Chapter 7

Investigation of Electronic Properties of Model Ternary Polypeptide Chains
In our previous chapter (Chapter 6), we performed theoretical investigations on model binary polypeptide chains made of glycine and alanine units, to study their electronic properties with the help of genetic algorithm. In the present study, we extend our calculations to model ternary polypeptides comprising of glycine, alanine and serine as amino acid residues. Their chemical structures are shown in Fig. 7.1 a-c. In our calculations, we have used the *ab initio* Hartree-Fock crystal orbital results obtained using Clementi’s minimal basis (MB) set [1-4]. The effect of use of a double zeta (DZ) basis set on the electronic structure and conduction properties of the model aperiodic polypeptide chains has been investigated along with studies including, effect of change in secondary structure (viz. α-helix and anti-parallel β-pleated sheet structures) and effect of hydration. We have also performed calculations to study the effect of H\(^+\) and Li\(^+\) ion-binding to the peptide group, in the presence of negative point charges. A comparison is also drawn between the results obtained for the two cationic adducts. The protonated adduct is expected to withdraw more negative charge from the polypeptide chain due to smaller size of H\(^+\). We have also investigated the change brought about in electronic properties of the polypeptide, if one of the amino acids, alanine, is replaced by leucine (Fig. 7.1 (d)).

7.1. Electronic properties and DOS of model ternary polypeptide chains

In the investigations on model ternary polypeptide chains, \(A_xB_yC_z\) comprising of the three homopolypeptides, \((A)_x\) (polyglycine), \((B)_y\) (polyalanine) and \((C)_z\) (polyserine), using the GA technique, our purpose is to find their relative concentrations such that the resulting polypeptide chain formed has a minimum band gap value and maximum
electronic delocalization. As an extension to the studies carried out in chapter 6, on binary polypeptides, we now use the condition, $x + y + z = 100$. Hence the values of $x$, $y$, $z$ obtained in the solution after convergence of the GA would be the optimal percentages of glycine, alanine and serine respectively, in the ternary model polypeptide.

From *ab initio* treatment we obtain the band structure details for both valence as well as conduction bands (VB and CB respectively) of the homopolypeptides (polyglycine, polyalanine, polyserine and polyleucine). We then consider a polypeptide chain of a defined sequence of such homopolypeptides. From these band structure results (Fig. 7.2 a-g), we compute the $\alpha$ and $\beta$ values (Hückel parameters) for both VB and CB of each homopolypeptide and thereafter, use these parameters to construct a Hückel determinant for the polypeptide chain according to the sequence of the units in the chain.

To determine DOS, an energy grid of 0.001 eV has been consistently used in the calculations. The polypeptide systems (comprising of gly, ala and ser) which we have investigated in this study include, (i) $\beta$-chain conformation using MB set, (ii) $\alpha$-helix using MB set, (iii) $\beta$-chain using a DZ basis set, (iv) $\beta$-chain using MB set in the effective field of surrounding water molecules, (v) $\beta$-chain using MB set forming an adduct with $H^+$ ion, and (vi) $\beta$-chain using MB set forming an adduct with $Li^+$ ion. In the last two systems (v) and (vi) mentioned above, we also take into account a negative point charge in order to maintain the electro neutrality of the system. Thereafter, we have replaced alanine units by leucine units and investigated the electronic properties of the gly-leu-ser polypeptide for the $\beta$-chain form.
The values of IP, EA, $E_g$ and IPN for the optimum percentage composition obtained from GA in each system are listed in Table 7.1. Since there is random sequencing of amino acids (repeat units) in the polypeptide chain, the respective environments of gly, ala/leu and ser units keep on changing and therefore, their energy positions (peaks) are scattered over a wide range, as is evident from the DOS plots. The details of the effects of various factors which affect the electronic properties of the polypeptide chains have been discussed below.

7.1.1. Effect of conformation

To investigate the effect of change in conformation of polypeptide chain on the electronic DOS, the DOS for both β-helix (Fig. 7.3 (a)) as well as, α-helix (Fig. 7.3 (b)) forms of the polypeptide chain have been calculated using Clementi’s MB set. Band gap value (Table 7.1) in α-helix structure is found to be slightly lower (13.932 eV) than in the case of β-pleated chains (14.696 eV) because of comparatively lower value of EA, while the IP remains almost similar. The same trend is evident from the DOS curves as well. A comparison of the optimum solutions obtained in the two cases shows that while $A_6B_1C_{93}$ is the ideal percentage composition of the polypeptide in case of anti-parallel β-pleated sheet structure, serine being the highest in amount in the chain, $A_1B_{85}C_{14}$ is optimum composition in case of α-helix with alanine being present in the maximum amount.

