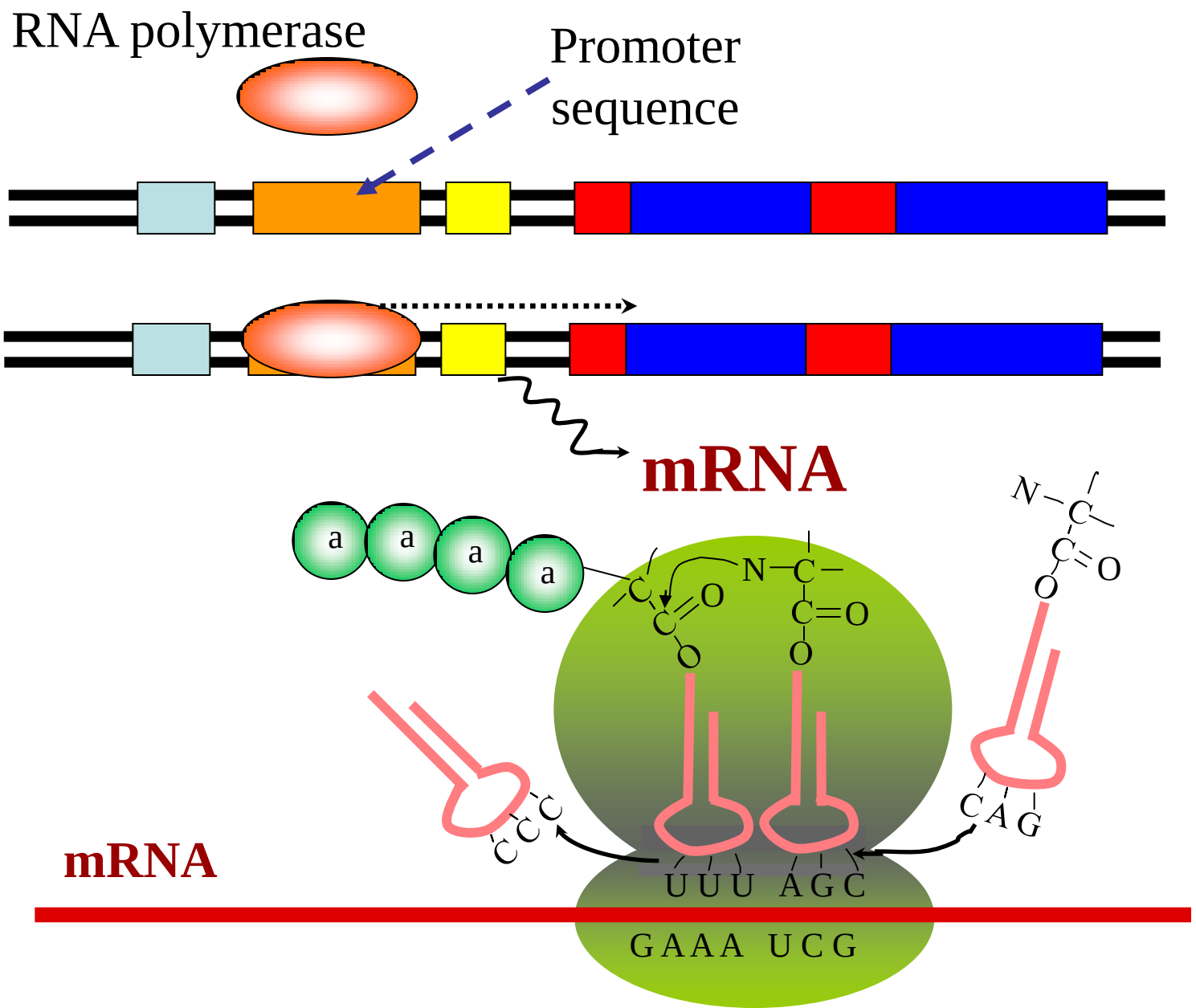
Structural Studies of Non-canonical Base pairs in RNA

Dhananjay Bhattacharyya
Biophysics Division
Saha Institute of Nuclear Physics
Kolkata
dhananjay.bhattacharyya@saha.ac.in

Cellular functions: DNA → RNA → Protein



SINP HOME

HD-RNAS

Hierarchical Database of RNA Structures

Quick Search Option:

(Enter PDB-IDs separated by comma)

Search

Search the Database

Advanced Search

Non-redundant Dataset

Contact us

External links:

Dhananjay Bhattacharyya (Home page)

Other Bioinformatics Resources at Biophysics Division, SINP

Structural/Functional Classification of 1363 RNA Chains

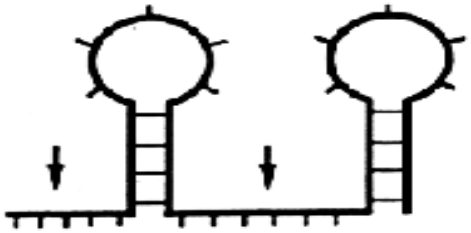
- mRNA *(62 RNA Chain/s)*
- Ribonuclease *(10 RNA Chain/s)*
- Riboswitch *(68 RNA Chain/s)*
- Ribozyme *(141 RNA Chain/s)*
- rRNA *(433 RNA Chain/s)*
- SRP-RNA *(25 RNA Chain/s)*
- tRNA *(193 RNA Chain/s)*
- Unclassified-Function *(431 RNA Chain/s)*

