Semi empirical analysis of proteins and analysis of long chain polymers(bonding via π stacking and H-bonding)

INTRODUCTION

my work basically comprised of analysis of protein molecules and studying their interactions when they are in long chains . It also initially involved construction of amino acids from scratch and then joining together by various possible peptide bonds and analyzing their DLMO's and bonding pattern .

the programs and tools used for this analysis were gauss view and gaussian03 . gauss view is basically used to formulate and visualize a molecular skeleton in 3-d . It has all the tools for editing and creating bonds and structure , altering bond lengths , bond angles and all other stereo chemical properties . In short its an interactive platform for creating and editing molecules . Gaussian , on the other hand was used for optimizing structures created in gauss view by using various techniques such as Hartree fork and semi empirical methods .

CREATING MONOMER UNITS

(Tree Branch Method)

(***excerpt from modelling of interfaces in emulsion explosives PhD thesis***)

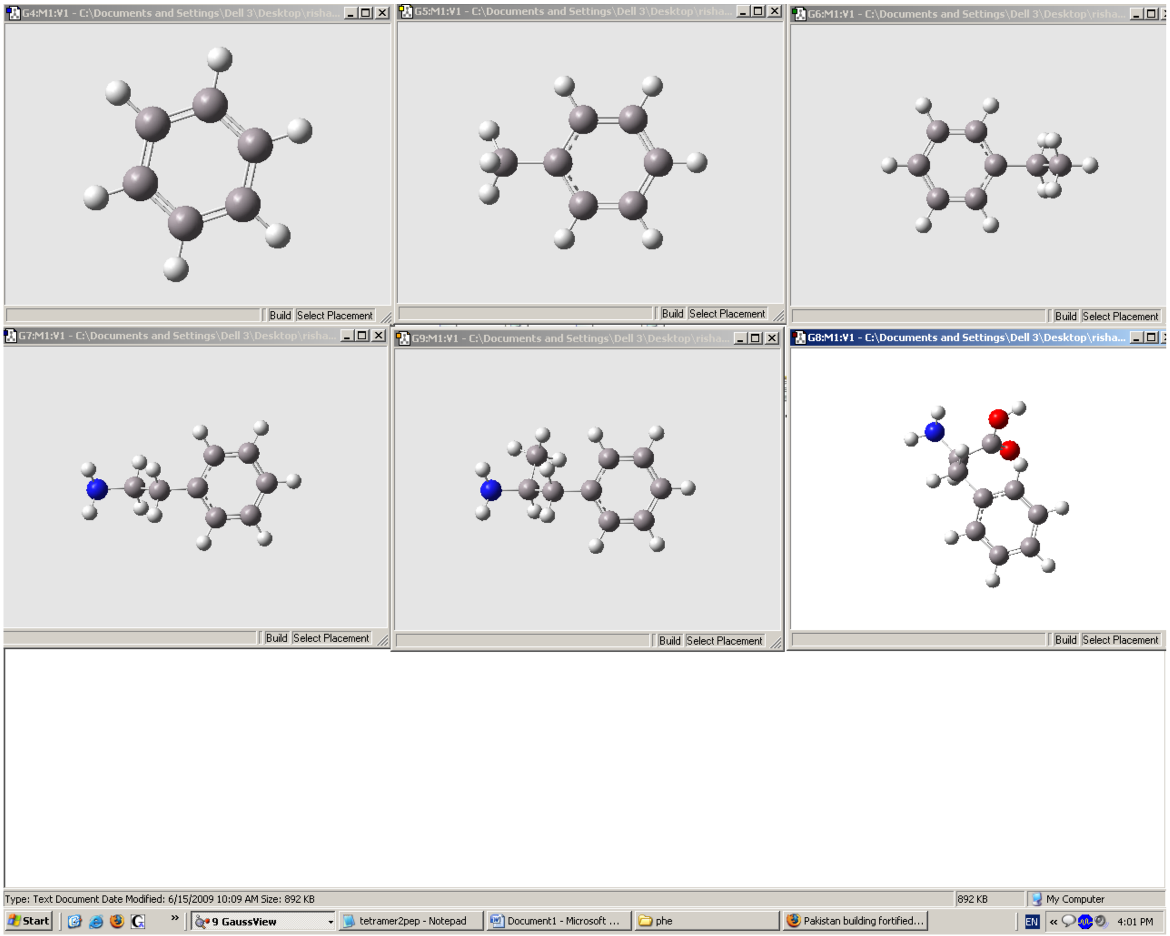
The potential energy functions used in molecular mechanics calculations , ensure that a large number of minima (other than the true global minimum) are present on the potential energy surface the presence of a large number of minima , makes it impossible to identify the global minimum from a single minimization . Hence the proposed method is the tree branch method, by which extensive computational times are avoided by using minimum energy conditions to eliminate large portions of the tree, and the number of structures to be evaluated .

The proposed method begins by minimizing a small precursor of the target molecule , which is identified from a functional group in the molecule . For example , in case of oleic acid , the smallest precursor are formic acid , ethylene or ethane . A minimum is first defined for one of the precursors , and the molecule is grown one heavy atom (non hydrogen) at a time . Each new heavy atom is entered at the position of one of the hydrogens , on the atom it is being attached to . The appropriate number of hydrogens are added to the new atom , and structure minimized . The procedure is repeated for all other hydrogen positions on older substrate , and only the lowest energy structure is retained to continue the growth . For a sp3 hybridized atom such as carbon in R-CH3 , three minimizations are needed for each new substituted atom . By selecting the lowest energy conformer , a very specific path is followed along the energy trees .

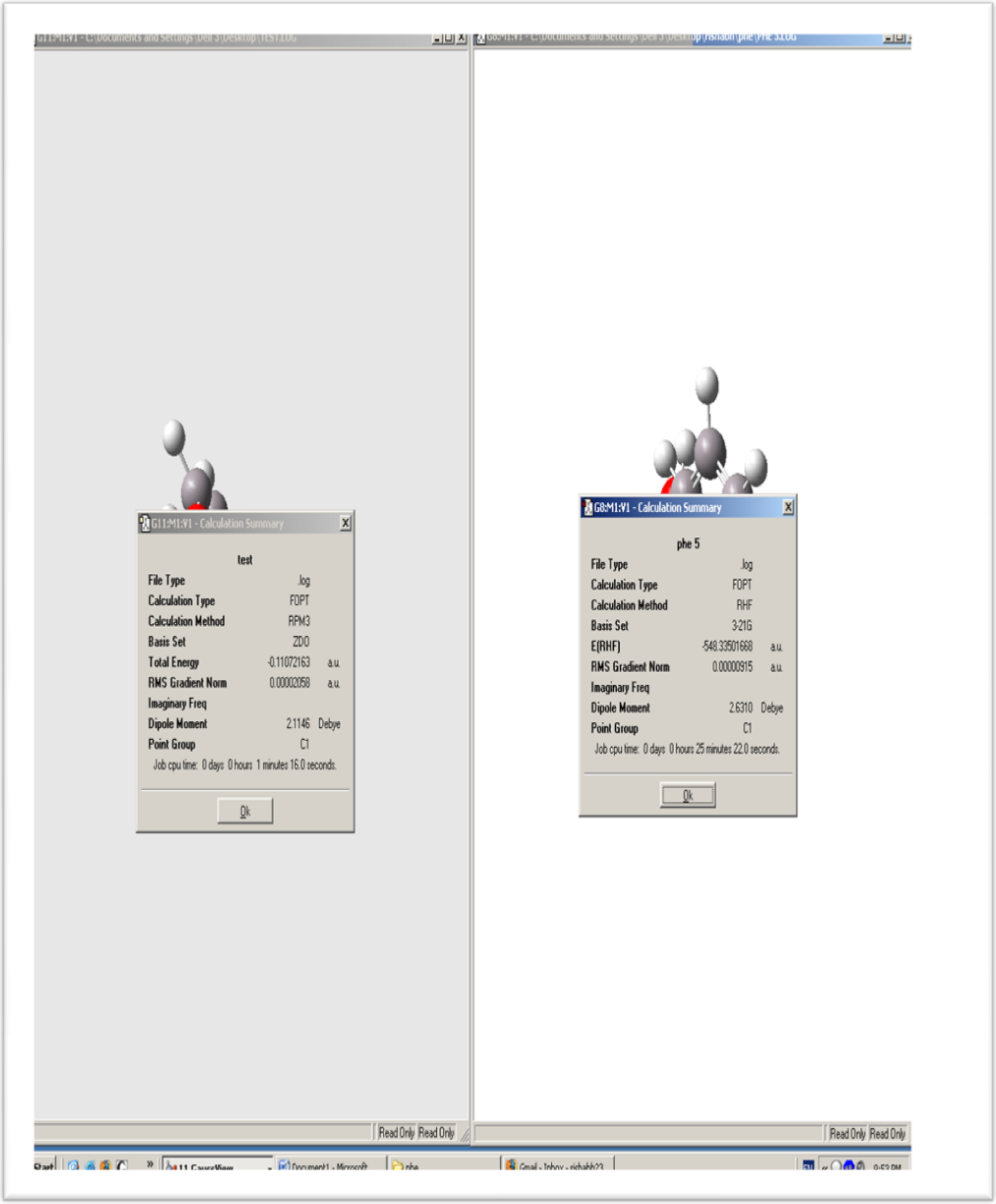
The proposed method does not necessarily identify the overall global minimum , or all the the lowest energy minima . However , it has a very high probability of identifying one of the lower energy conformers for target molecule .