7.1.2. Effect of basis set

A comparison of the results obtained (Table 7.1) for the β-pleated sheets using Clementi’s DZ basis set with those obtained using MB set shows that the fundamental
band gap decreases (from 14.696 to 13.527 eV) with the use of a better basis set. The decrease in the fundamental band gap value with the use of a better basis set (DZ) is also depicted from the fact that the VB shifts up significantly (Fig. 7.3 (c)) leading to a decrease in the band gap (as compared to MB set (Fig. 7.3 (a))). A lower value of IPN (0.007607) in case of DZ basis set also shows a higher level of delocalization in the conjugated carbon chain as a result of an extended basis set.

7.1.3. Effect of hydration

We have shown herein, the effects of the solvation shell on the band structures of the polypeptides in terms of band shifts as well as change in the band widths and the subsequent effect on the DOS plots. The simulations on biopolymers, in the presence of solvent molecules, as reported by Liegener et al. [3] have been used as input for the calculations. They performed Monte-Carlo simulation on the polymer surrounded with water molecules with Metropolis sampling scheme at 300K and a water density of 1.0 g/cc assuming Periodic Boundary Condition. The band structures of the copolymer with water interaction (W) are depicted in Fig. 7.2 (d). From the DOS curves (Fig. 7.3 (d)) we can easily infer that the bands get shifted by the water point charges. It can be seen that hydration has a much stronger influence on the width of the DOS curves. In Fig. 7.3 (a), the VB ranges from around -12.5 to -11 eV and CB from 3.5 to 5 eV while in Fig. 7.3 (d) the VB spreads from almost -14.5 to -11 eV and CB from 0.5 to 5 eV. These trends in the band widths indicate that the extent of delocalization in the chain increases due to hydration.

Moreover, the GA run suggests that the optimum polypeptide chain should have maximum percentage of alanine ($A_1B_{91}C_8$) rather than serine ($A_6B_1C_{93}$ without water).
A large downward shift can be observed in the CB in the presence of surrounding water molecules. This is obviously due to the reason that the different chemical nature of the side chain (-H in glycine, -CH$_3$ in alanine and –CH$_2$OH in serine) leads to an altogether different arrangement of the water molecules in their vicinity, which has different effects on the corresponding energy bands. As a result the $E_g$ also decreases drastically and comes down from 14.696 to 11.952 eV. Thus, we can conclude that hydration also strongly influences the intrinsic conduction properties of aperiodic polypeptide chains.

### 7.1.4. Effect of metal ion binding

Protons and metal ions play a very important role in the biophysical chemistry of proteins and peptides. They are also involved in a large variety of biological processes such as gene expression, catalytic activity etc. As an input, we have used the *ab initio* results obtained by Bakhshi *et al.* [4] which involve crystal orbital calculations to study the effect of H$^+$ and Li$^+$ binding to the peptide group. They have investigated the effect of ion binding in the presence of anions represented by point charges on the band structure of the homopolypeptides poly(gly), poly(ala) and poly(ser). Their band alignments are shown in Fig. 7.2 (e) and Fig. 7.2 (f). Although the percentage composition of the optimum solution (A$_3$B$_1$C$_{94}$) is similar to what it is in the $\beta$-pleated polypeptide (A$_6$B$_1$C$_{93}$ with MB set) there is ~6 eV change in IP value. Both the VB and CB in the H$^+$ adduct are lowered by 6 to 8 eV as can be seen in Table 7.1. The DOS curves (Fig. 7.3 (e)) show that the VB becomes significantly broader. The result of this change in band width and positioning of the bands is that the $E_g$ value decreases by almost 3 units.
Adduct formed when Li\(^+\) ion is bound to the peptide group also produces qualitatively similar results with the optimum percentage composition being A\(_4\)B\(_1\)C\(_95\). A lowering in energy of about 4.5 to 5.5 eV is seen in both the valence and conduction bands (Fig. 7.3 (f)). A critical comparison of the two cationic adducts shows that the adduct of H\(^+\) + Θ with the polypeptide chain has very high EA (5.048 eV, in comparison to 1.638 eV of Li\(^+\) + Θ adduct) which is possibly due to the large electron affinity of H\(^+\) ion. Hence, it is expected that the H\(^+\) + Θ adduct withdraws much more negative charge from the peptide group than the Li\(^+\) + Θ adduct because of the smaller size of H\(^+\) ion, as a result of which it acts as a strong electron acceptor than Li\(^+\). Our results are in very good agreement with the \textit{ab initio} studies carried out earlier [5] on the same amino acid residues.