- Viral RNA
- Signaling RNA
- miRNA
- siRNA

Please send your bug reports to: dhananjay.bhattacharyya@saha.ac.in

a. DUPLEXES

b. SINGLE STRANDED REGIONS



c. HAIRPINS

HAIRPIN LOOP

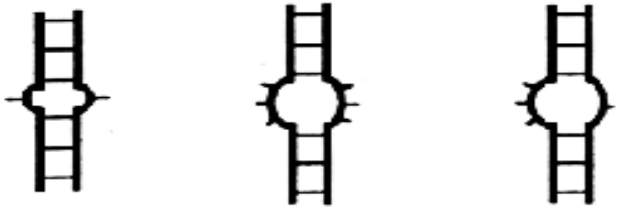
HAIRPIN STEM

d. BULGES

BULGE

SINGLE-BASE BULGE

e. INTERNAL LOOPS

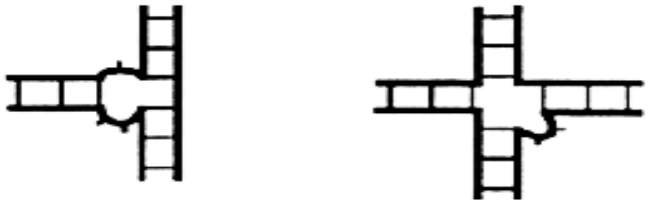


MISMATCH

SYMMETRIC INTERNAL LOOP

ASYMMETRIC INTERNAL LOOP

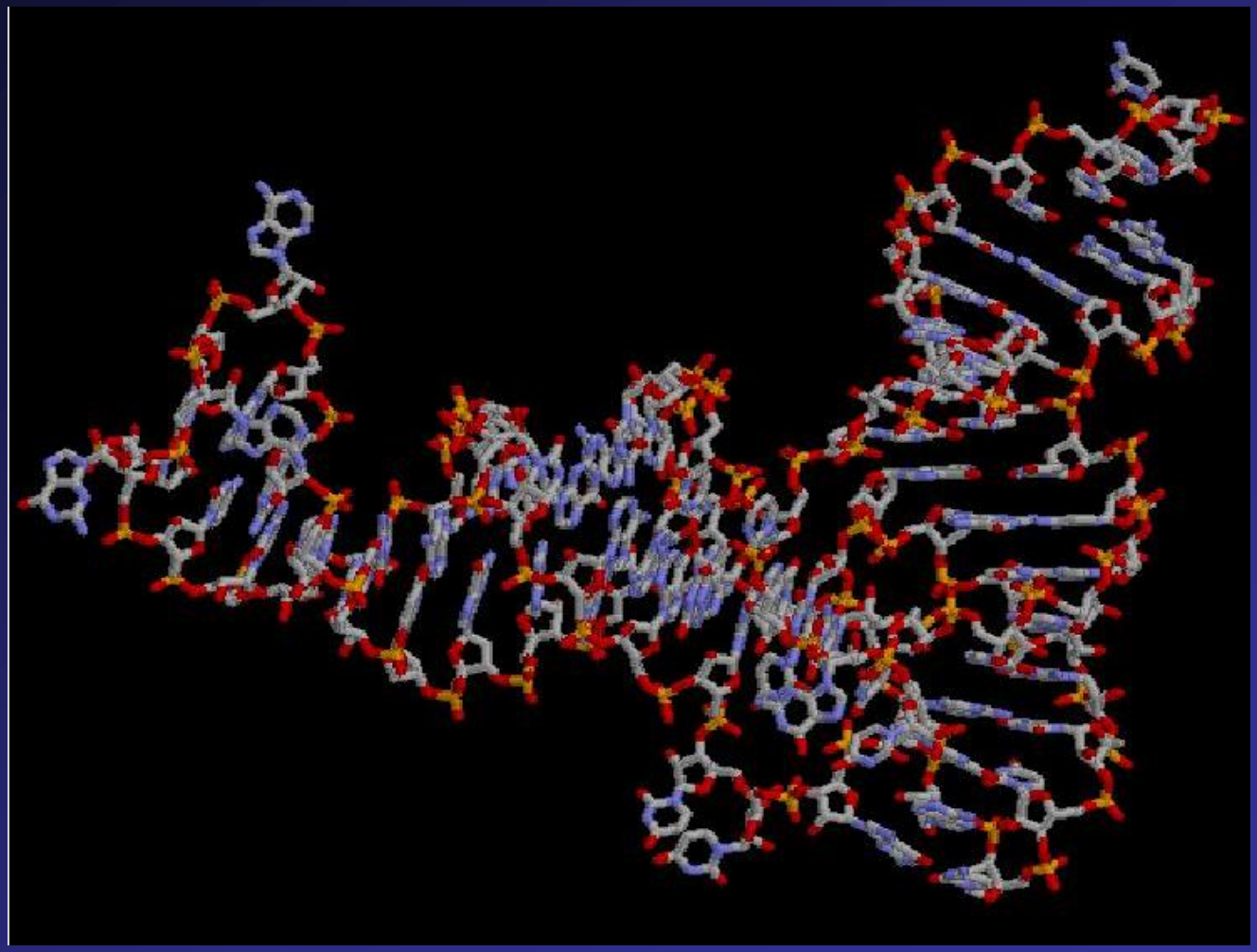
f. JUNCTIONS

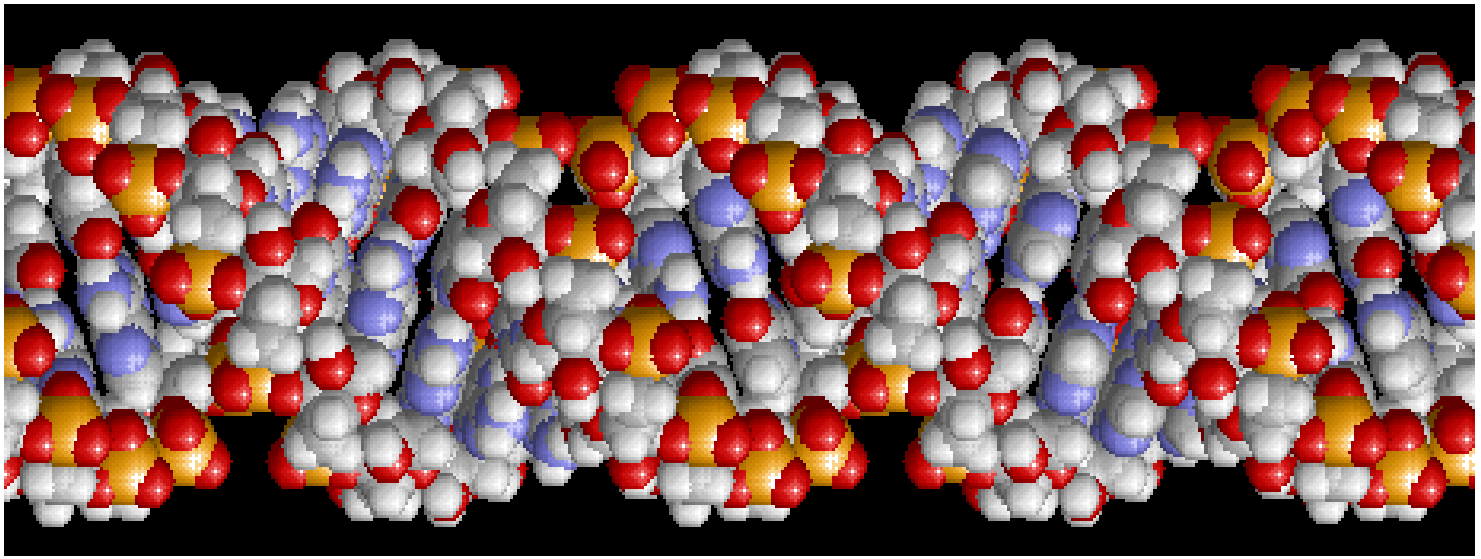


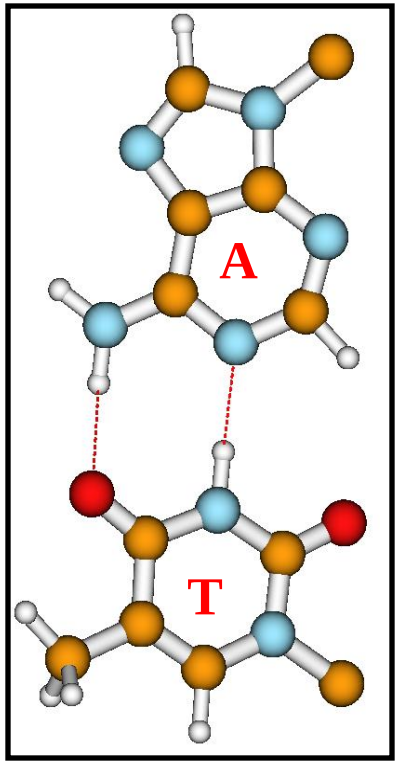
THREE STEM

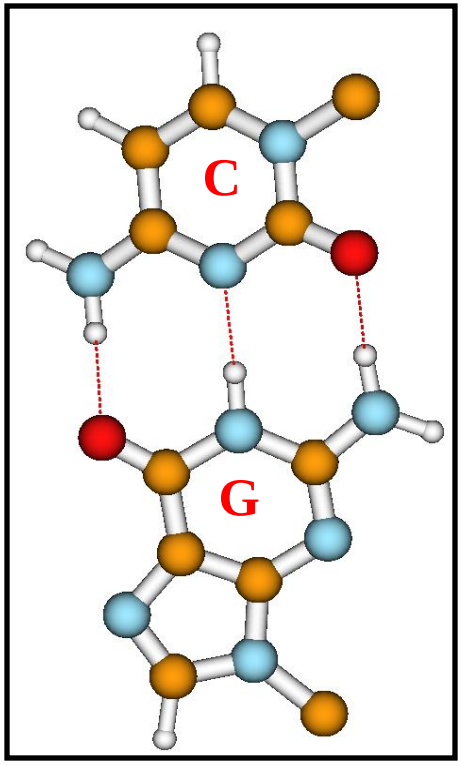
FOUR STEM

$tRNA^{Ile}$ Crystal Structure (PDB ID: 1QU2)

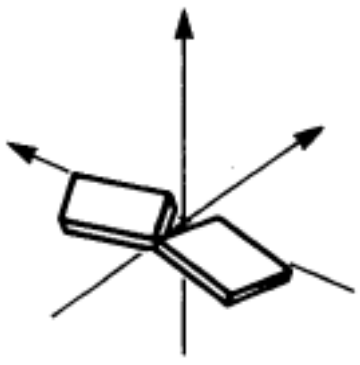




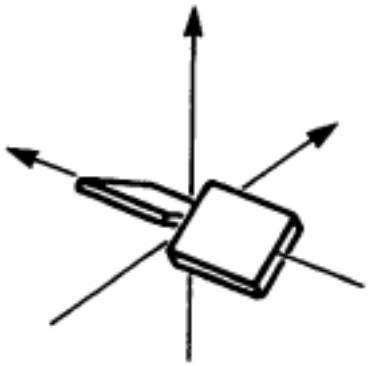

A
T


C
G

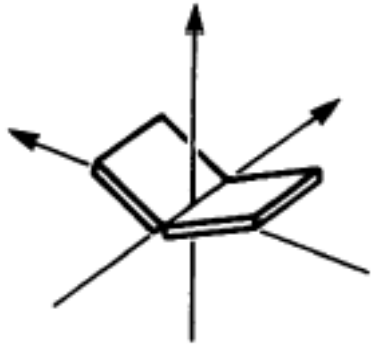
Basepair Parameters (IUPAC/IUB recommendation



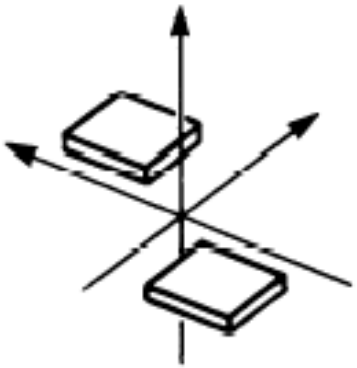
Opening (σ)



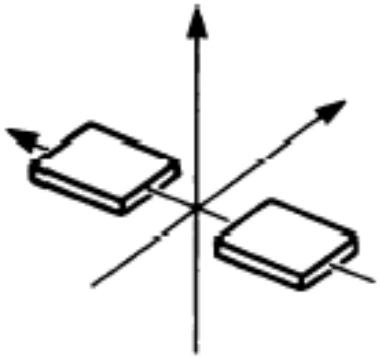
Propeller twist (ω)



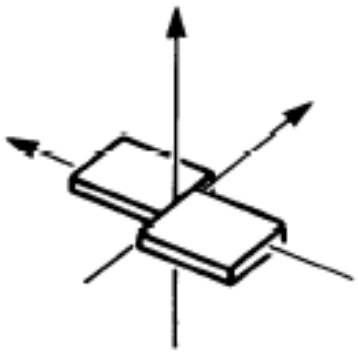
Buckle (κ)



Stagger (Sz)

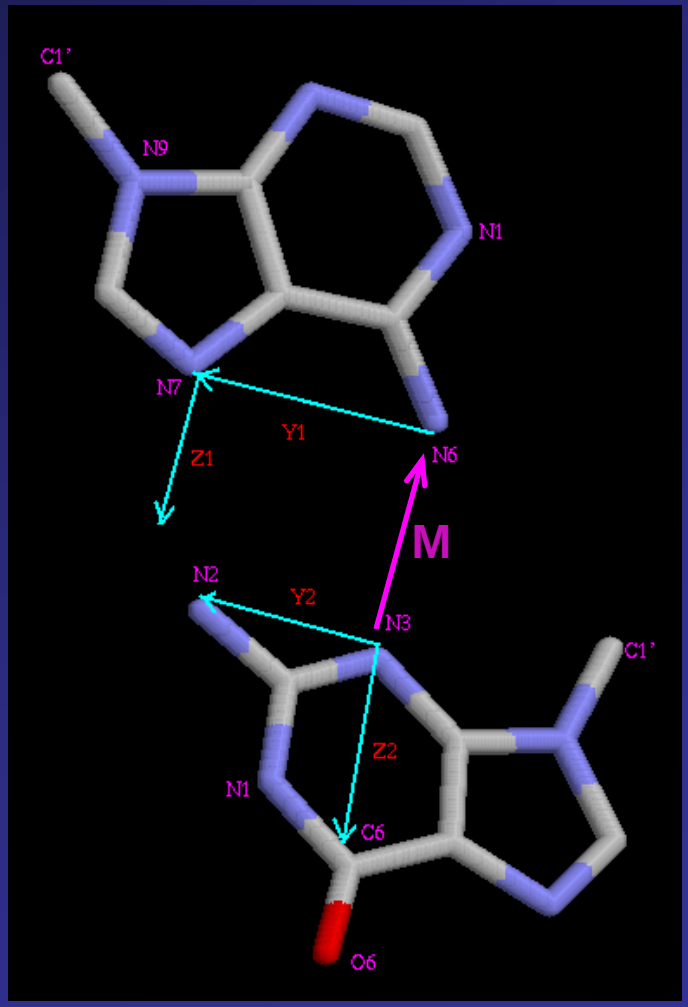


Stretch (Sy)



Shear (Sx)

Base pair parameter definition in NUPARM



Buckle = $2 \sin^{-1}(\mathbf{Zm} \cdot \mathbf{Y_1})$

Opening = $2 \sin^{-1}(\mathbf{Zm} \cdot \mathbf{X_1})$

Propeller = $\cos^{-1} ((\mathbf{X_1} \quad \mathbf{Zm}) \cdot (\mathbf{X_2} \quad \mathbf{Zm}))$

Shear = $-\mathbf{Xm} \cdot \mathbf{M}$

Stagger = $\mathbf{Ym} \cdot \mathbf{M}$

Stretch = $\mathbf{Zm} \cdot \mathbf{M}$

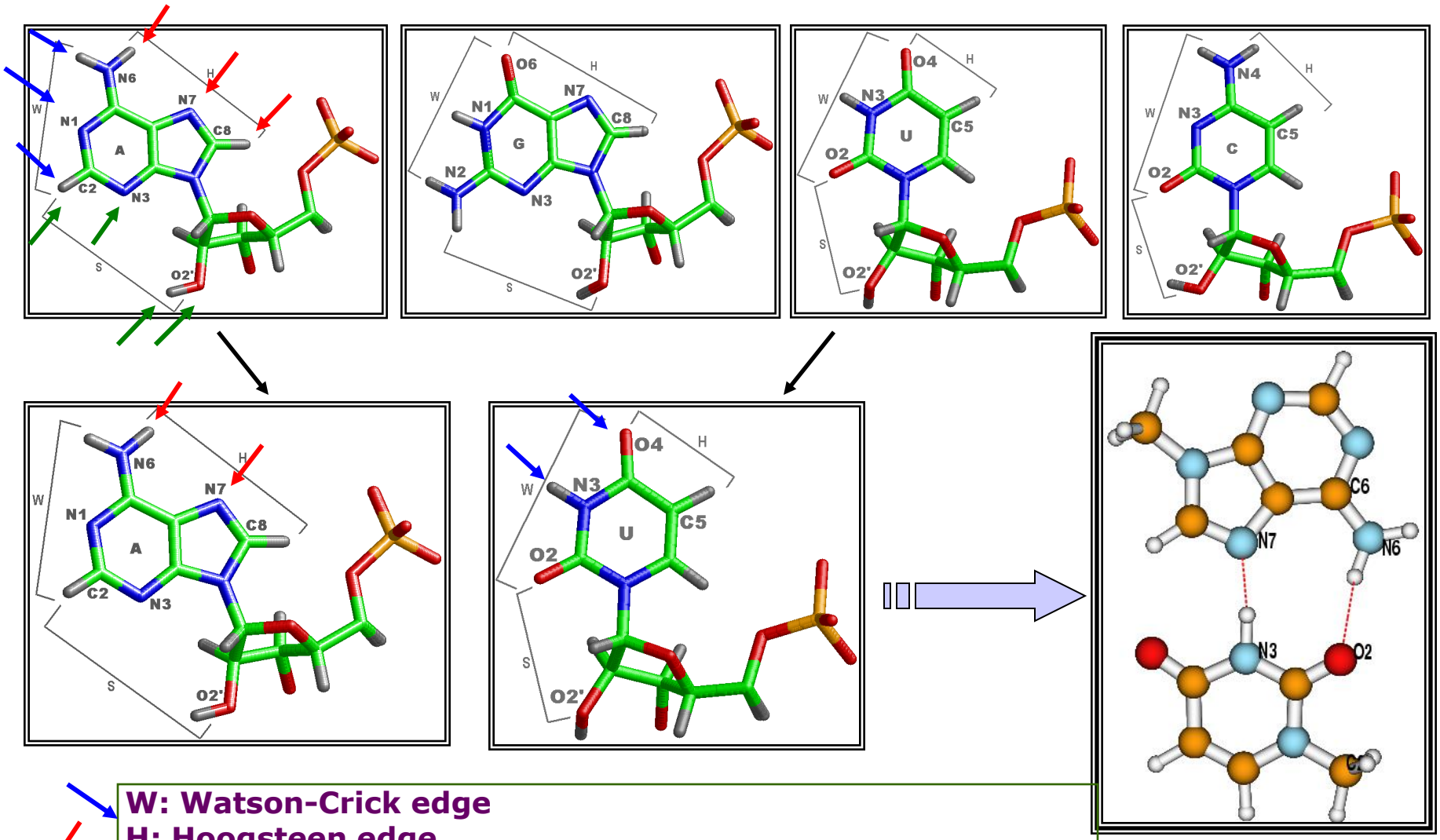
$$\mathbf{Xm} = (\mathbf{X_1} + \mathbf{X_2}) / \mid (\mathbf{X_1} + \mathbf{X_2}) \mid$$

$$\mathbf{Ym} = (\mathbf{Y_1} + \mathbf{Y_2}) / \mid (\mathbf{Y_1} + \mathbf{Y_2}) \mid$$

$$\mathbf{Zm} = \{(\mathbf{X_1} + \mathbf{X_2}) \times (\mathbf{Y_1} + \mathbf{Y_2})\} / \{\mid (\mathbf{X_1} + \mathbf{X_2}) \mid \mid (\mathbf{Y_1} + \mathbf{Y_2}) \mid\}$$