STUDY OF INTERACTION BETWEEN MONOMER UNITS FORMED BY JOINING PHENYLALANINE AND GLUTAMIC ACID BY PEPTIDE BONDS

As the first step of analyzing the peptide linkage between the two chosen molecules , phenylalanine was constructed step by step using tree branch method .



The phenylalanine was optimized in Gaussian 03 using semi empirical pm 3 method . It was finally optimized in water medium for bringing it in its zwitter ion form . Also this procedure very beautifully demonstrates the validity and usefulness of tree branch method . The inbuilt structure when optimized shows energy much greater then the structure obtained from the tree branch optimization



```
G11:M1:V1 - Calculation Summary

                 test
File Type                     .log
Calculation Type              FOPT
Calculation Method            RPM3
Basis Set                     ZDO
Total Energy           -0.11072163   a.u.
RMS Gradient Norm       0.00002058   a.u.
Imaginary Freq
Dipole Moment               2.1146   Debye
Point Group                   C1
 Job cpu time:  0 days  0 hours  1 minutes 16.0 seconds.

                  Ok
```

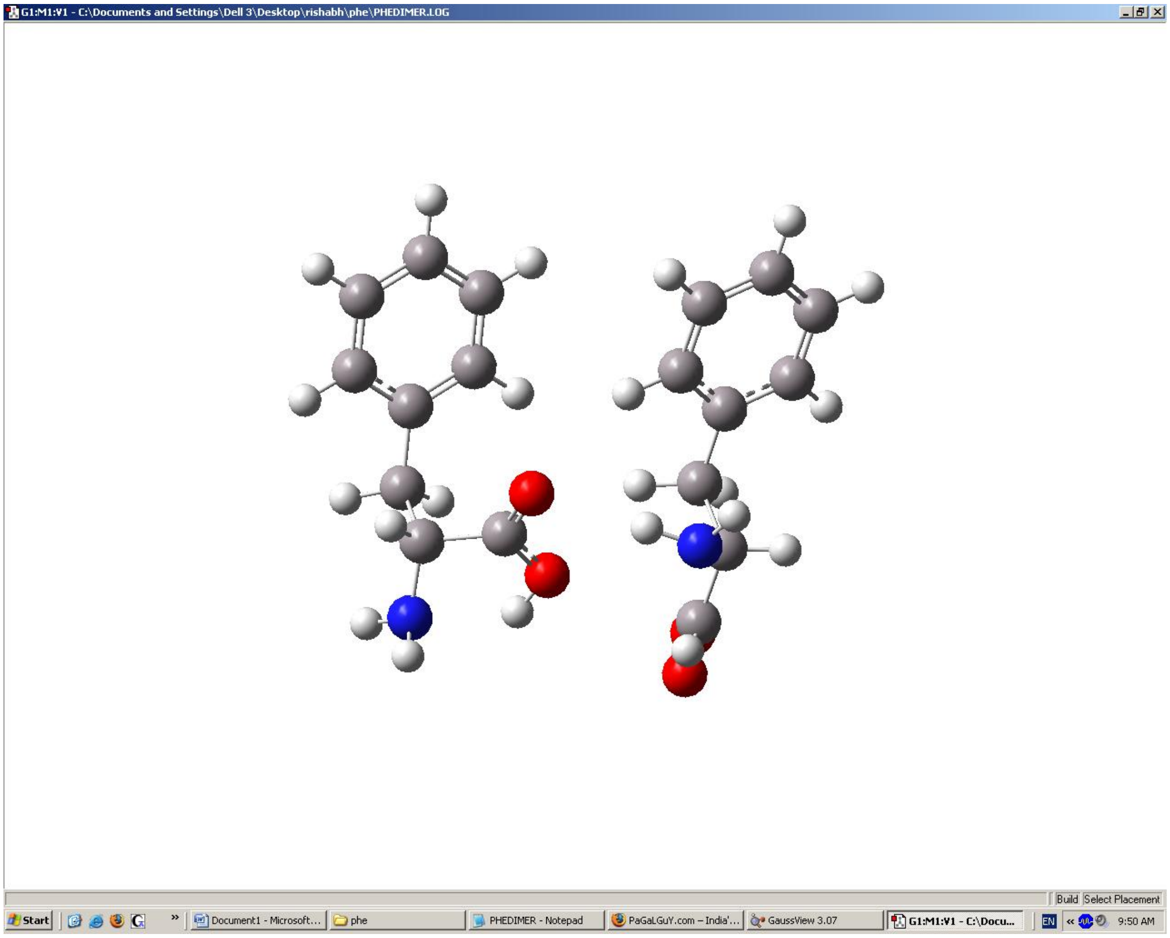
```
G8:M1:V1 - Calculation Summary

                 phe 5
File Type                     .log
Calculation Type              FOPT
Calculation Method            RHF
Basis Set                     3-21G
E(RHF)               -548.33501668   a.u.
RMS Gradient Norm       0.00000915   a.u.
Imaginary Freq
Dipole Moment               2.6310   Debye
Point Group                   C1
 Job cpu time:  0 days  0 hours 25 minutes 22.0 seconds.

                  Ok
```

a

In an initial bid to analyze the pie stacking and the nature of bonding between phen , a dimer and then trimer of phen was also optimized , to see the pie stacking and the nature of h-bonding between the monomer units .

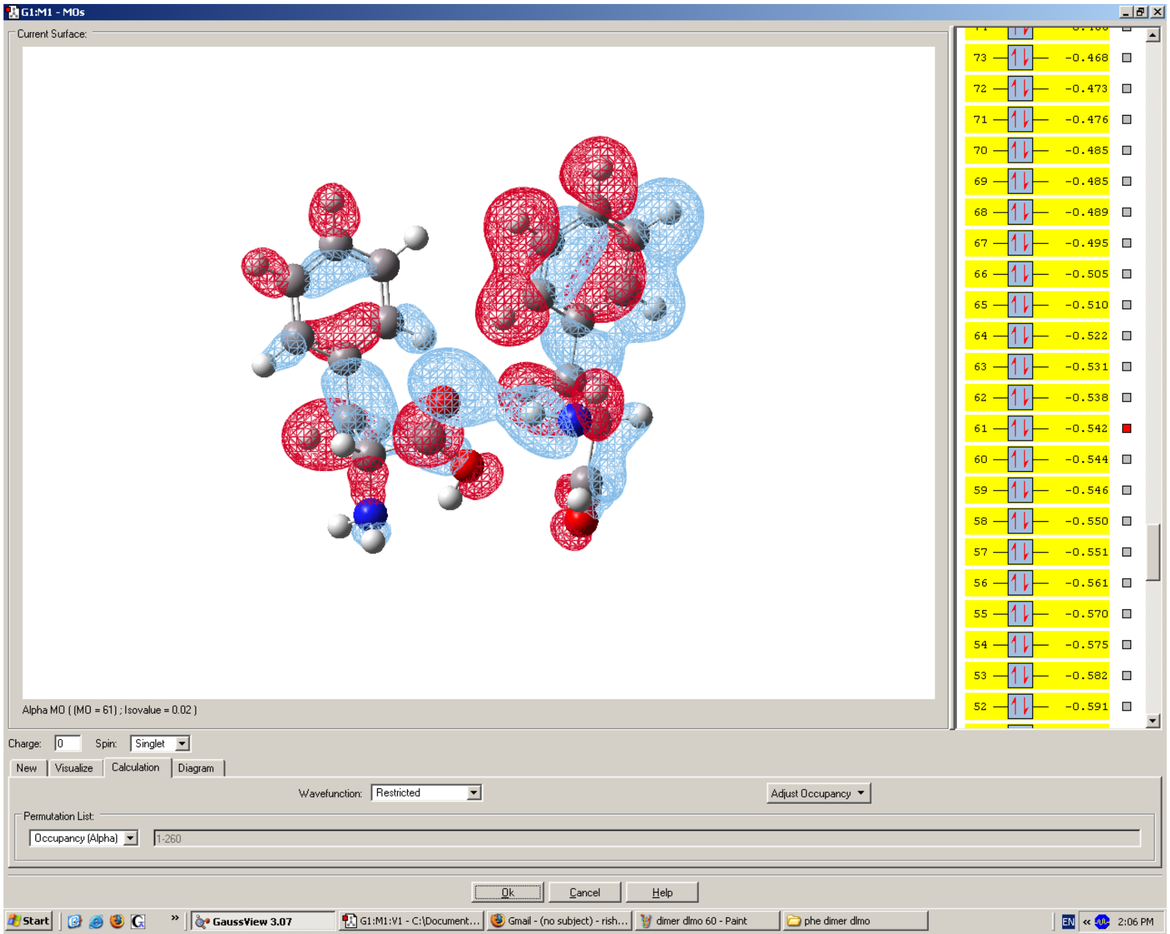


The result of analysis of dimer of phen showed following results

ENERGY : -1096.69451969 a.u.

R.M.S Gradient norm : 0.00001098 a.u.