### 7.1.5. Effect of replacing alanine with leucine

We intend to investigate the variation in electronic properties of the β-pleated polypeptide chains on replacement of alanine units with leucine units. The band structure details of the polypeptide of gly-leu-ser are depicted in Fig. 7.2 (g). A comparison of the results obtained (Table 7.1) for gly-ala-ser chains with those of gly-leu-ser chains shows that the fundamental band gap comes down drastically by almost 6 eV (from 14.696 to 8.743 eV) due to presence of leucine residues in the polypeptide. We come to a conclusion that this happens due to a substantial upward shift in VB, as can be observed from the IP values. The IP for gly-ala-ser chains is 11.101 eV while for gly-leu-ser chains it is 5.117 eV. The EA remains more or less same in both cases and hence does not contribute much in the lowering of E\(_g\) value. The above trends in electronic properties are also confirmed from the DOS plots.
We also observe a lowering in the IPN for the gly-leu-ser polypeptide. IPN for gly-ala-ser chains is 0.008018 while for gly-leu-ser chains is 0.005294. A lower value of IPN suggests that there is a higher level of delocalization in the conjugated polypeptide chain containing leucine units. In addition, according to the solution obtained from GA i.e., $A_2D_9C_1$ (where A, D, and C denote gly, leu, and ser, respectively) we conclude that the optimum polypeptide chain is the one which contains 97% units of leucine. In other words, leucine units should be present in maximum amount in order to generate a polypeptide which possesses minimum band gap and maximum electronic delocalization in the chain.

### 7.2. Conclusions

In this study we have investigated the effect of change in secondary structure, change of basis set, effect of hydration/solvation shell as well as the effects of $H^+$ and $Li^+$ binding to the peptide group (in the presence of anions represented by point charges, $\Theta$) on the electronic structure and conduction properties of aperiodic poly(gly, ala, ser) chains. We have also investigated the effect of replacing alanine residue in the polypeptide chain with leucine residue.

It was observed that the band gap value decreased on changing the chain conformation from the anti-parallel $\beta$-pleated to $\alpha$-helical form and also when a double zeta basis set was employed in the calculations. The gap value was also found to decrease in presence of a solvation shell and in presence of cations in the vicinity of the polypeptide chain. Moreover, remarkable lowering in band gap was seen when we replaced alanine with leucine in the polypeptide. The optimum solutions obtained at the end of the GA run also varied accordingly in each case. The DOS distribution
curves further lend a qualitative perspective to the results obtained from GA. Although, a drawback of the \textit{ab initio} Hartree-Fock crystal orbital method is the neglect of correlation between motions of electrons with opposite spins; with the result, this method slightly overestimates the band gap values.

A systematic analysis of all possible percentages of the three components involved, starting from 1\% and going up in steps of 1\% each time, is computationally very expensive. The GA technique helps to automatically find out the optimal solutions through intelligent searches with selective sampling of values in the entire configuration space. Hence the results obtained verify that the GA methodology is a very efficient tool for generating an optimal solution to the problem of theoretical “tailoring” of polypeptides chains with pre-specified properties.
References

Table 7.1: Calculated electronic properties—IP, EA and $E_g$ of the aperiodic ternary polypeptides, along with the optimum percentage compositions and IPN values as obtained from GA [A = gly, B = ala, C = ser, D = leu].