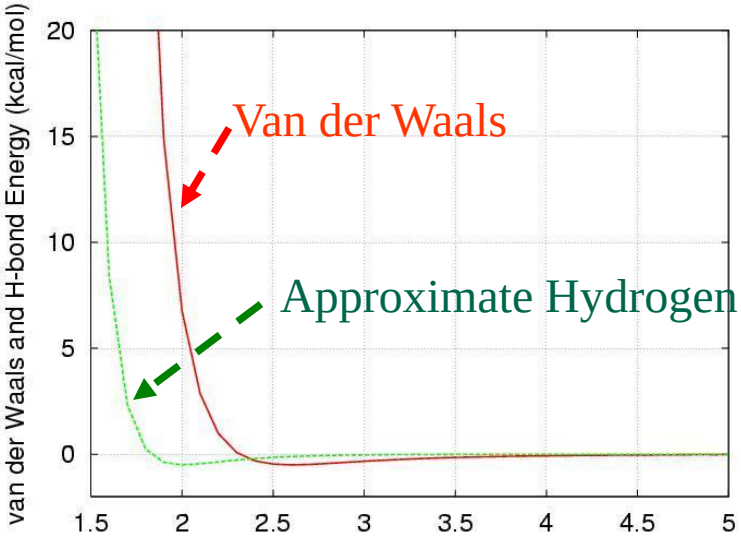
S. Mukherjee, M. Bansal; D. Bhattacharyya (2006) *J. Comp. Aided Mol. Des.* **20**; 629

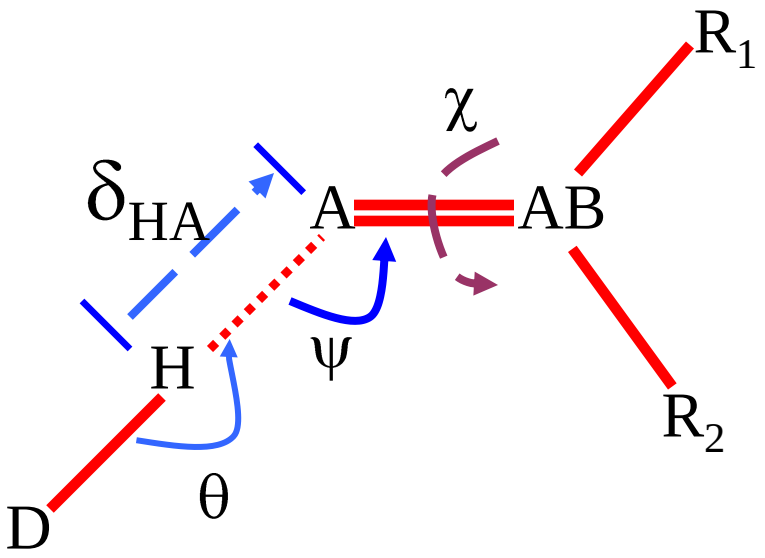
Non-canonical Basepairing

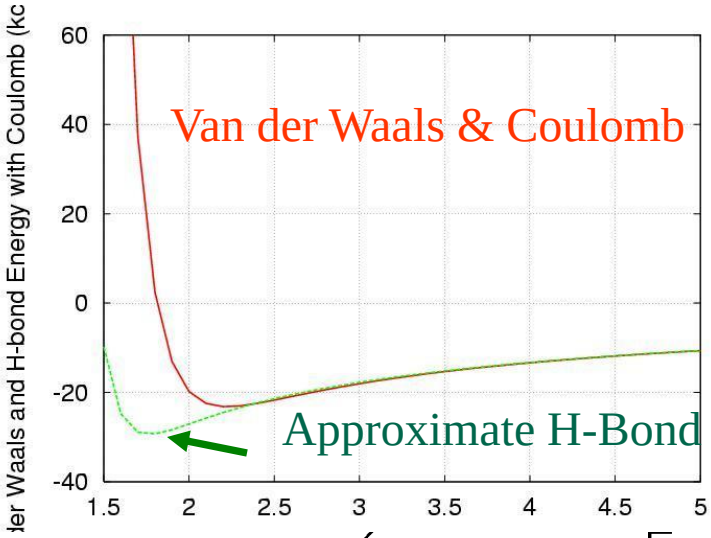


A:U H:WT

W: Watson-Crick edge
H: Hoogsteen edge
S: Sugar edge
w/h/s: Involves weak C-H…O/N interaction
C/T: *Cis*- or *Trans*-orientation of the two glycosidic bonds

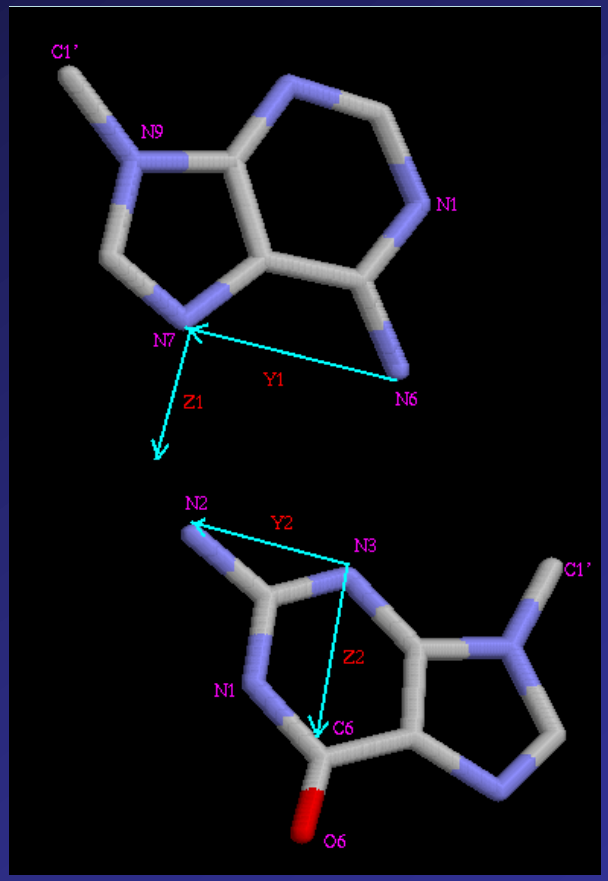






$$V(r)=\sum_{H-bonded}\left(\frac{q_i q_j}{\varepsilon r_{ij}}+d_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12}-\left(\frac{R_{min,ij}}{r_{ij}}\right)^{10}\right]\right)f(\theta,\psi,\chi)$$

Base Pair Finder



- ✓ Took a base edge
- ✓ Identify the H-bonding centers (N3G & N2G)
- ✓ Look for H-bond partner through distance calculation (N6A & N7A)
- ✓Calculate pseudo-angles (such as C6G-N3G-N6A, N3G-N6A-N1A, N1G-N2G-N7A, N2G-N7A-N9A in figure) for planarity
- ✓ Confirm orientation through angle calculation
- ✓Calculate $E=\Sigma_i(d_i-3.0)^2 + \frac{1}{2}\Sigma_k(\theta_k-\pi)^2$; i are for two H-bond distances and k are for four pseudo angles

Gives rise to:
6959 A:U W-W(C);
21965 G:C W-W(C) and
2786 G:U W-W(C) base pairs

Das, Mukherjee, Mitra & Bhattacharyya (2006) *J Biomol Struct Dynam, 24, 149-161*

Applications Places System 1:04 PM

Basepairs involving Hoogsteen edge of Adenine and Sugar edge of Guanine are shown - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.saha.ac.in/biop/www/db/pairinfo/AH_GS_cis_trans.html

Go den linux download

Red Hat Red Hat Magazine Red Hat Network Red Hat Support

Basepairs involving Hoogsteen edge of Adenine (H) and Sugar edge of Guanine (S) are shown. Highlighted examples are found from RNA crystal structures obtained from PDB.

Base pairs stabilized by N-H…N/O type of hydrogen bonds only

Basepair (type & orientation)	*Frequency*	*Example*	*Hydrogen bonding atoms and their precursors*
A:G H:S Trans	2303	Residue Nos. 21(B)-44(B) PDB ID 1MZP	N1-N6 ..... N3-C6 N9-N7 ..... N2-N1

Base pairs stabilized by C-H…N/O interactions along with N-H…N/O hydrogen bond

Basepair (type & orientation)	*Frequency*	*Example*	*Hydrogen bonding atoms and their precursors*
A:G h:s Cis	11	Observed but not shown	C5-N7 ..... N2-N1