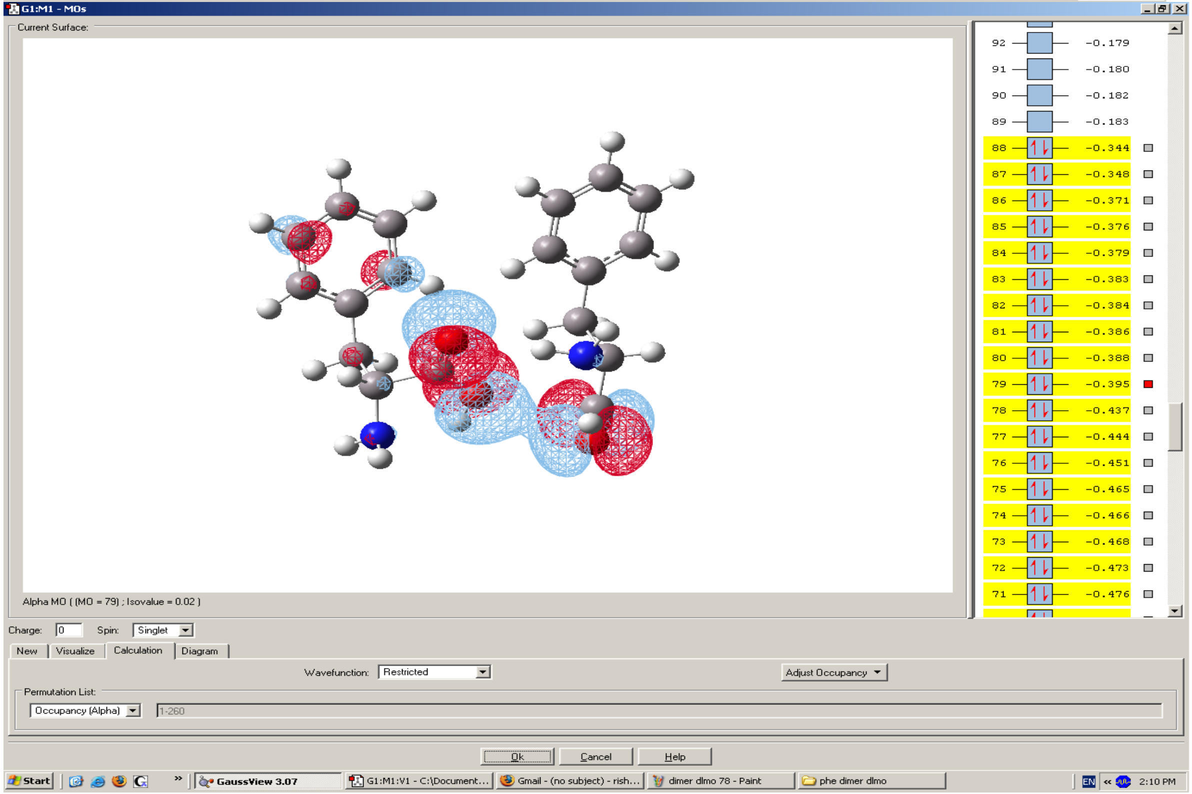
On generating the MO’s of dimer , mainly two types of MO’s are seen . The s orbitals wavefunctions and the p orbitals wavefunction .



Alpha MO ((MO = 61) ; Isovalue = 0.02)

Charge: 0
Spin: Singlet
Wavefunction: Restricted
Permutation List:
Occupancy (Alpha) 1-260

73	↑↓	-0.468
72	↑↓	-0.473
71	↑↓	-0.476
70	↑↓	-0.485
69	↑↓	-0.485
68	↑↓	-0.489
67	↑↓	-0.495
66	↑↓	-0.505
65	↑↓	-0.510
64	↑↓	-0.522
63	↑↓	-0.531
62	↑↓	-0.538
61	↑↓	-0.542
60	↑↓	-0.544
59	↑↓	-0.546
58	↑↓	-0.550
57	↑↓	-0.551
56	↑↓	-0.561
55	↑↓	-0.570
54	↑↓	-0.575
53	↑↓	-0.582
52	↑↓	-0.591



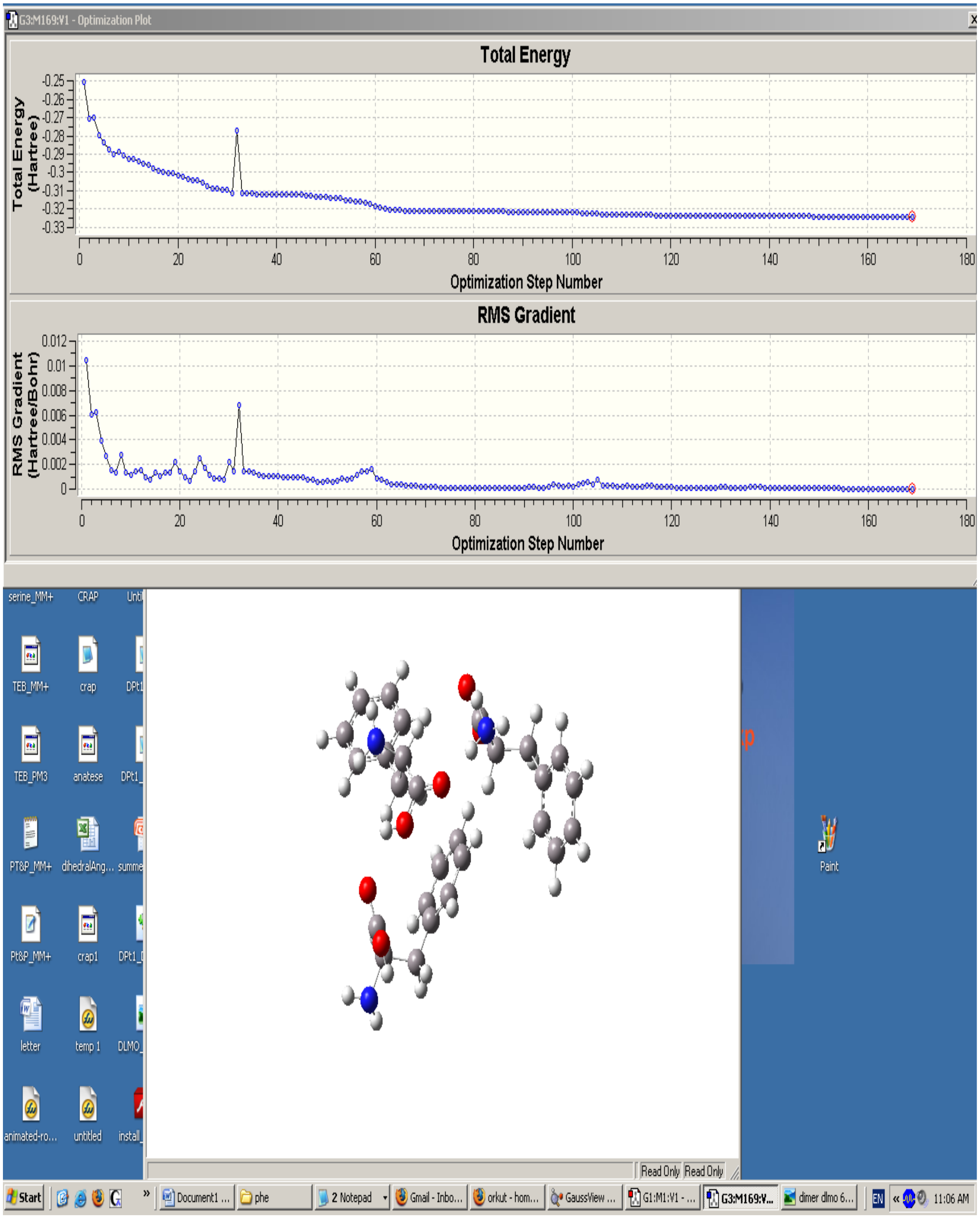
Alpha MO ((MO = 79) ; Isovalue = 0.02)

Charge: 0
Spin: Singlet
Wavefunction: Restricted
Permutation List:
Occupancy (Alpha) 1-260

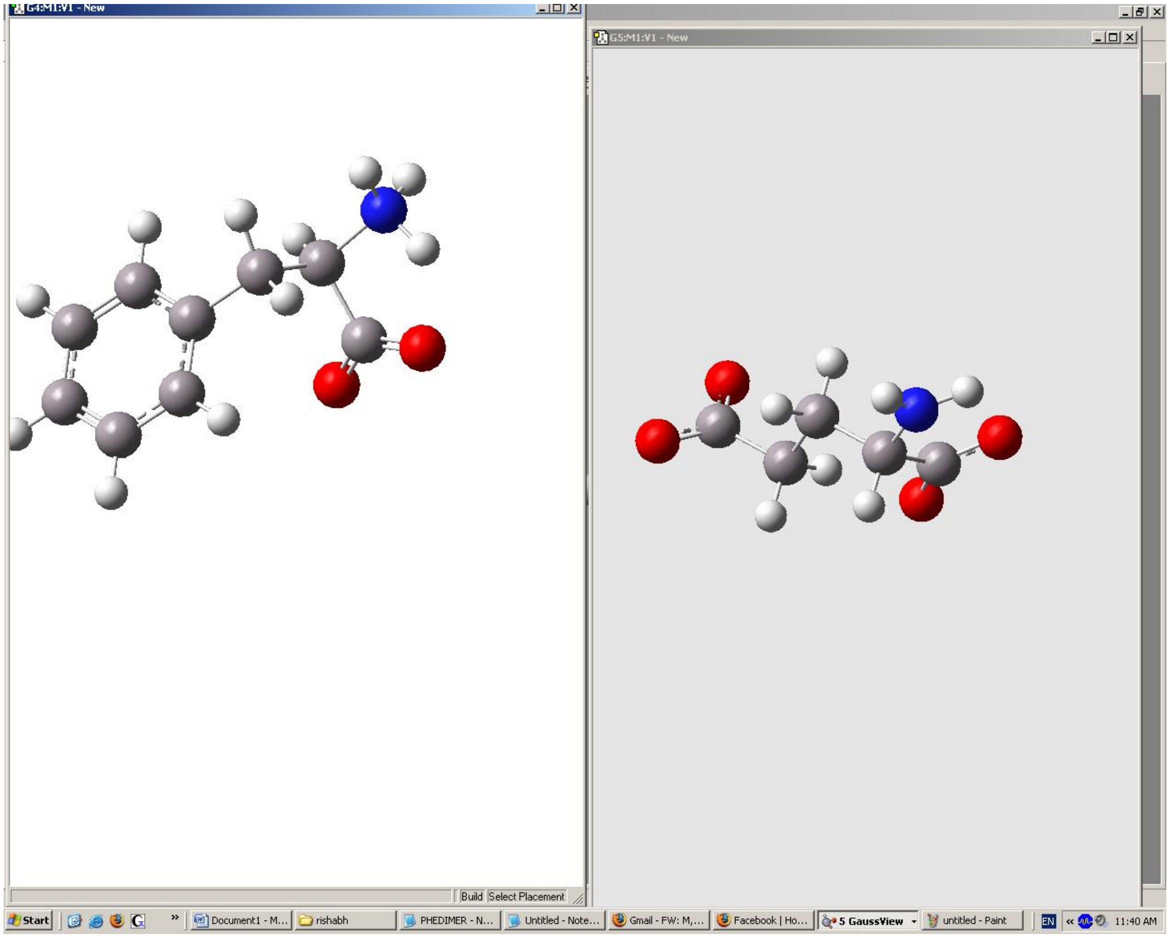
92		-0.179
91		-0.180
90		-0.182
89		-0.183
88	↑↓	-0.344
87	↑↓	-0.348
86	↑↓	-0.371
85	↑↓	-0.376
84	↑↓	-0.379
83	↑↓	-0.383
82	↑↓	-0.384
81	↑↓	-0.386
80	↑↓	-0.388
79	↑↓	-0.395
78	↑↓	-0.437
77	↑↓	-0.444
76	↑↓	-0.451
75	↑↓	-0.465
74	↑↓	-0.466
73	↑↓	-0.468
72	↑↓	-0.473
71	↑↓	-0.476

