pi\pi^*}^{21}$	$V_{\pi\pi^*\pi\pi^*}^{21}$	$V_{n\pi^*\pi\pi^*}^{22}$	$E_{\pi\pi^*}^2$
$(\varphi/^\circ)$	$(\psi/^\circ)$	$\bar{E} (\sigma)$ cm^{-1}									
-120	180	43700 (370)	-117 (1500)	52200 (200)	-5 (640)	-200 (1500)	44100 (320)	-960 (1400)	-225 (460)	1200 (1370)	52700 (230)
-60	180	44200 (370)	450 (820)	51100 (300)	150 (750)	-650 (840)	44500 (300)	1100 (900)	-1000 (370)	-120 (1300)	53000 (340)
-120	120	43900 (560)	780 (490)	52900 (490)	19 (920)	2500 (600)	43500 (630)	1200 (1000)	-160 (400)	-850 (840)	51300 (290)
-120	60	44500 (360)	-36 (350)	50500 (670)	25 (200)	560 (480)	44300 (430)	-840 (780)	-390 (520)	970 (680)	51900 (240)
-120	0	44300 (360)	-32 (1300)	52800 (260)	-49 (160)	1270 (560)	44300 (300)	620 (960)	-260 (180)	400 (1100)	51600 (150)
-60	0	44500 (320)	100 (400)	51000 (420)	80 (180)	580 (350)	44400 (380)	360 (960)	-160 (190)	1000 (330)	51400 (230)
-60	-60	44500 (730)	210 (850)	50700 (830)	-105 (290)	-820 (580)	44500 (820)	-540 (1090)	-290 (430)	230 (1100)	52200 (320)
-74	-4	44900 (200)	-410 (320)	52000 (360)	-240 (350)	500 (266)	45100 (100)	-590 (580)	-49 (130)	810 (500)	52400 (440)
-48	-57	45000 (190)	-32 (380)	51600 (290)	-27 (280)	-280 (440)	45300 (150)	-167 (590)	-200 (310)	-39 (840)	52600 (730)
-135	135	46600 (290)	67 (1000)	53200 (370)	180 (480)	1400 (450)	46100 (220)	-18 (1300)	-370 (210)	1000 (740)	53000 (350)

4. Conclusion

Our work with the exciton Hamiltonian matrix for a dipeptide is far from over. There are still some off-diagonal elements which require more optimization or may be addition of new criterion to help us narrow the possible solutions to our matrices. However, it is possible to find many exciton matrices, using the Monte Carlo search, that diagonalise to give transition energies close to the *ab initio* energies. Evidently, the present work shows that we can achieve some solutions for the diagonal elements very close to what they should look like. In addition, other off-diagonal matrix elements are well-defined. The transition moments provide some useful additional information. It is an

additional post processing criterion to assess the similarity between the *ab initio* transition dipole moments and those from the exciton method, by calculating the dot product between the *ab initio* electric transition dipole moments and the transformed electric transition dipole moments of the dimer. More work will be needed to reduce the number of the solutions to construct our exciton Hamiltonian matrix for a dipeptide.

5. Future work

Final construction of the exciton Hamiltonian matrix for a dipeptide is not achieved in the present work. Perhaps, this could be achieved by some attempts to apply

or add some additional criterion in between or after the initial guess of our matrix, for example, the sign or magnitude of elements of Hamiltonian matrix. We suppose that will help us to identify physically meaningful solutions. When this planned work is finished, we plan to study the overlap-dependent term, comprising exchange and penetration, is the prime candidate for improving the exciton calculations beyond current approximations. Whilst exchange and Coulomb interactions are usually opposite in sign, penetration effects can reinforce the coulombic interaction. Exchange and penetration can become significant at short separations. A simple exponential function, of the form $A \exp(-2\alpha R)$, where A and α are constant has been shown to model this interaction acceptably. We propose to fit the values of A and α based on the discrepancies between the exciton calculations and the fully ab initio CASSCF calculations. The off-diagonal elements of the exciton Hamiltonian will be updated accordingly.

6. References

1. Darby N. J., and Creighton T. E., (1993), Protein structure, IRL Press at Oxford University Press Oxford, UK.
2. Langel U., Cravatt B. F., Graslund A., Von Heijne N., Zorko M., Land T. and Niessen S., (2009). Introduction to peptides and proteins. CRC Press.
3. Ramachandran G. N., Ramakrishnan C. and Sasisekharan V., (1963). Stereochemistry of polypeptide chain configurations. *Journal of molecular biology*. 7, 95-99.
4. Kleywegt G. J., and Jones T. A., (1996). Phi/psi-chology: Ramachandran revisited. *Structure* (London, England : 1993). 4, 12, 1395-1400.
5. Ho B. K., Thomas A., and Brasseur R., (2003). Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. *Protein science : a publication of the Protein Society*. 12, 11, 2508-2522.
6. Nordén B., Rodger A. and Dafforn T., (2010), Linear dichroism and circular dichroism, The Royal Society of Chemistry.
7. Berova N. and Nakanishi K., (2000), Circular dichroism: principles and applications, John Wiley and Sons.
8. Rosenfeld, L. (1929). Quantenmechanische Theorie der natürlichen optischen Aktivität von Flüssigkeiten und Gasen. *Zeitschrift für Physik*. 52, 161-174.