Done

Terminal | Basepairs invo... | [xterm] | Downloads | xterm | xterm

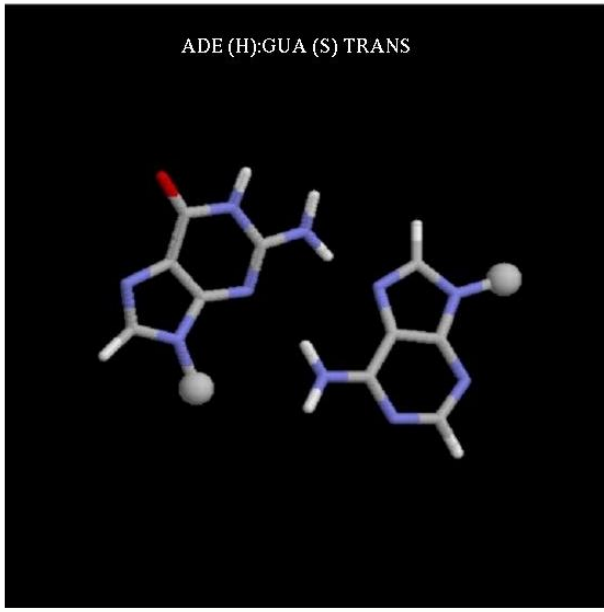
Applications Places System

AG_HST_1MZP_21B_44B.jpg (JPEG Image, 552

File Edit View Go Bookmarks Tools Help

http://www.saha.ac.in/biop/www/db/pairinfo/AG_HS

Red Hat Red Hat Magazine Red Hat Network Red Hat Support



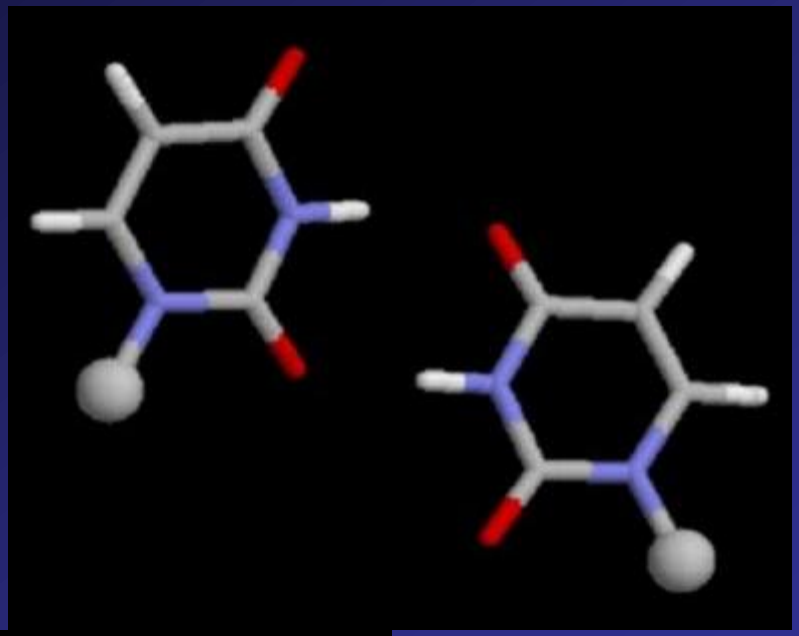
Done

Terminal | AG_HST_1MZP... | [xterm] | Download

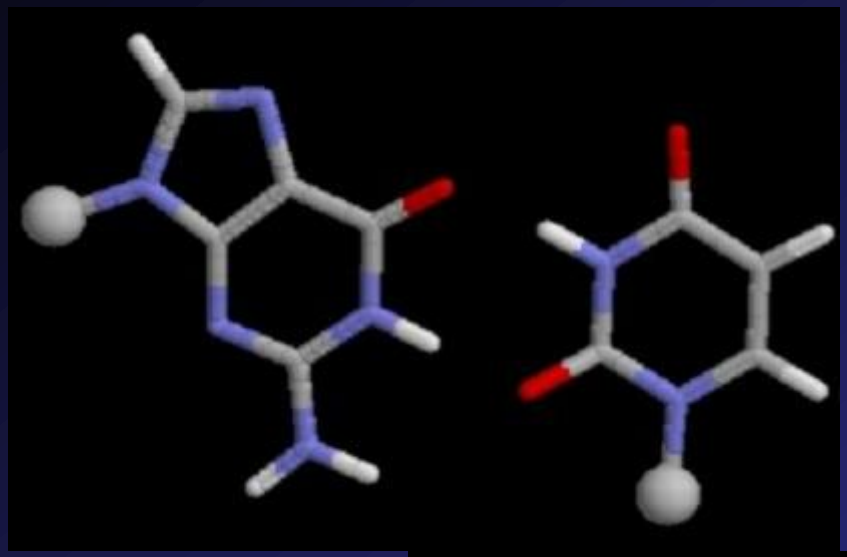
Different types of possible base pairs occurring between nucleotides are tabulated. Numbers of geometries that can occur between any two base edges are hyper-linked. Examples of specific base pairing type obtained from PDB structures and involving regular bases are prepared with RASMOL. Those involving protonated bases are prepared with MOLDEN. The base pairs and base triples are detected by BPFIND and from a set of structures determined by X-ray crystallography. The outputs are created in a suitable way so that their structural parameters can be calculated by NUPARM (the older version of NUPARM can also be downloaded from PDB.

	Ade W	Ade H	Ade S	GuaW	Gua H	Gua S	Cyt W	Cyt H	Cyt S	Ura W	Ura H	Ura S
Ade W	two	two	two	one	two	two	four		two	two	two	two
Ade H		two	two	one	-	two	three		one	two	two	one
Ade S			one	one	two	one	one			two	one	
Gua W				one	two	one	three		two	two	one	two
Gua H					one	three	two	two				
Gua S						one	three		one	two	one	one
Cyt W							three	two	two	four	one	two
Cyt H									two			two
Cyt S												

U:U W:W Cis (84)

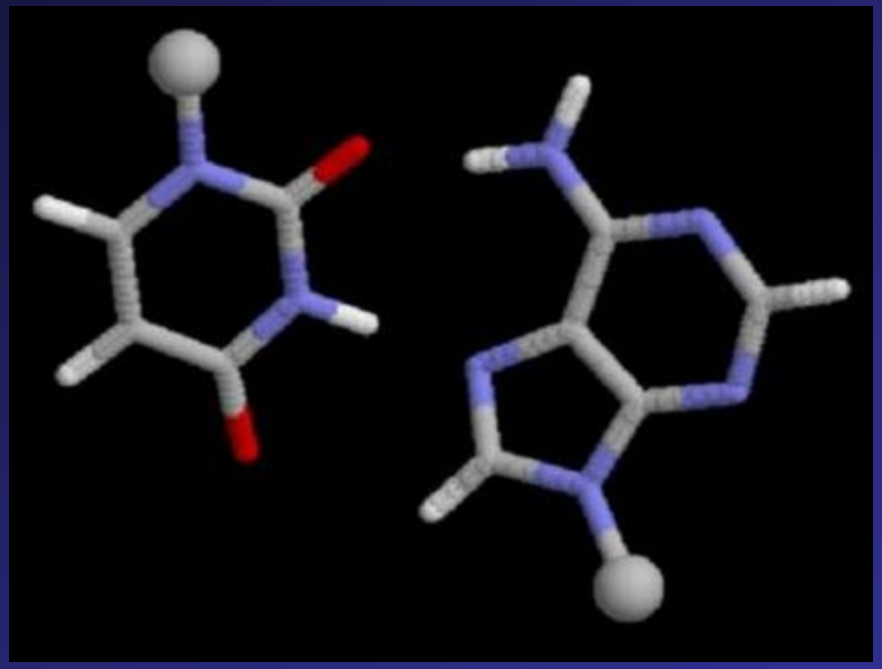


G:U W:W Cis (846)

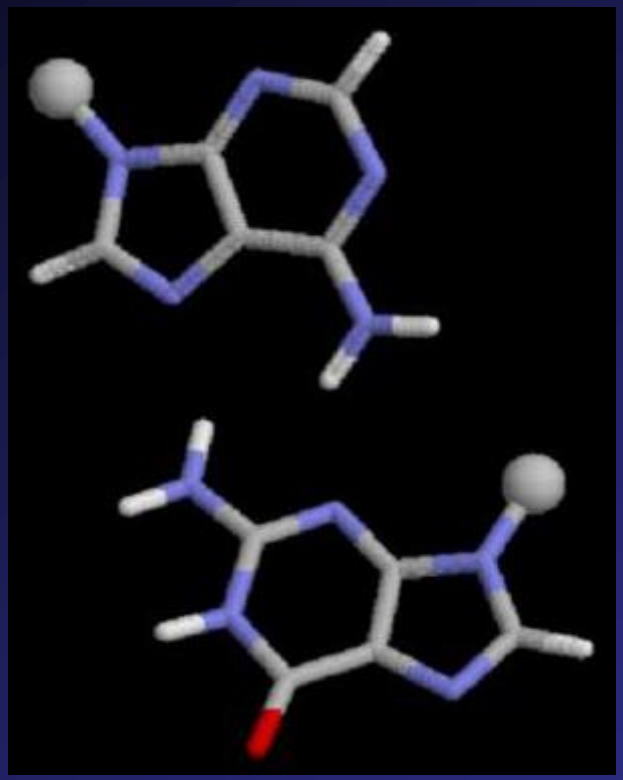


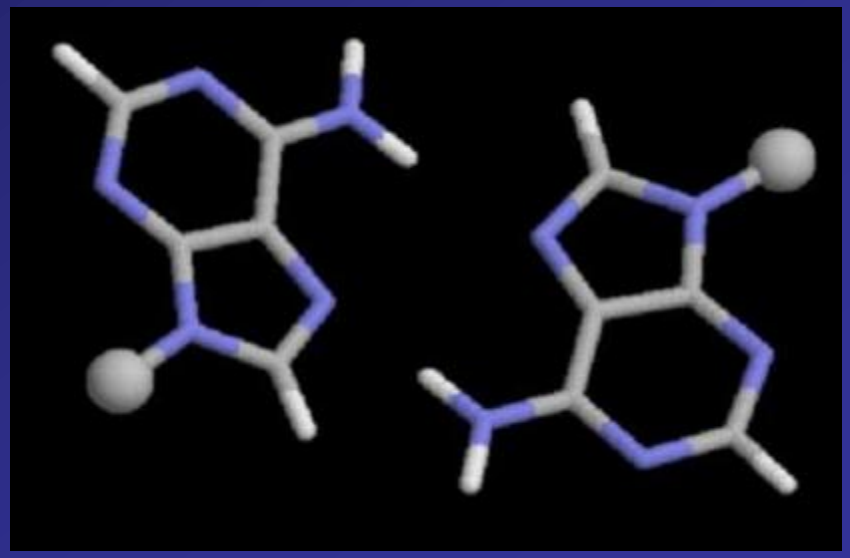
A:G W:W Cis (150)

A:U H:W Trans (410)



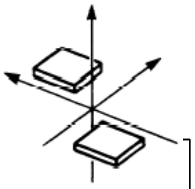
A:G H:S Trans (558)

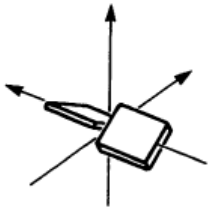


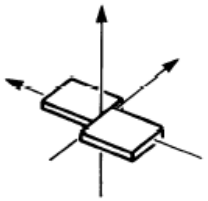


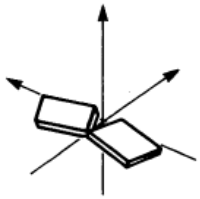
A:A H:H Trans (109)