on Further scrutiny of the MO's it was found that initial MO's were the ones that had nothing to do with the molecular interactions and were restricted to the monomer and hence we can omit the initial MO structure in further analysis .

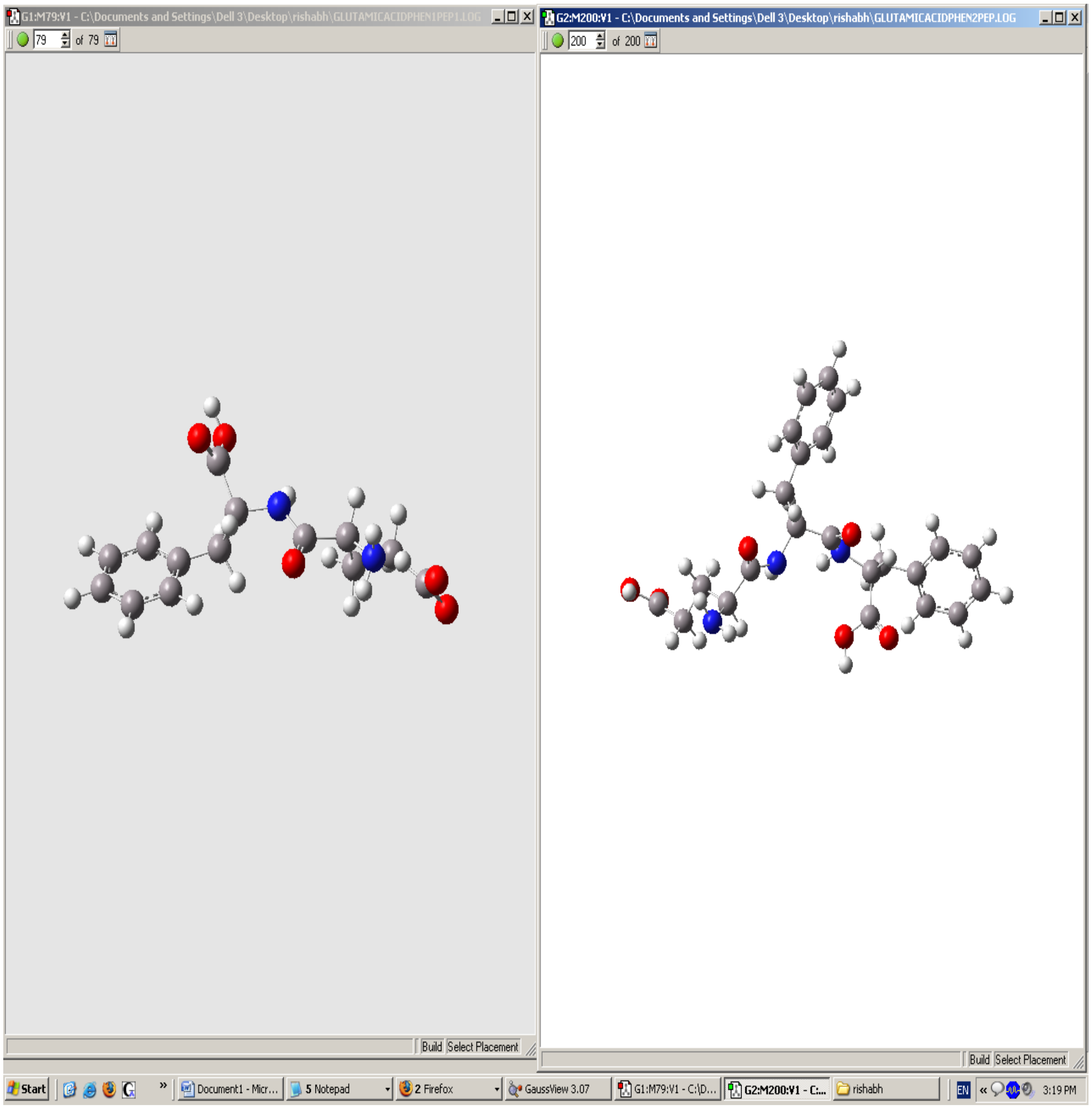
In the similar fashion the trimer of phen was also optimized to study the h-bonding and to explore the possibility of π stacking , if any , in between the monomer interactions .



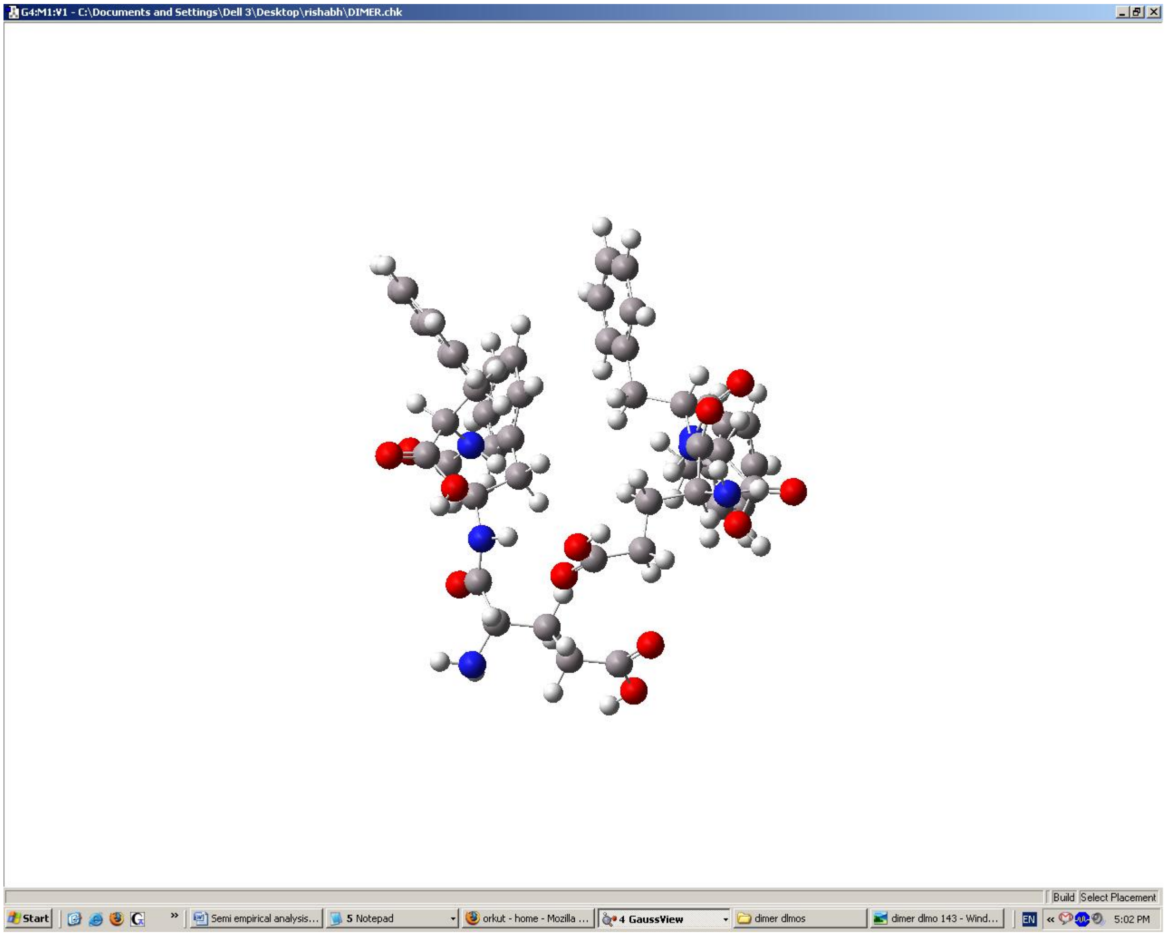
In a similar fashion glutamic acid was also optimized , and now we had the two structures ready for analysis of stacking of monomer units linked by peptide bonds .