9. Mark T. Oakley and Jonathan D. Hirst., (2006). Charge-Transfer Transitions in Protein Circular Dichroism Calculations, *Journal of the American Chemical Society*. 128, 38, 12414-12415.
10. Yang J. T., Wu C. S., and Martinez H. M., (1986). Calculation of protein conformation from circular dichroism. *Methods in enzymology*. 130, 208-269.
11. Besley N. A., Oakley M. T., Cowan A. J., and Hirst, J. D., (2004). A sequential molecular mechanics/quantum mechanics study of the electronic spectra of amides. *Journal of the American Chemical Society*. 126, 41, 13502-13511.
12. Bulheller B. M., Miles A. J., Wallace B. A., and Hirst, J. D. (2008). Charge-transfer transitions in the vacuum-ultraviolet of protein circular dichroism spectra. *The journal of physical chemistry. B*, 112, 6, 1866-1874.
13. Kaminsky J., Kubelka J., and Bour P., (2011). Theoretical modeling of peptide α -helical circular dichroism in aqueous solution. *The journal of physical chemistry. A*, 115, 9, 1734-1742.
14. Jiang J., and Mukamel S., (2011). Two-dimensional near-ultraviolet spectroscopy of aromatic residues in amyloid fibrils: a first principles study. *Physical chemistry chemical physics : PCCP*. 13, 6, 2394-2400.
15. Robert W. Woody, and Ignacio Tinoco., (1967). Optical Rotation of Oriented Helices. III. Calculation of the Rotatory Dispersion and Circular Dichroism of the Alpha- and 310-Helix. *The Journal of Chemical Physics*. 46, 12, 4927-4945.
16. Oakley M.T., Bulheller B.M., and Hirst J. D., (2006). First-principles calculations of protein circular dichroism in the far-ultraviolet and beyond†. *Chirality*. 18, 340-347.
17. Davydov A., (1971), *Theory of molecular excitations*. Plenum press. New York, 7, 387-394.
18. Matile S., Berova N., Nakanishi K., Fleischhauer J. and. Woody R. W., (1996). Structural Studies by Exciton Coupled Circular Dichroism over a Large Distance: Porphyrin Derivatives of Steroids, Dimeric Steroids, and Brevetoxin B₁. *Journal of the American Chemical Society*. 118, 22, 5198-5206.
19. Gawroński J., Brzostowska M., Kacprzak K., Kołbon H., and Skowronek, P., (2000), Chirality of aromatic bis-imides from their circular dichroism spectra. *Chirality*. 12, 263-268.

20. Oakley M. T., Guichard G. and Hirst J. D., (2007), Calculations on the Electronic Excited States of Ureas and Oligoureas. *The Journal of Physical Chemistry B.* 111, 12, 3274-3279.
21. Nachtigallová D., Hobza P., and Ritze H. H., (2008). Electronic splitting in the excited states of DNA base homodimers and-trimers: an evaluation of short-range and Coulombic interactions. *Physical Chemistry Chemical Physics.* 10, 37, 5689-5697.
22. Tsubaki K., (2007). Synthesis and properties of the chiral oligonaphthalenes. *Organic and Biomolecular Chemistry.* 5, 14, 2179-2188.
23. Vlaming S. M., Augulis R., Stuart M. C. A., Knoester J., and Van Loosdrecht P. H. M., (2009). Exciton spectra and the microscopic structure of self-assembled porphyrin nanotubes. *Journal of physical chemistry B.* 113, 8, 2273-2283.
24. Tinoco, I., (1996). Nucleic acid structures, energetics, and dynamics. *The Journal of Physical Chemistry.* 100, 31, 13311-13322.
25. Hirst J. D., (1998). Improving protein circular dichroism calculations in the far-ultraviolet through reparametrizing the amide chromophore. *The Journal of Chemical Physics.* 109, 2, 782-788.
26. Seibt J., Lohr A., Würthner F., and Engel V., (2007). Circular dichroism and absorption spectroscopy of merocyanine dimer aggregates: molecular properties and exciton transfer dynamics from time-dependent quantum calculations. *Physical Chemistry Chemical Physics.* 9, 47, 6214-6218.
27. Kenny E. P., and Kassal I., (2016). Benchmarking calculations of excitonic couplings between bacteriochlorophylls. *The Journal of Physical Chemistry B.* 120, 1, 25-32.
28. Bayley P. M., Nielsen E. B., and Schellman J. A., (1969). Rotatory properties of molecules containing two peptide groups: theory. *The Journal of Physical Chemistry.* 73, 1, 228-243.
29. Roos B. O., (1992). *Lecture notes in quantum chemistry*, Springer-Verlag Berlin.
30. Goldberg, D. E., and Deb, K., (1991). A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of genetic algorithms.* 1, 69-93, Elsevier.
31. Benjamin M. Bulheller, and Jonathan D. Hirst, (2009). DichroCalc—circular and

linear dichroism online, Bioinformatics.
25, 4, 539-540.

32. Besley N. A., and Hirst J. D., (1999).
Theoretical studies toward quantitative
protein circular dichroism calculations.
Journal of the American Chemical
Society. 121, 41, 9636-9644.
33. Hayashi T., Zhuang W., and Mukamel S.,
(2005). Electrostatic DFT map for the
complete vibrational amide band of
NMA. The Journal of Physical
Chemistry A. 109, 43, 9747-9759.