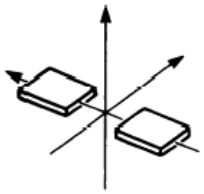
Property of good and stable Base Pair

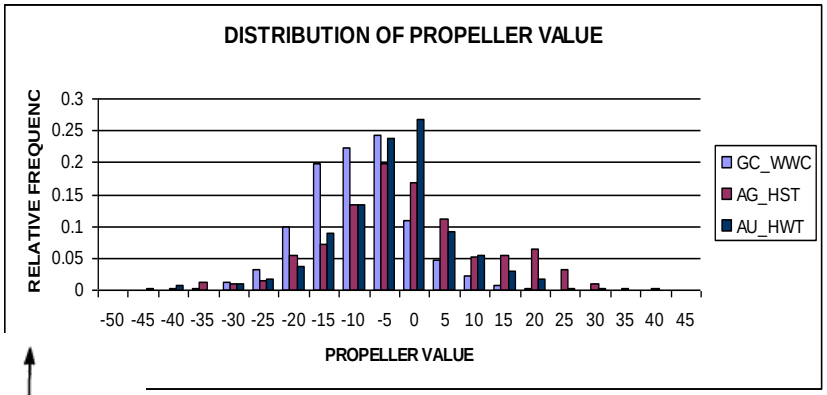


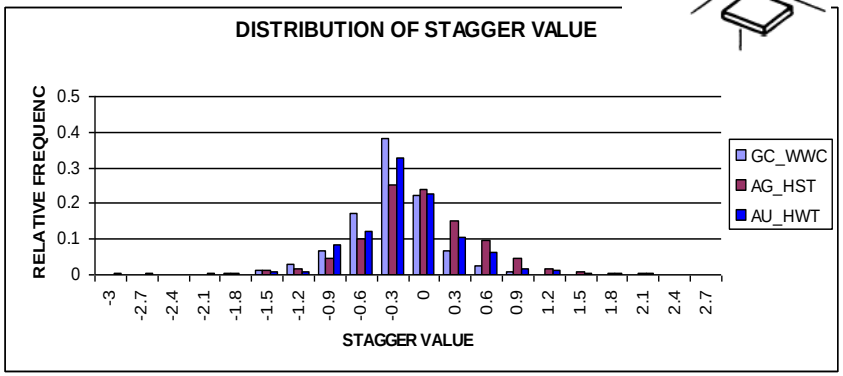


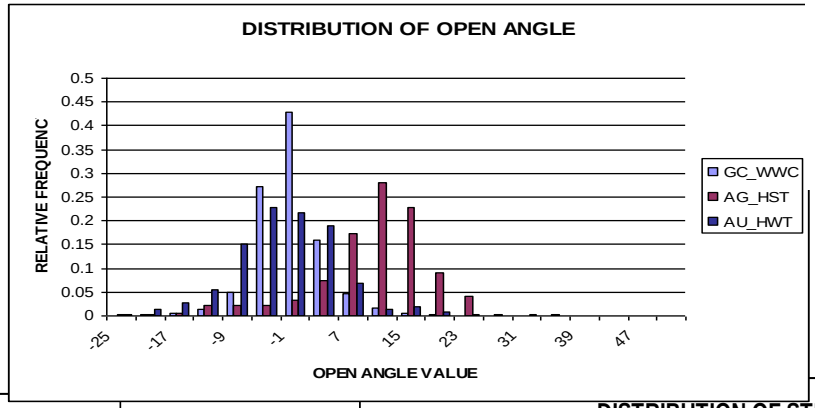


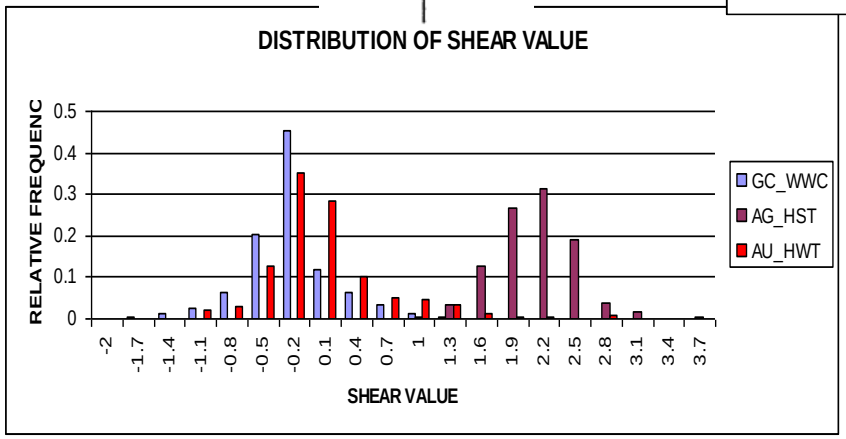












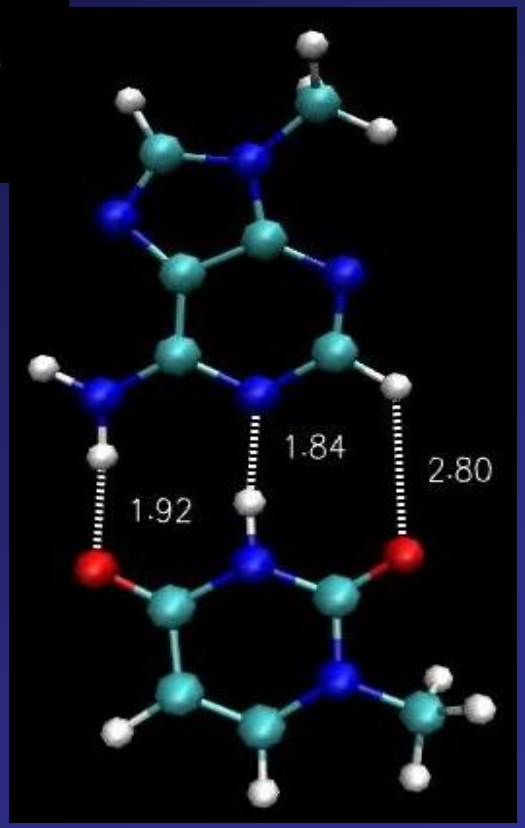
Geometry Optimization by different Methods

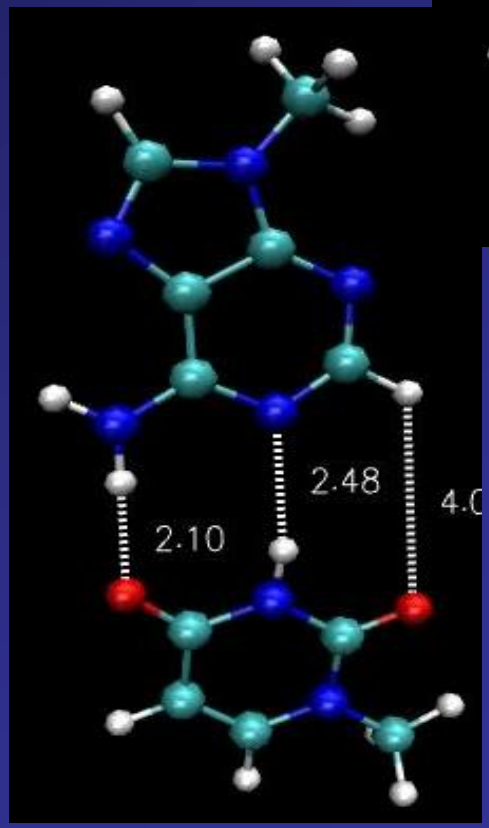
- Selected structures of BPs from PDB

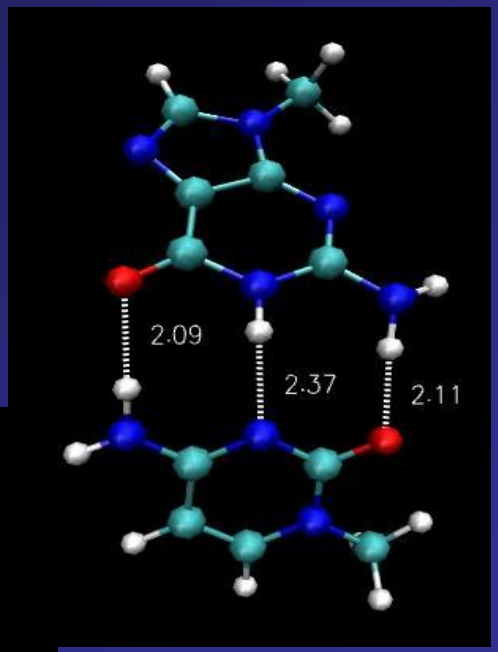
- Optimized Structures by B3LYP/6-31G**
- Optimized by MP2/6-31G**
- Optimized by HF/CC-pVDZ
- Optimized by GGA-BW91/DZP
- Optimized by semi-empirical methods (AM1, PM3)
- Optimized by AMBER force-field

- Compared structure and dynamics with Molecular Dynamics Simulations

Failure of AM1 (most popular semi-empirical method)

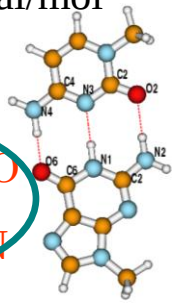






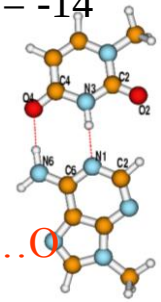
Strengths of different H-bonds from 33 non-canonical Base Pairs

G:C W:W C
ΔE = -26 kcal/mol



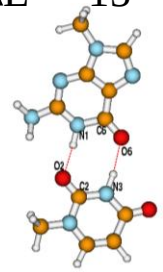
2=>NH…O
1=>NH...N

A:U W:W C
ΔE = -14



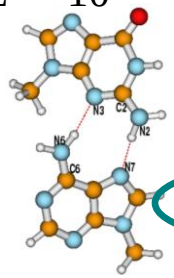
1=>NH…O
1=>NH…N

G:U W:W C
ΔE = -15



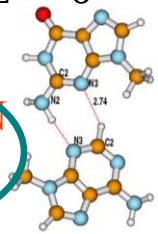
2=>NH...O

A:G H:S T
ΔE = -10



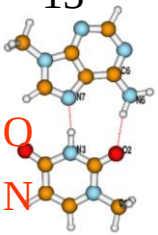
2=>NH...N

A:G s:s T
ΔE = -6



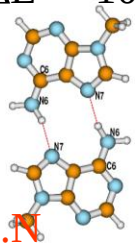
1=>NH…N
1=>CH...O

A:U H:W T



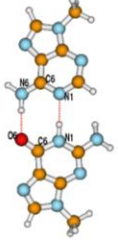
1=>NH...O
1=>NH...N

A:A H:H T
ΔE = -10



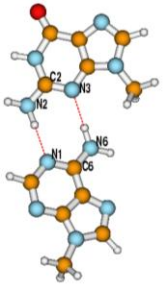
2=>NH...N

G:A W:W C
ΔE = -15



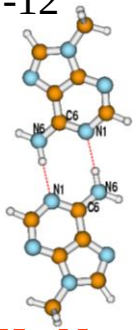
1=>NH...O
1=>NH…N

G:A S:W T
ΔE = -11



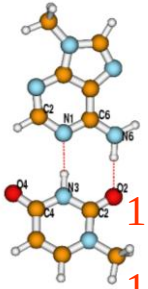
2=>NH...N

A:A W:W T
ΔE = -12



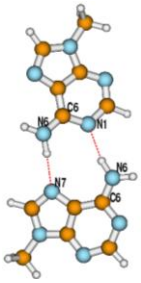
2=>NH...N

A:U W:W T
ΔE = -13



1=>NH..O
1=>NH...N

A:A H:W T
ΔE = -11



1=>NH...O
1=>NH…N

Ray, Panigrahi, Bhattacharyya & Bhattacharyya (2008) J. Phys. Chem. **B112, 3786.**

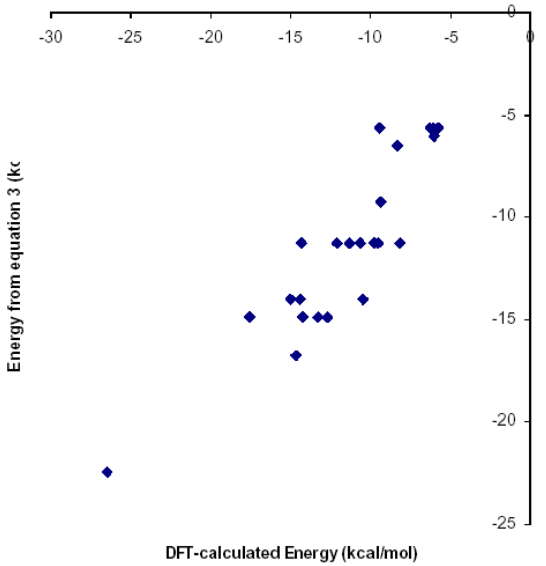
Considered Energy components, E^{NHO}, E^{NHN}, etc are additive. Additional stabilities, δ_i may come from van der Waals, dipole-dipole etc interactions.

$$E_{\text{int}}^{i} = n_i^{NHO} E^{NHO} + n_i^{NHN} E^{NHN} + n_i^{OHN} E^{OHN} + n_i^{CHO} E^{CHO} + n_i^{CHN} E^{CHN} + \delta_i$$

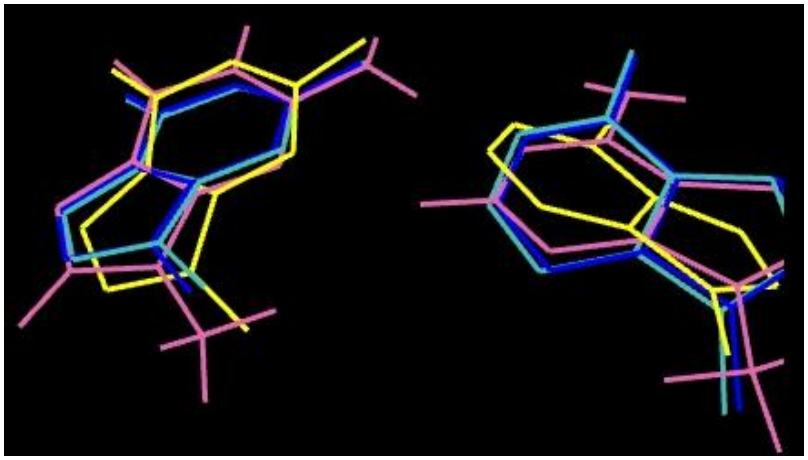
Least Squares Fit indicates δ_i, errors should be smallest for best Fit

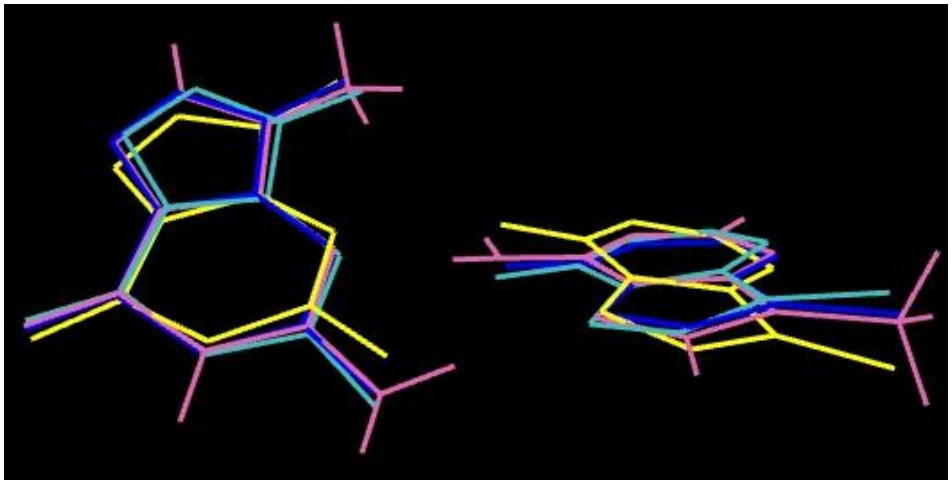
$$\sum_i \delta_i^2 = \sum_i \left(E_{\text{int}}^{i} - n_i^{NHO} E^{NHO} - n_i^{NHN} E^{NHN} - n_i^{OHN} E^{OHN} - n_i^{CHO} E^{CHO} - n_i^{CHN} E^{CHN} \right)^2$$