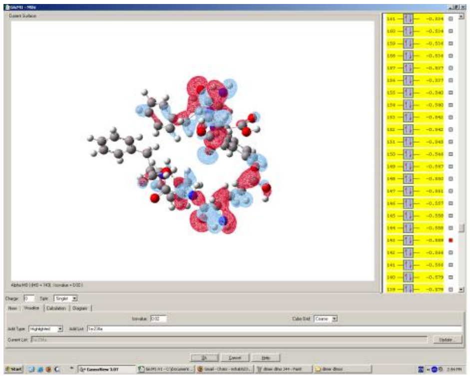
Now if we closely analyze both the structure (phen and glutamic acid) , we will find that there are 3 peptide bonds possibe between the two molecules . Also due to availability of sites on glutamic acid , two phen molecules were also joined on single one .

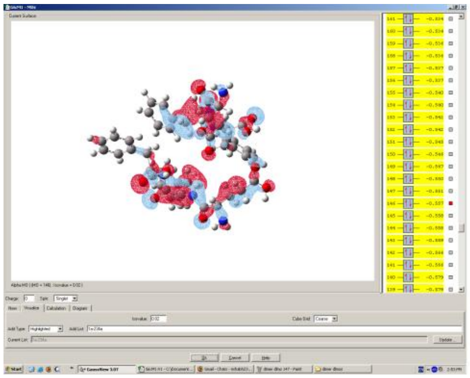


in the first pic , one molecule of phen and glutamic acid are joined end to end . While in second structure two phen are joine on to single phen molecule .



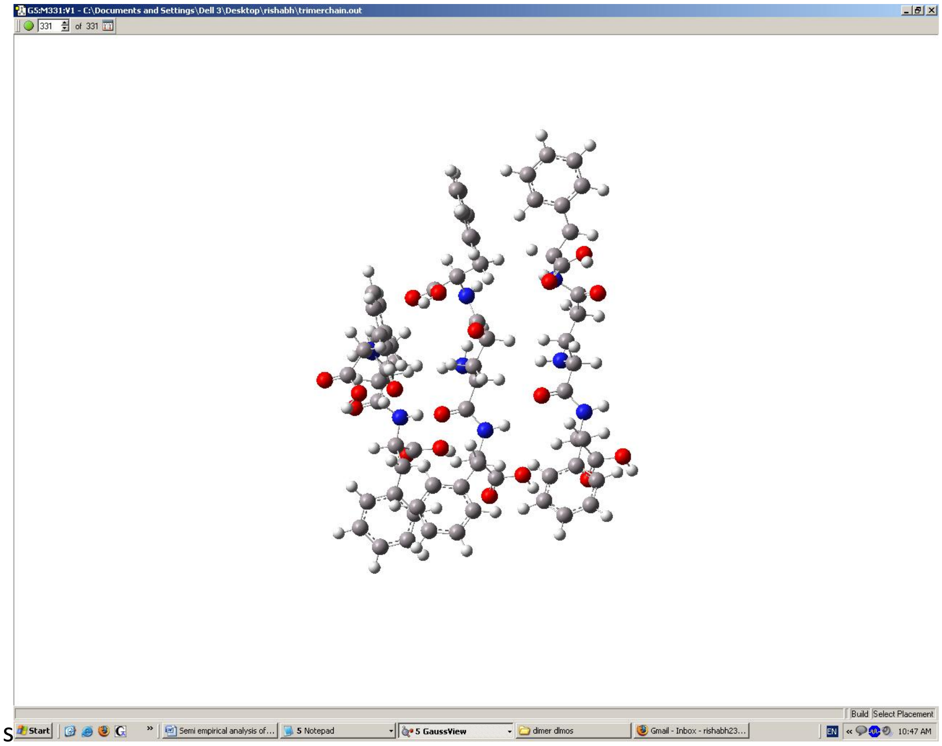
DIMER OF TWO PEPTIDE BOND MONOMER





the MO's show the H-bonding betweeb the monomer units . The lower orbitals were excluded because they are mainly s wavefunctions which do not participate in molecular bonds .

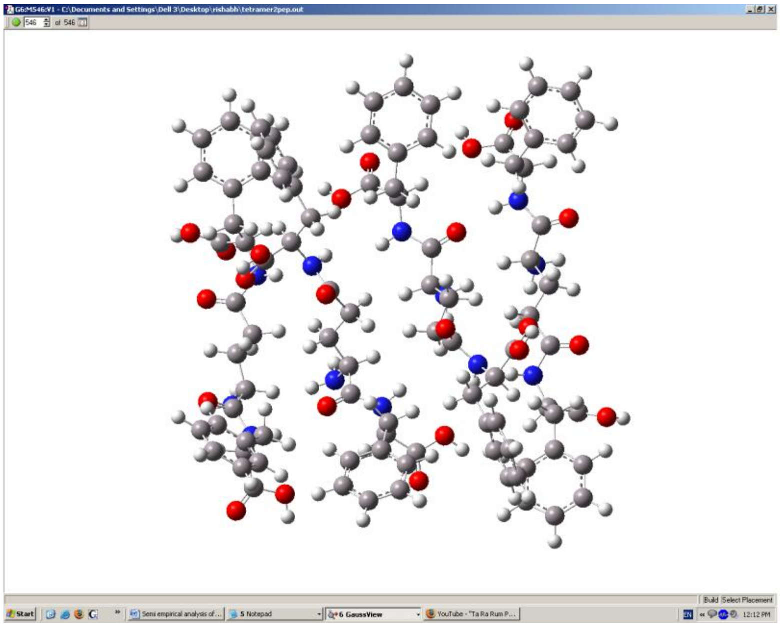
in the similar fashion trimer was also analyzed for the same monomer. they lined up side on and



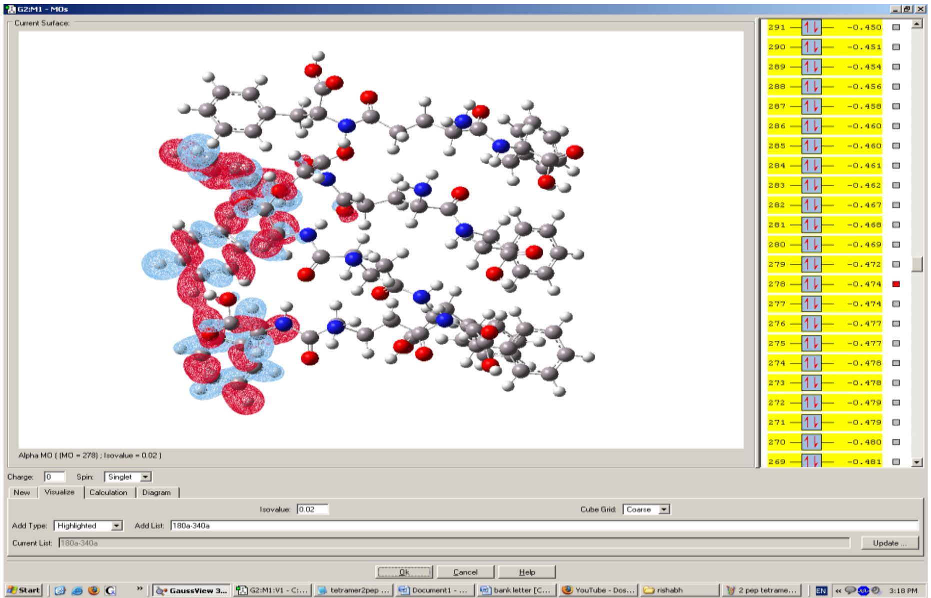
s

provided a very good insight into the possiblity of their forming long chain structures . It was evident from their Molecular analysis structures that they primarily combine via H-bonds . When the structure was optimized for a tetramer , it showed similar result with the molecules lining side by side .

Also the molecular orbital analysis showed them to be connected by H-bonds and the there is also significant twisting in the molecule as a whole



The molecule also showed some amount of π stacking . Due to twisting of the tetramer , it was not probably captured at its best , but certainly there was a ceratin amount of π stacking present .



291	-0.450	☐
290	-0.451	☐
289	-0.454	☐
288	-0.456	☐
287	-0.458	☐
286	-0.460	☐
285	-0.460	☐
284	-0.461	☐
283	-0.462	☐
282	-0.467	☐
281	-0.468	☐
280	-0.469	☐
279	-0.472	☐
278	-0.474	■
277	-0.474	☐
276	-0.477	☐
275	-0.477	☐
274	-0.478	☐
273	-0.478	☐
272	-0.479	☐
271	-0.479	☐
270	-0.480	☐
269	-0.481	☐