<table>
<thead>
<tr>
<th>System</th>
<th>Optimum composition</th>
<th>IP (eV)</th>
<th>EA (eV)</th>
<th>$E_g$ (eV)</th>
<th>IPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-chain (MB$^a$)</td>
<td>A$_6$B$<em>1$C$</em>{93}$</td>
<td>11.101</td>
<td>-3.595</td>
<td>14.696</td>
<td>0.008018</td>
</tr>
<tr>
<td>α-helix (MB)</td>
<td>A$<em>1$B$</em>{85}$C$_{14}$</td>
<td>11.347</td>
<td>-2.584</td>
<td>13.932</td>
<td>0.008144</td>
</tr>
<tr>
<td>β-chain (DZ$^b$)</td>
<td>A$_1$B$<em>1$C$</em>{98}$</td>
<td>10.591</td>
<td>-2.936</td>
<td>13.527</td>
<td>0.007607</td>
</tr>
<tr>
<td>β-chain (MB + W$^c$)</td>
<td>A$_1$B$_9$C$_8$</td>
<td>11.246</td>
<td>-0.706</td>
<td>11.952</td>
<td>0.008195</td>
</tr>
<tr>
<td>β-chain (H$^+$ + Θ)$^d$</td>
<td>A$_3$B$<em>1$C$</em>{94}$</td>
<td>16.927</td>
<td>5.048</td>
<td>11.879</td>
<td>0.008009</td>
</tr>
<tr>
<td>β-chain (Li$^+$ + Θ)</td>
<td>A$_4$B$<em>1$C$</em>{95}$</td>
<td>15.796</td>
<td>1.638</td>
<td>14.158</td>
<td>0.007851</td>
</tr>
<tr>
<td>β-chain (MB)</td>
<td>A$<em>2$D$</em>{97}$C$_1$</td>
<td>5.117</td>
<td>-3.626</td>
<td>8.743</td>
<td>0.005294</td>
</tr>
</tbody>
</table>

$m$MB: Minimal basis set; $^b$DZ: Double zeta basis set; $^c$W: Polypeptide in presence of water; $^d$Θ: Negative point charges on the cationic adduct.
Fig. 7.1: Chemical structures of the amino acids comprising the polypeptide chain, (a) glycine, (b) alanine, (c) serine, and (d) leucine.
Fig. 7.2 (a): Band alignments for the three homopolypeptides, poly(gly), poly(ala), poly(ser) in the β-chain conformation using MB set (all values are in eV).
Fig. 7.2 (b): Band alignments for the three homopolypeptides, poly(gly), poly(ala), poly(ser) in the α-helix conformation using MB set (all values are in eV).
Fig. 7.2 (c): Band alignments for the three homopolypeptides, poly(gly), poly(ala), poly(ser) in the β-chain conformation using DZ basis set (all values are in eV).
Fig. 7.2 (d): Band alignments for the three homopolypeptides, poly(gly), poly(ala), poly(ser) in the effective field of surrounding water molecules (W) (all values are in eV).
Fig. 7.2 (e): Band alignments the three homopolypeptides, poly(gly), poly(ala), poly(ser), in the β-chain conformation using MB set, forming an adduct with H⁺ ion in the presence of negative point charge Θ (all values are in eV).
Fig. 7.2 (f): Band alignments for the three homopolypeptides, poly(gly), poly(ala), poly(ser), in the β-chain conformation using MB set, forming an adduct with Li⁺ ion in the presence of negative point charge Θ (all values are in eV).
Fig. 7.2 (g): Band alignments for the three homopolypeptides, poly(gly), poly(leu), poly(ser), in the β-chain conformation using MB set (all values are in eV).
Fig. 7.3 (a): DOS of the VB and CB of poly (gly, ala, ser) peptide in the β-chain conformation using MB set, for the solution A₆B₁C₉₃.
Fig. 7.3 (b): DOS of the VB and CB of poly (gly, ala, ser) peptide in the \( \alpha \)-helix conformation using MB set, for the solution \( A_1B_{85}C_{14} \).
Fig. 7.3 (c): DOS of the VB and CB of the poly (gly, ala, ser) peptide in the β-chain conformation using DZ basis set, for the solution A₁B₁C₉₈.
Fig. 7.3 (d): DOS of the VB and CB of the poly (gly, ala, ser) peptide in the β-chain conformation in the effective field of surrounding water molecules (W), for the solution $A_1B_9C_8$. 
Fig. 7.3 (e): DOS of the VB and CB of the poly (gly, ala, ser) peptide in the β-chain conformation using MB set, forming an adduct with H⁺ ion in the presence of negative point charge Θ, for the solution A₅B₁C₉₄.
Fig. 7.3 (f): DOS of the VB and CB of the poly (gly, ala, ser) peptide in the β-chain conformation using MB set, forming an adduct with Li$^+$ ion in the presence of negative point charge Θ, for the solution $A_4B_1C_{95}$. 
Fig. 7.3 (g): DOS of the VB and CB of poly (gly, leu, ser) peptide in the β-chain conformation using MB set, for the solution $A_2D_{97}C_1$. 