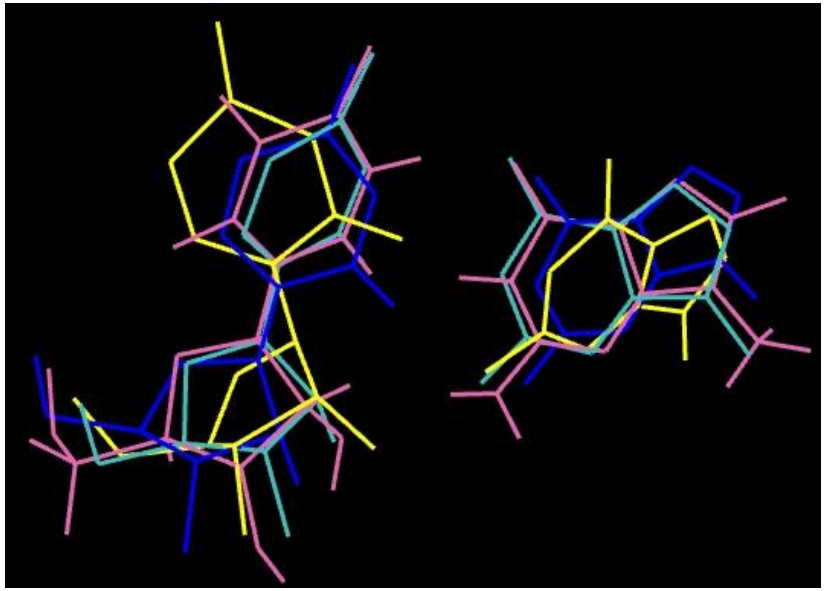
Type of H-bond	ΔE (kcal/mol)
N-H…O	-7.82
N-H…N	-5.62
O-H…N	-6.89
C-H…O	-1.33
C-H…N	-0.67



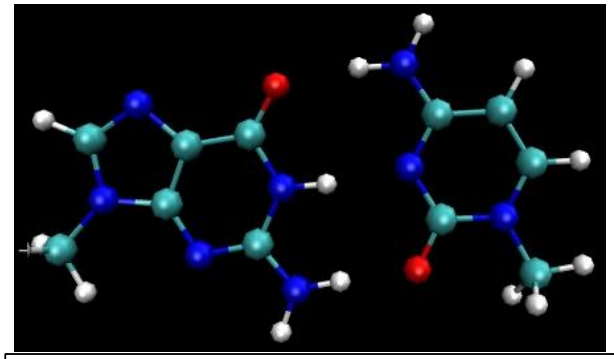
Roy, Bhattacharyya, Panigrahi, Bhattacharyya, (2008) J. Phys. Chem. B B112, 3786

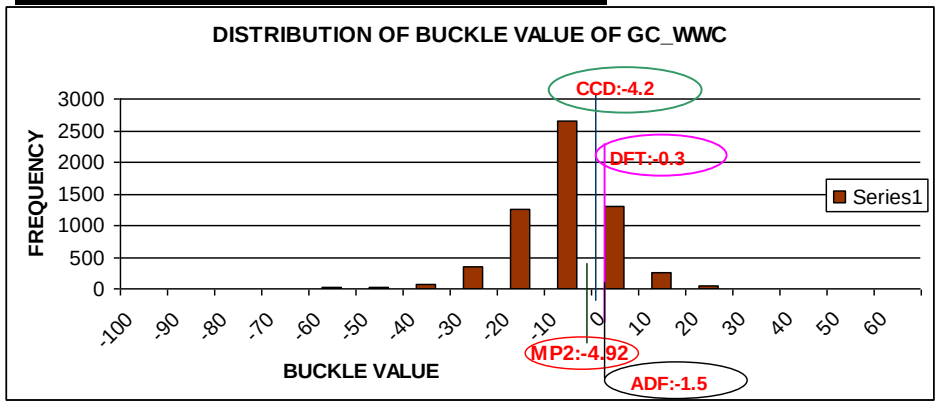


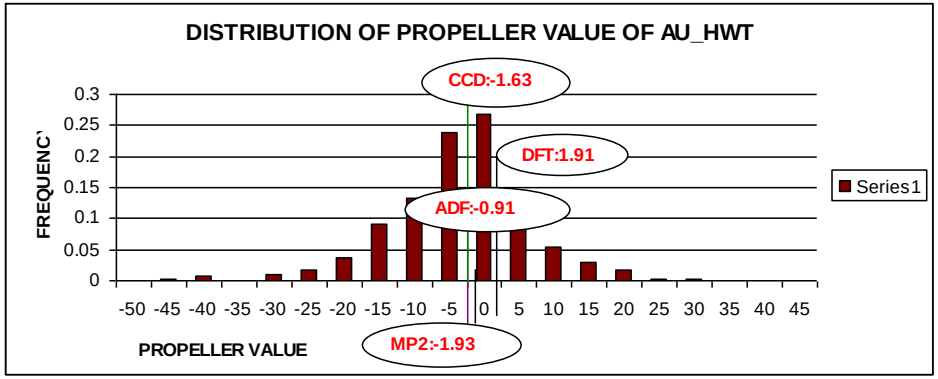


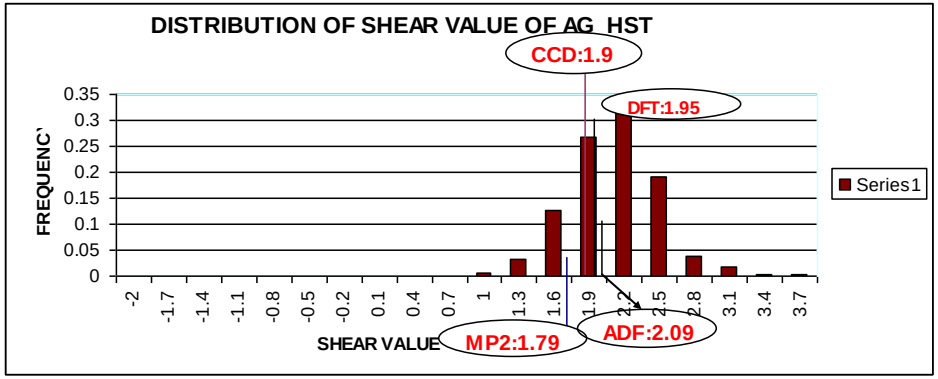


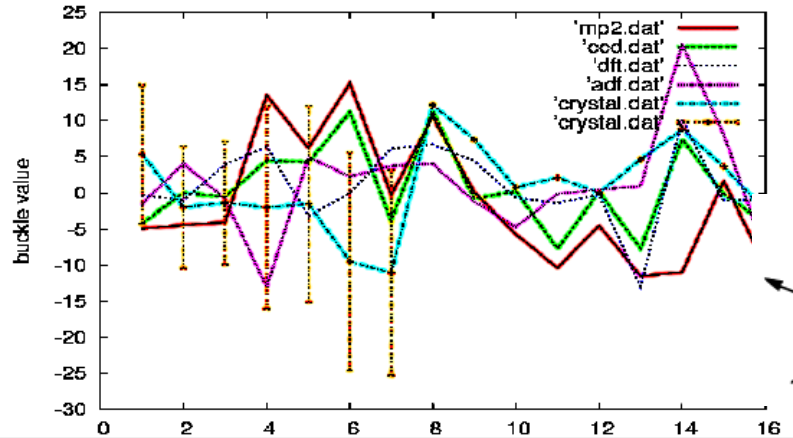
Comparison with X-ray

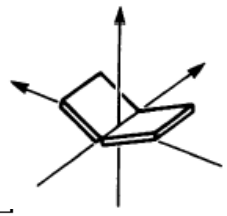


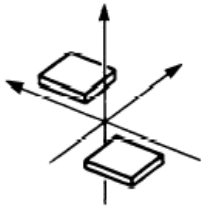


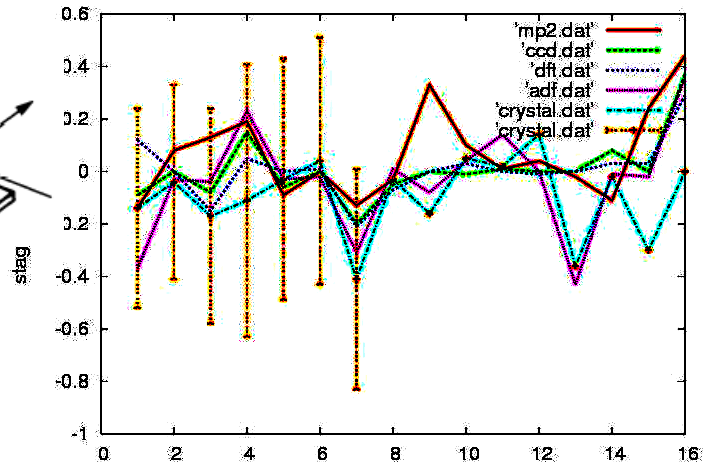


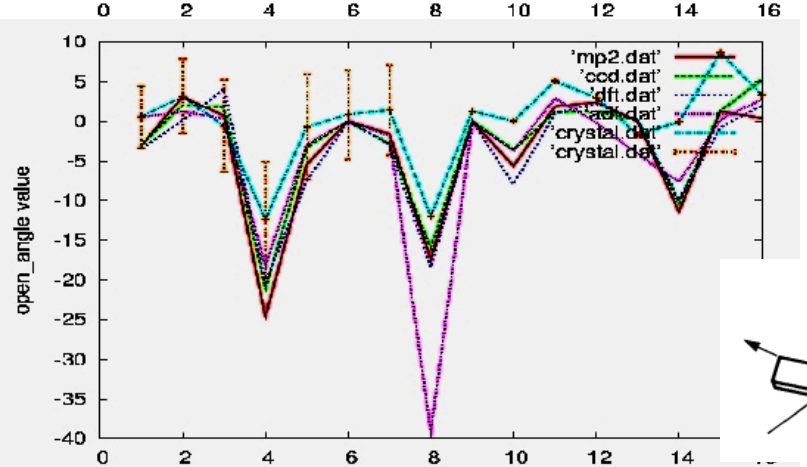


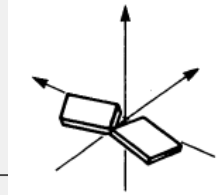

buckle value
'mp2.dat'
'ccd.dat'
'dft.dat'
'adf.dat'
'crystal.dat'
'crystal.dat'

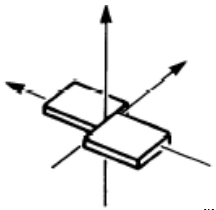


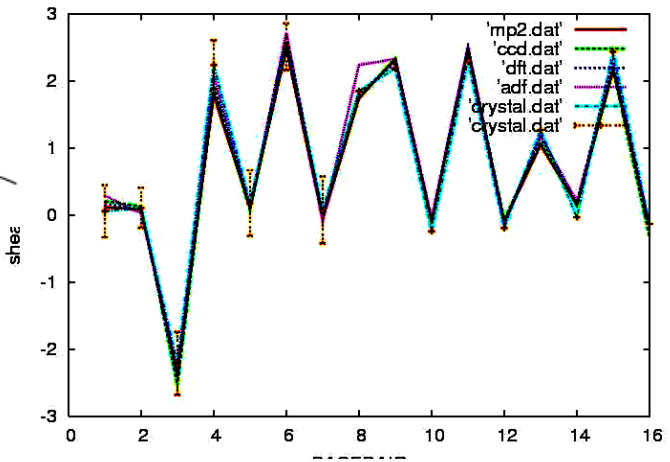


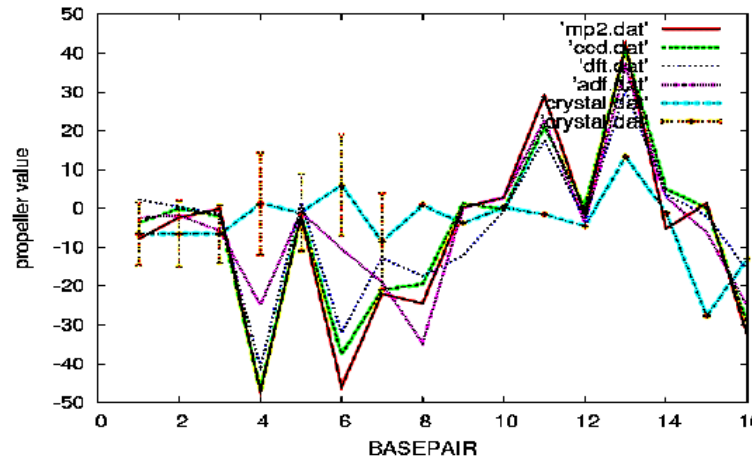

stag
'mp2.dat'
'ccd.dat'
'dft.dat'
'adf.dat'
'crystal.dat'
'crystal.dat'