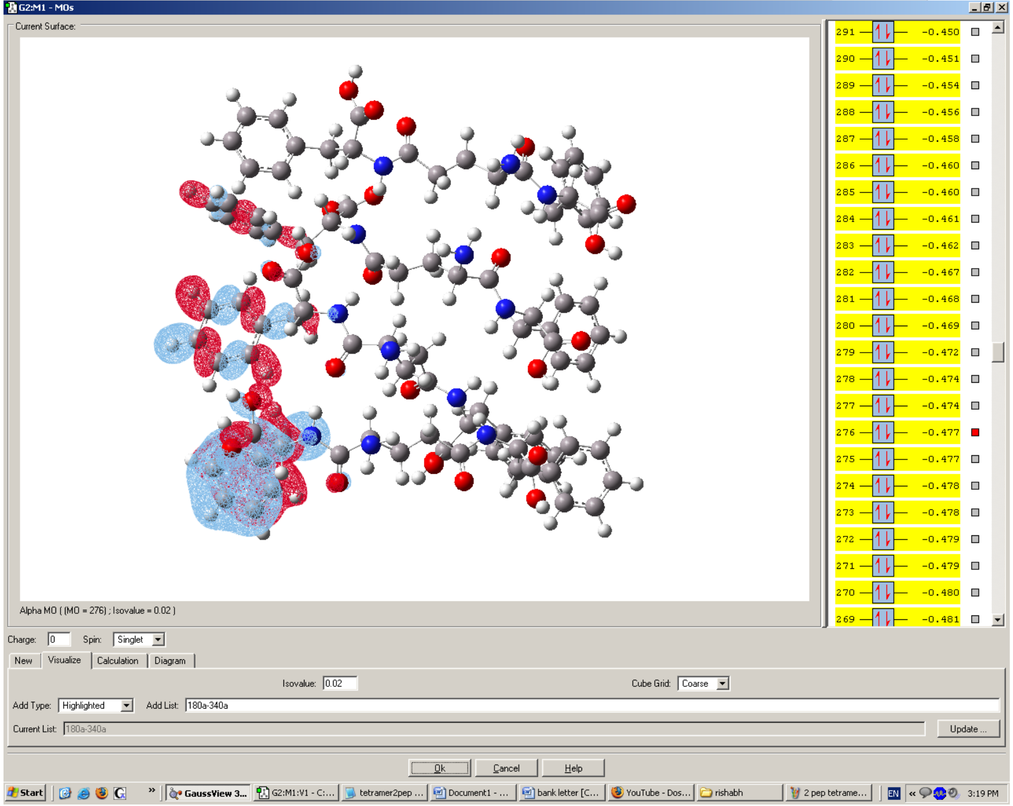
Alpha MO ((MO = 278) ; Isovalue = 0.02)

Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a



291	-0.450	☐
290	-0.451	☐
289	-0.454	☐
288	-0.456	☐
287	-0.458	☐
286	-0.460	☐
285	-0.460	☐
284	-0.461	☐
283	-0.462	☐
282	-0.467	☐
281	-0.468	☐
280	-0.469	☐
279	-0.472	☐
278	-0.474	☐
277	-0.474	☐
276	-0.477	■
275	-0.477	☐
274	-0.478	☐
273	-0.478	☐
272	-0.479	☐
271	-0.479	☐
270	-0.480	☐
269	-0.481	☐

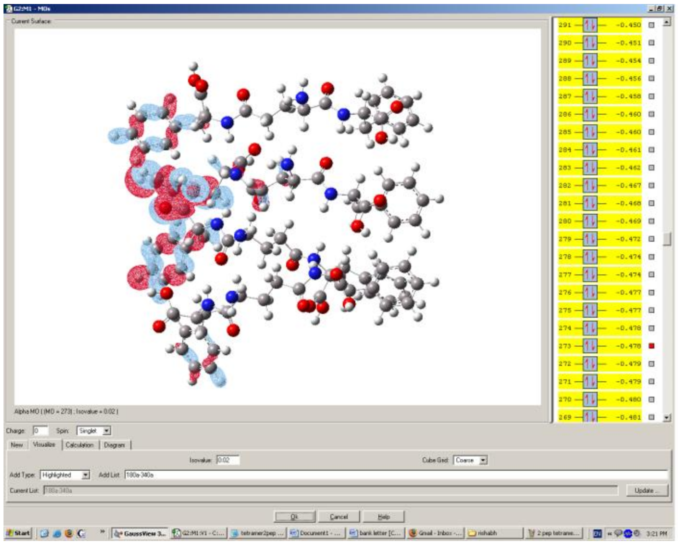
Alpha MO ((MO = 276) ; Isovalue = 0.02)

Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a



Alpha MO ((MO = 273) ; Isovalue = 0.02)

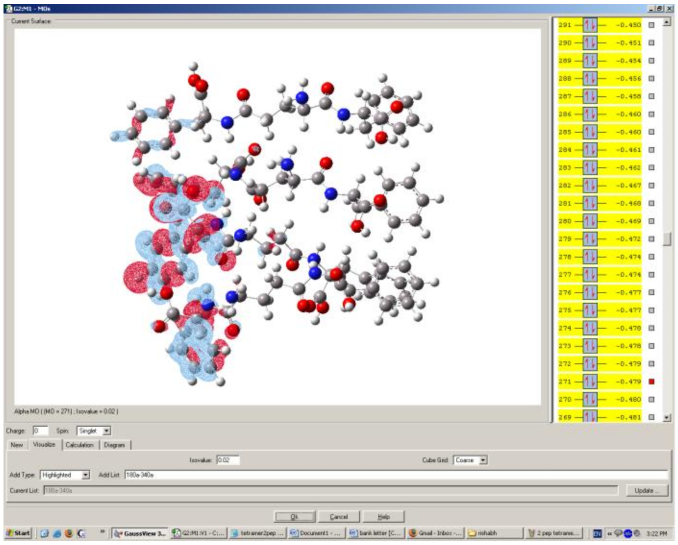
Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a

291	-0.450	☐
290	-0.451	☐
289	-0.454	☐
288	-0.456	☐
287	-0.458	☐
286	-0.460	☐
285	-0.460	☐
284	-0.461	☐
283	-0.462	☐
282	-0.467	☐
281	-0.468	☐
280	-0.469	☐
279	-0.472	☐
278	-0.474	☐
277	-0.474	☐
276	-0.477	☐
275	-0.477	☐
274	-0.478	☐
273	-0.478	■
272	-0.479	☐
271	-0.479	☐
270	-0.480	☐
269	-0.481	☐



Alpha MO ((MO = 271) ; Isovalue = 0.02)

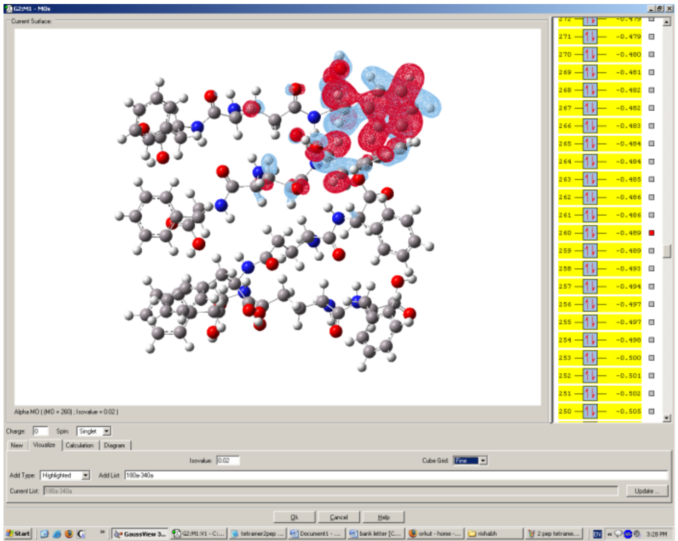
Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a

291	-0.450	☐
290	-0.451	☐
289	-0.454	☐
288	-0.456	☐
287	-0.458	☐
286	-0.460	☐
285	-0.460	☐
284	-0.461	☐
283	-0.462	☐
282	-0.467	☐
281	-0.468	☐
280	-0.469	☐
279	-0.472	☐
278	-0.474	☐
277	-0.474	☐
276	-0.477	☐
275	-0.477	☐
274	-0.478	☐
273	-0.478	☐
272	-0.479	☐
271	-0.479	■
270	-0.480	☐
269	-0.481	☐



Alpha MO ((MO = 260) ; Isovalue = 0.02)

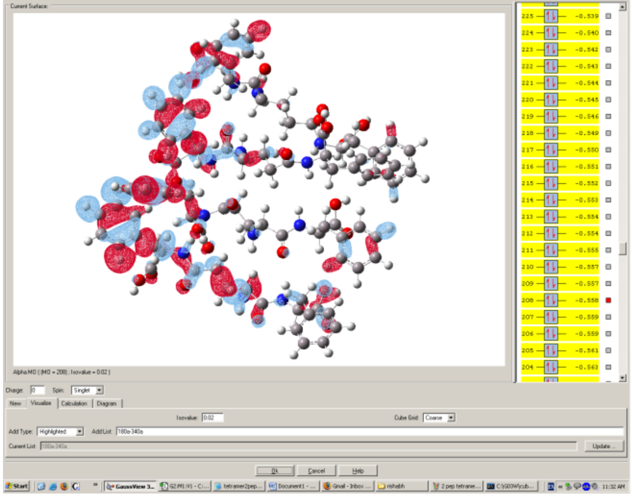
271	-0.479	☐
270	-0.480	☐
269	-0.481	☐
268	-0.482	☐
267	-0.482	☐
266	-0.483	☐
265	-0.484	☐
264	-0.484	☐
263	-0.485	☐
262	-0.486	☐
261	-0.486	☐
260	-0.489	■
259	-0.489	☐
258	-0.493	☐
257	-0.494	☐
256	-0.497	☐
255	-0.497	☐
254	-0.498	☐
253	-0.500	☐
252	-0.501	☐
251	-0.502	☐
250	-0.505	☐

Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Fine

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a



Alpha MO ((MO = 208) ; Isovalue = 0.02)

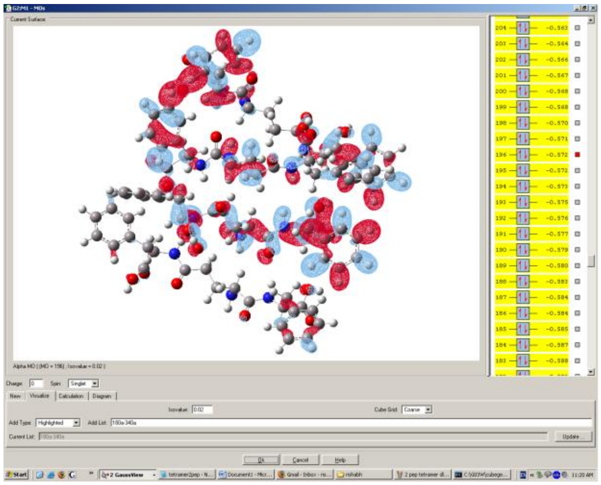
225	-0.539	☐
224	-0.540	☐
223	-0.542	☐
222	-0.543	☐
221	-0.544	☐
220	-0.545	☐
219	-0.546	☐
218	-0.549	☐
217	-0.550	☐
216	-0.551	☐
215	-0.552	☐
214	-0.553	☐
213	-0.554	☐
212	-0.554	☐
211	-0.555	☐
210	-0.557	☐
209	-0.557	☐
208	-0.558	■
207	-0.559	☐
206	-0.559	☐
205	-0.561	☐
204	-0.563	☐

Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a



Alpha MO ((MO = 196) ; Isovalue = 0.02)

204	-0.563	☐
203	-0.564	☐
202	-0.566	☐
201	-0.567	☐
200	-0.568	☐
199	-0.568	☐
198	-0.570	☐
197	-0.571	☐
196	-0.572	■
195	-0.572	☐
194	-0.573	☐
193	-0.575	☐
192	-0.576	☐
191	-0.577	☐
190	-0.579	☐
189	-0.580	☐
188	-0.583	☐
187	-0.584	☐
186	-0.584	☐
185	-0.585	☐
184	-0.587	☐
183	-0.588	☐

Charge: 0 Spin: Singlet

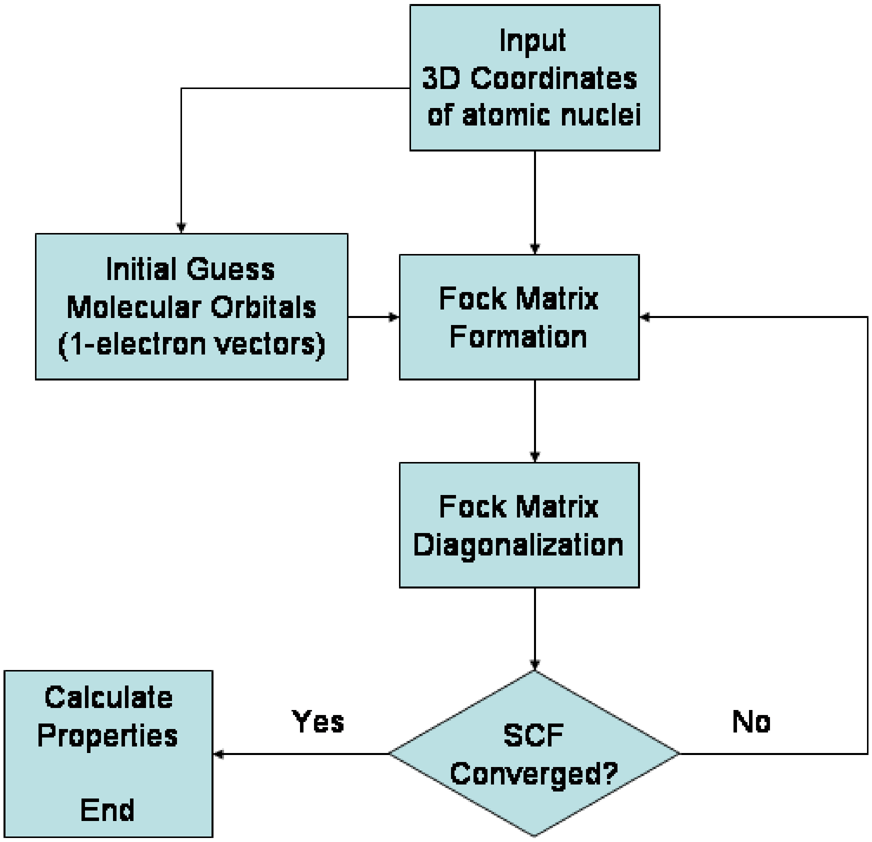
Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 180a-340a

Current List: 180a-340a

MATHEMATICAL PROCEDURES APPLIED AND USED

Two main methods used in analyzing these macromolecules. They were the Hartree fock and semi-empirical method. The Hartree–Fock method assumes that the exact, *N*-body wave function of the system can be approximated by a single permanent of N-spin orbitals . The Hartree–Fock method is typically used to solve the time-independent Schrödinger equation for a multi-electron atom or molecule as described in the Born Oppenheimer approximation .



The Hartree fock Method uses 5 major approximations in getting the wavefunctions.

1. The born oppenheimer approximation is inherently assumed. The full molecular wave function is actually a function of the coordinates of each of the nuclei, in addition to those of the electrons.
2. Relativistic effects are completely neglected.
3. The vibrational solution is always considered to linear combination of finite number of basis sets, which are usually chosen to be orthogonal.

4. Each energy eigenfunctions is assumed to be described by single Slater determinant. an anti symmetrized .
5. The mean field approximation is implied. Effects arising from his assumption, like that of electron correlation are completely ignored.

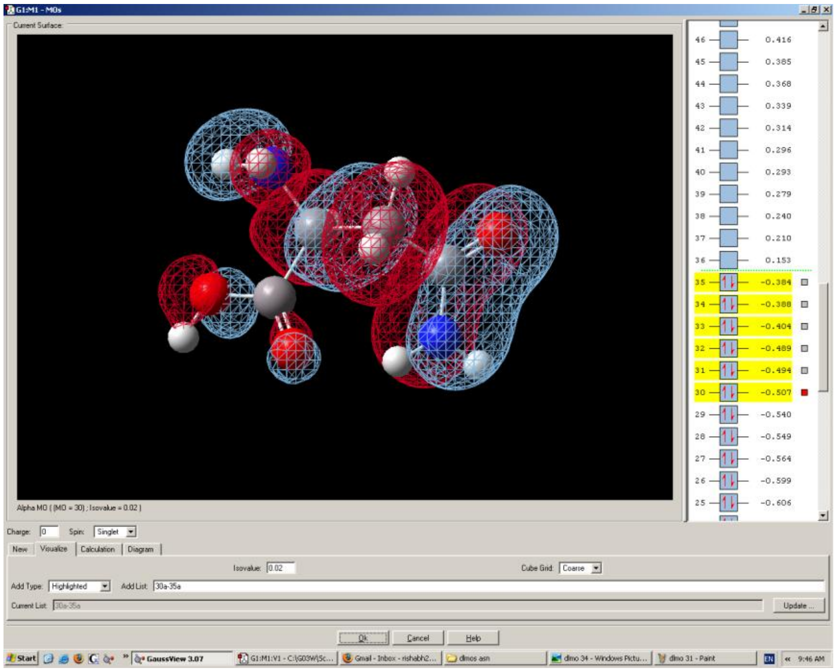
The other faster, somewhat less accurate than Hartree fock , but extremely useful method used was semi empirical PM 3 . It is based on neglect of differential diatomic overlap integral approximation.

Semi empirical method considerably reduces the number of integrals to be evaluated and is significantly faster than the Hartree fock method.

The main analysis lies in interpreting the DLMO's and finding a pattern or conclusion in the bonding pattern of long chain protein structures.

the analysis showed that as more and more monomer unit of a long chain protein structure is added , the structure gets more and more twisted . Also the processing time increases significantly with increase in the number of atoms , and so with larger molecules semi empirical PM3 is the option left

SOME OF THE DLMO's GENERATED (PICTURE APPENDIX)

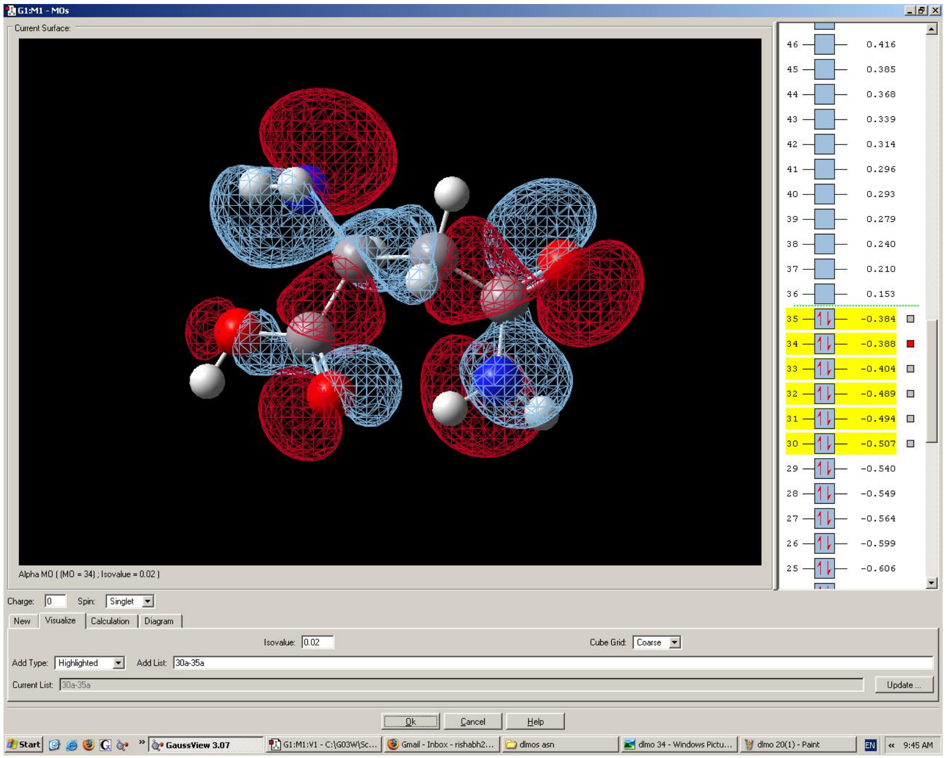


46	0.416
45	0.385
44	0.368
43	0.339
42	0.314
41	0.296
40	0.293
39	0.279
38	0.240
37	0.210
36	0.153
35	-0.384
34	-0.388
33	-0.404
32	-0.489
31	-0.494
30	-0.507
29	-0.540
28	-0.549
27	-0.564
26	-0.599
25	-0.606