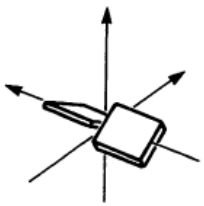

open_angle value
'mp2.dat'
'ccd.dat'
'dft.dat'
'adf.dat'
'crystal.dat'
'crystal.dat'

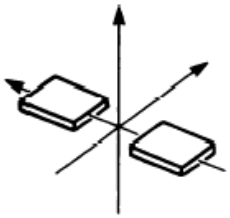


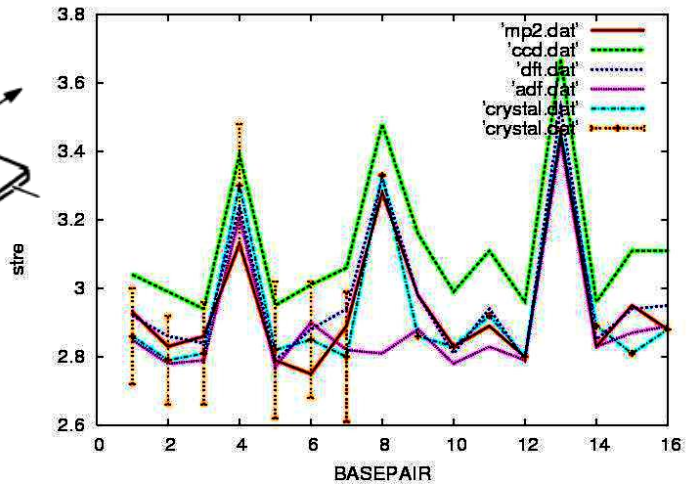



shea
'mp2.dat'
'ccd.dat'
'dft.dat'
'adf.dat'
'crystal.dat'
'crystal.dat'

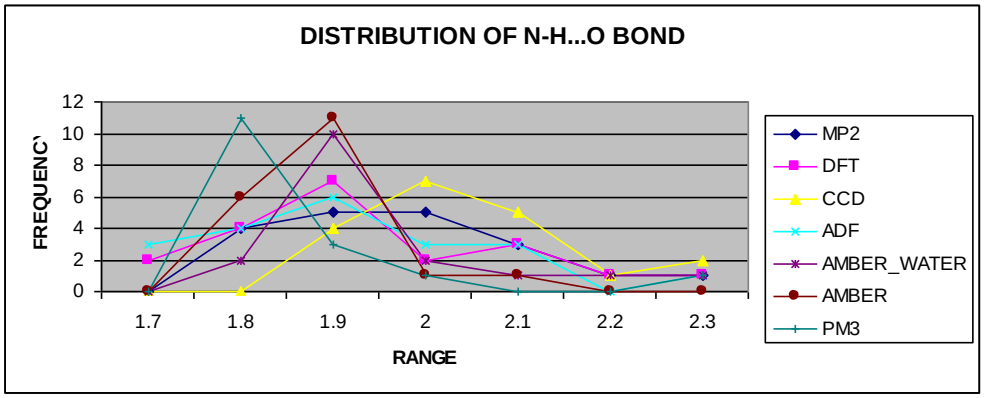

propeller value
'mp2.dat'
'ccd.dat'
'dft.dat'
'adf.dat'
'crystal.dat'
'crystal.dat'
BASEPAIR

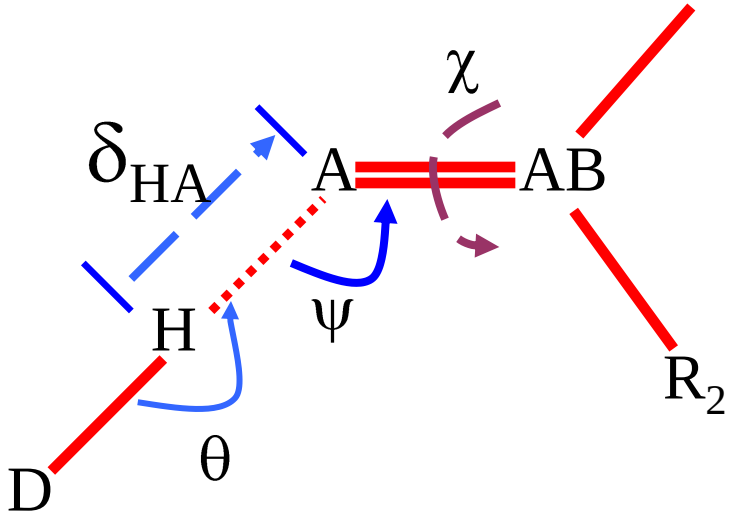


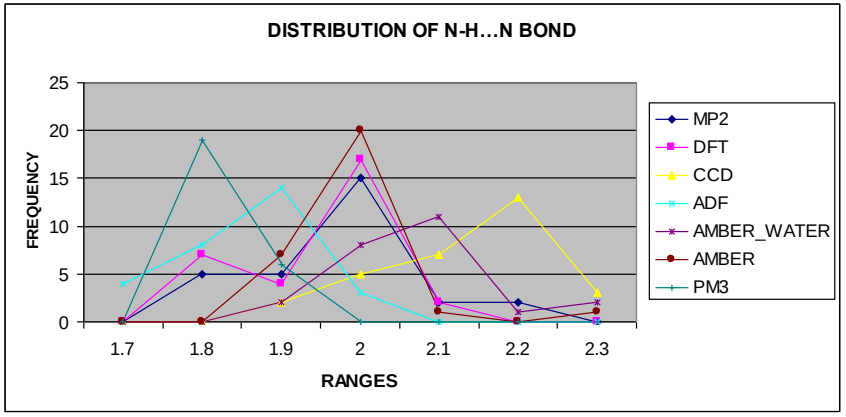


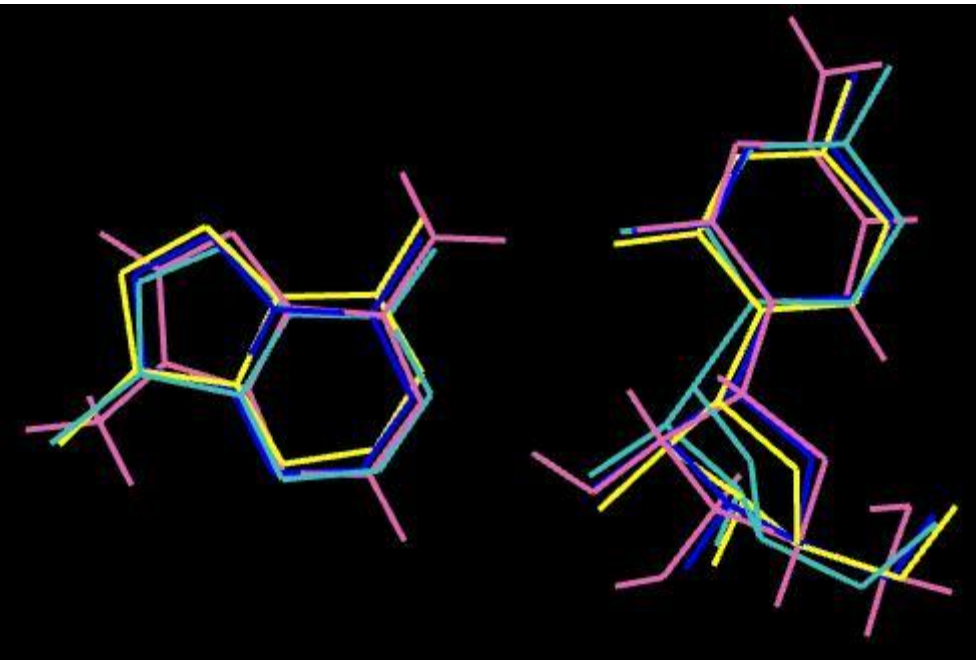

stre
'mp2.dat'
'ccd.dat'
'dft.dat'
'adf.dat'
'crystal.dat'
'crystal.dat'
BASEPAIR