Alpha MO ((MO = 30) ; Isovalue = 0.02)

Charge: 0
Spin: Singlet
Isovalue: 0.02
Cube Grid: Coarse
Add Type: Highlighted
Add List: 30a-35a
Current List: 30a-35a

THE s orbitals…not much significant



46	0.416
45	0.385
44	0.368
43	0.339
42	0.314
41	0.296
40	0.293
39	0.279
38	0.240
37	0.210
36	0.153
35	-0.384
34	-0.388
33	-0.404
32	-0.489
31	-0.494
30	-0.507
29	-0.540
28	-0.549
27	-0.564
26	-0.599
25	-0.606

Alpha MO ((MO = 34) ; Isovalue = 0.02)

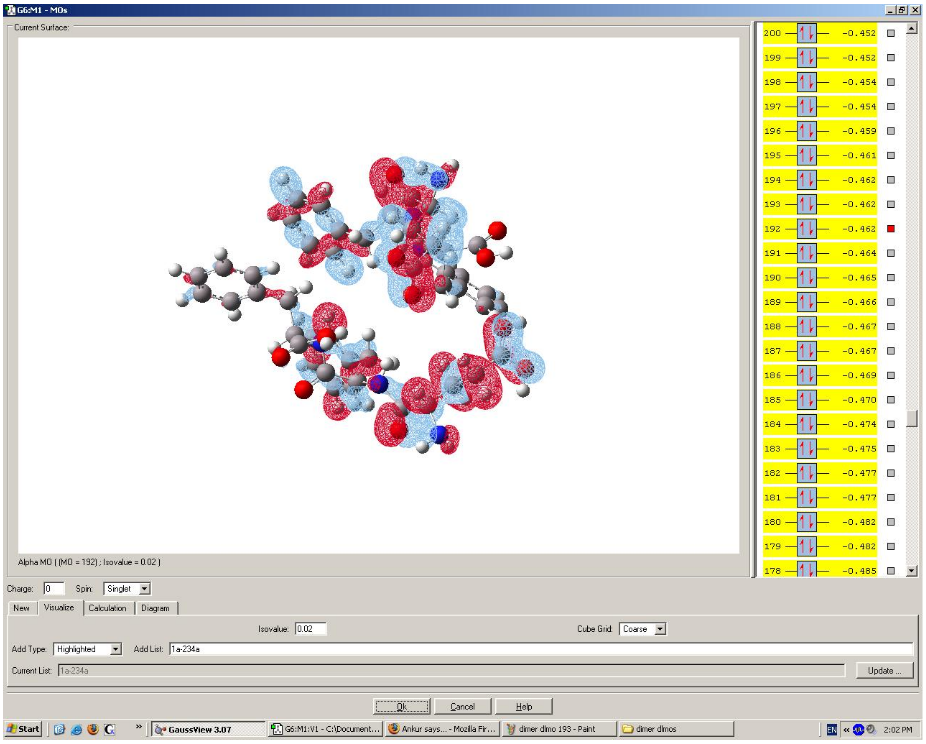
Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 30a-35a

Current List: 30a-35a

'p' wavefunction. clearly visible nodes



200	-0.452
199	-0.452
198	-0.454
197	-0.454
196	-0.459
195	-0.461
194	-0.462
193	-0.462
192	-0.462
191	-0.464
190	-0.465
189	-0.466
188	-0.467
187	-0.467
186	-0.469
185	-0.470
184	-0.474
183	-0.475
182	-0.477
181	-0.477
180	-0.482
179	-0.482
178	-0.485

Alpha MO ((MO = 192) ; Isovalue = 0.02)

Charge: 0 Spin: Singlet

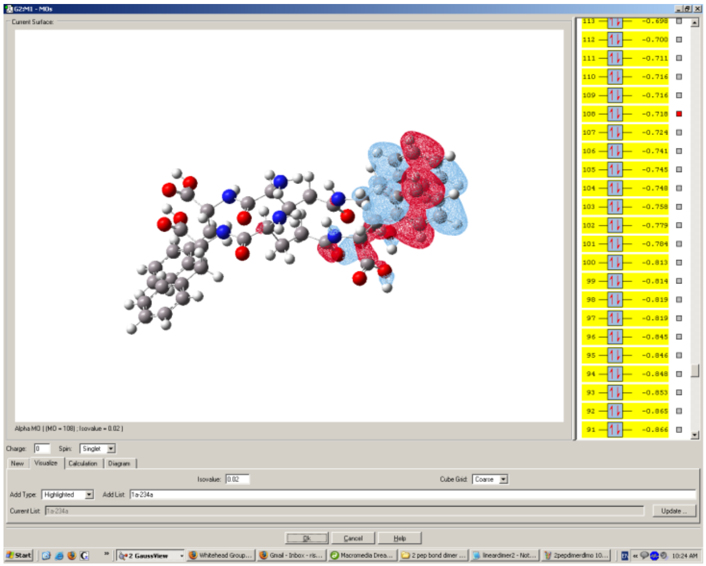
Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 1a-234a

Current List: 1a-234a

H-BONDING..BINGOO !!..:)

During the analysis of orbitals , for a typically large molecule we get dlmos in typical order of 300-400 and hence its pretty necessary to phase out unimportant lower wavefunctions and to consider only the higher and the more prominent ones involved in molecular bonding .



G2:M1 - MOs

Current Surface:

113	-0.698	☐
112	-0.700	☐
111	-0.711	☐
110	-0.716	☐
109	-0.716	☐
108	-0.718	■
107	-0.724	☐
106	-0.741	☐
105	-0.745	☐
104	-0.748	☐
103	-0.758	☐
102	-0.779	☐
101	-0.784	☐
100	-0.813	☐
99	-0.814	☐
98	-0.819	☐
97	-0.819	☐
96	-0.845	☐
95	-0.846	☐
94	-0.848	☐
93	-0.853	☐
92	-0.865	☐
91	-0.866	☐

Alpha MO ((MO = 108) ; Isovalue = 0.02)

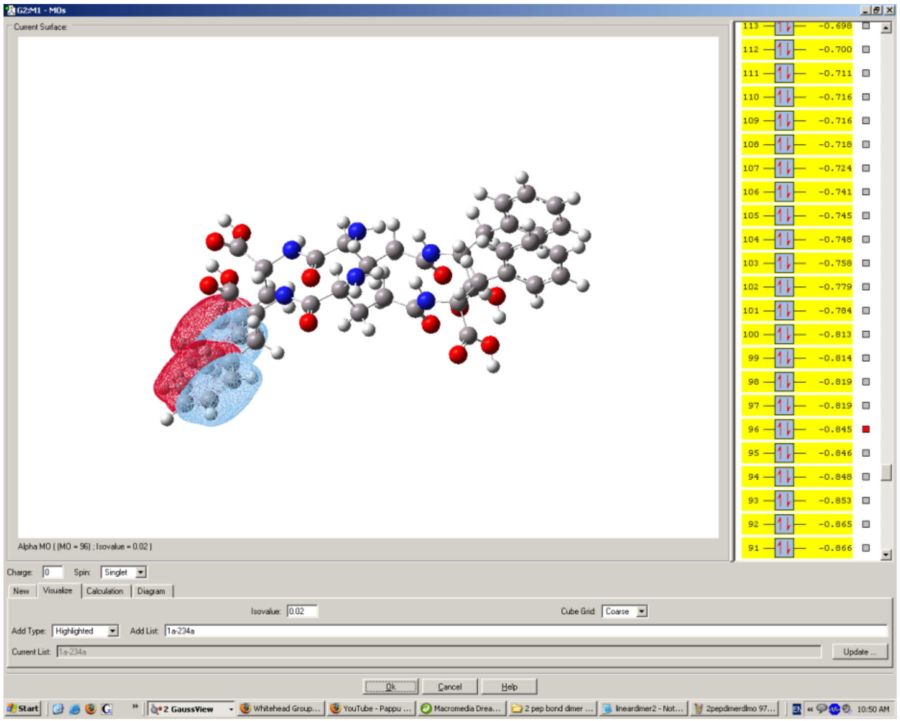
Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 1a-234a

Current List: 1a-234a

May be pie stacking..!



G2:M1 - MOs

Current Surface:

113	-0.698	☐
112	-0.700	☐
111	-0.711	☐
110	-0.716	☐
109	-0.716	☐
108	-0.718	☐
107	-0.724	☐
106	-0.741	☐
105	-0.745	☐
104	-0.748	☐
103	-0.758	☐
102	-0.779	☐
101	-0.784	☐
100	-0.813	☐
99	-0.814	☐
98	-0.819	☐
97	-0.819	☐
96	-0.845	■
95	-0.846	☐
94	-0.848	☐
93	-0.853	☐
92	-0.865	☐
91	-0.866	☐

Alpha MO ((MO = 96) ; Isovalue = 0.02)

Charge: 0 Spin: Singlet

Isovalue: 0.02 Cube Grid: Coarse

Add Type: Highlighted Add List: 1a-234a

Current List: 1a-234a