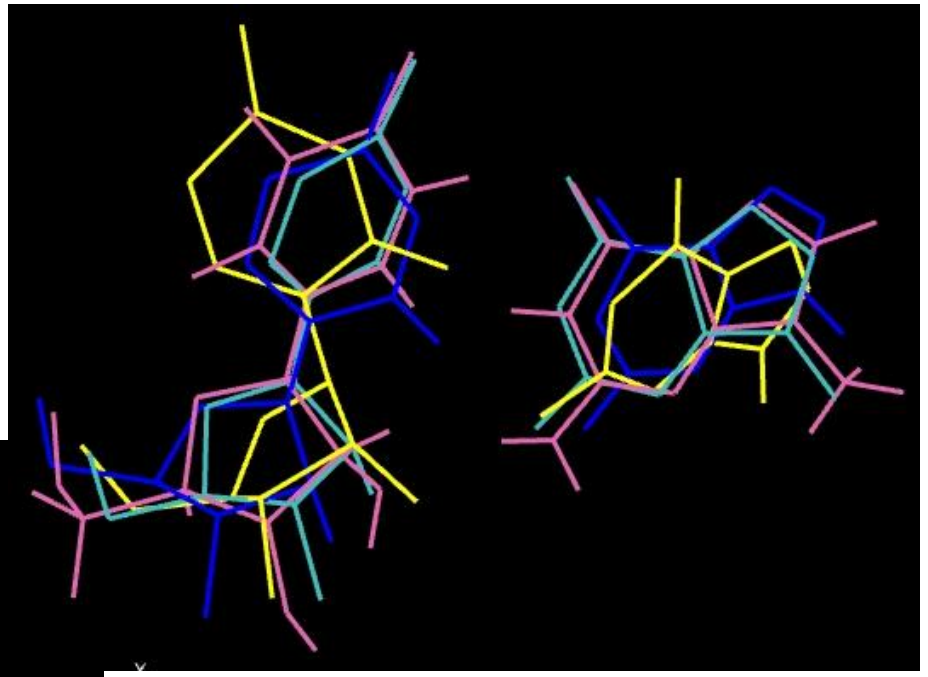
Hydrogen Bond Geometries

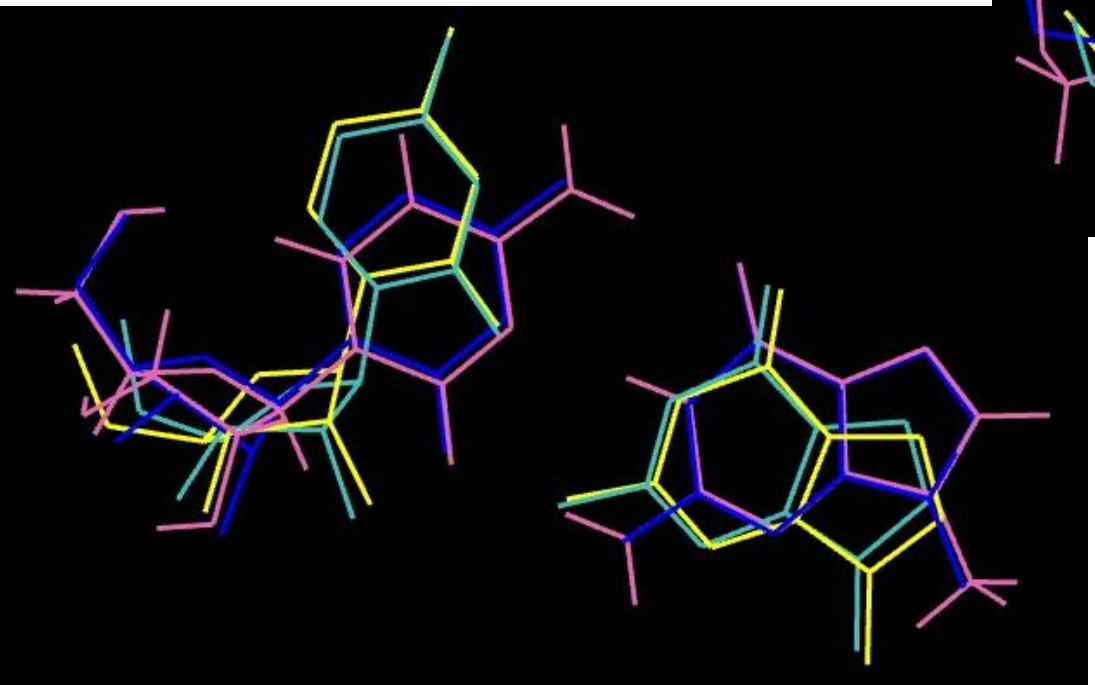




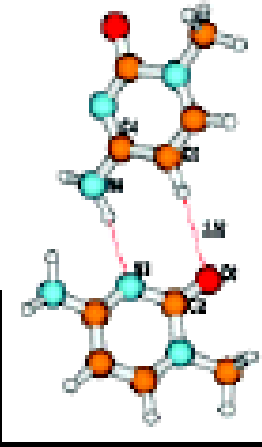


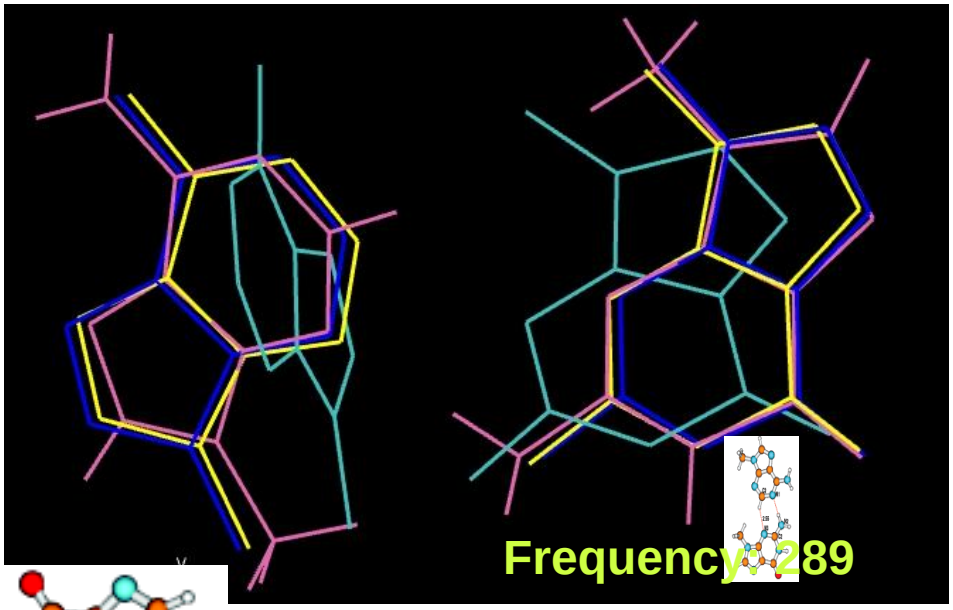


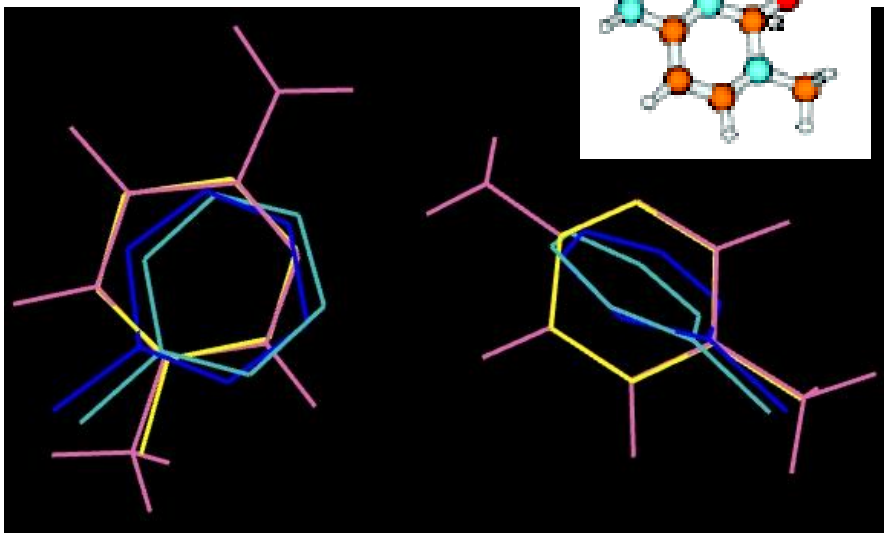


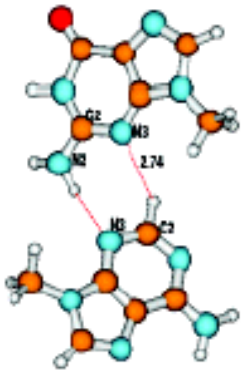


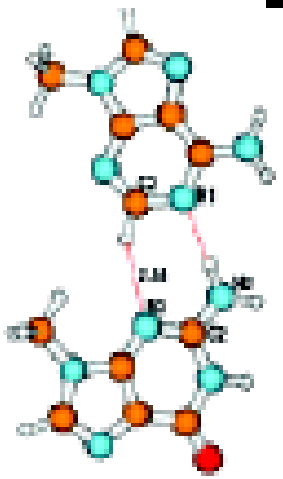
Non-polar Base Pairs

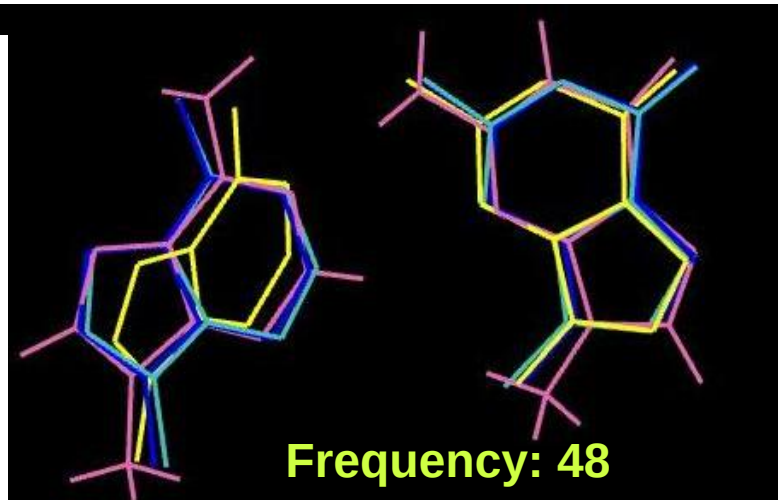




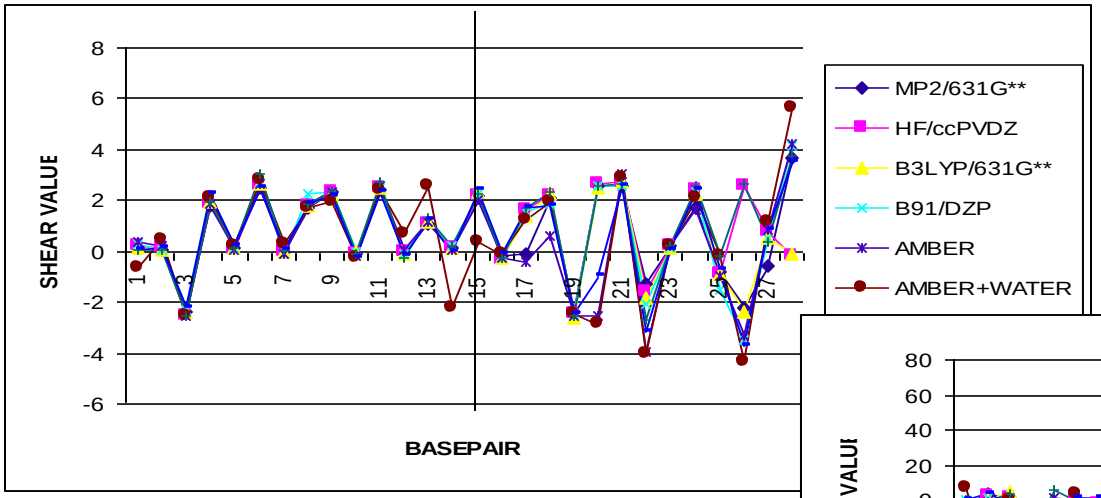


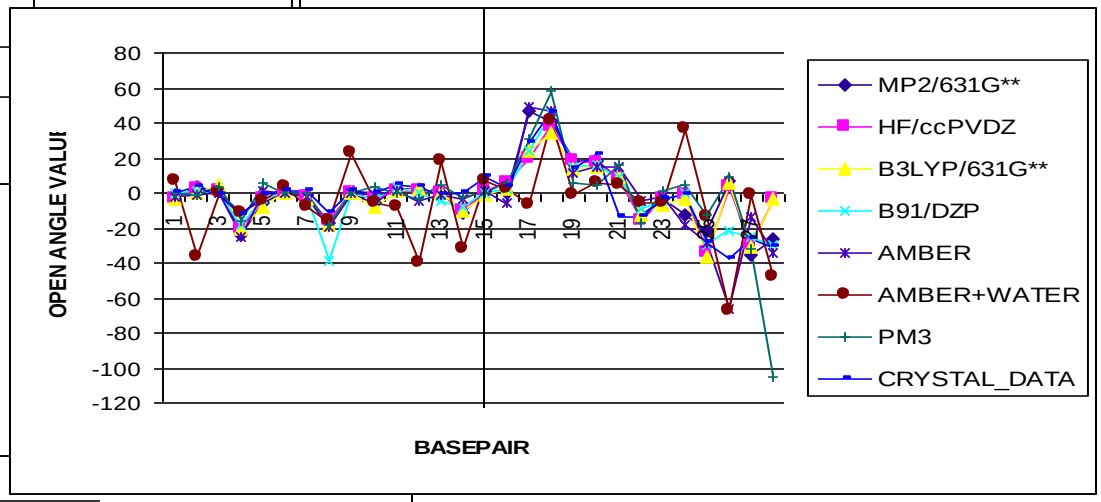


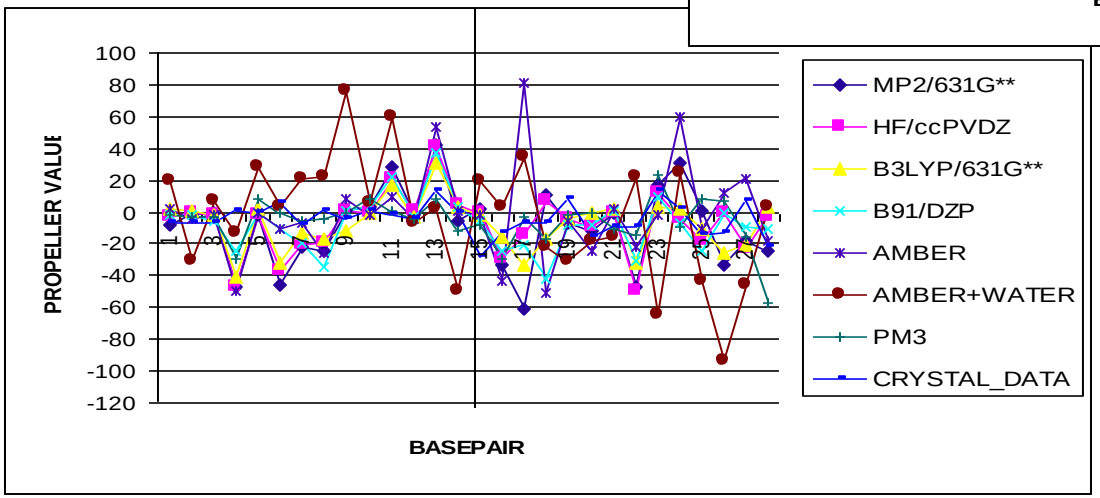




Structures of others with weaker H-bonds







Theory of Harmonic Vibration

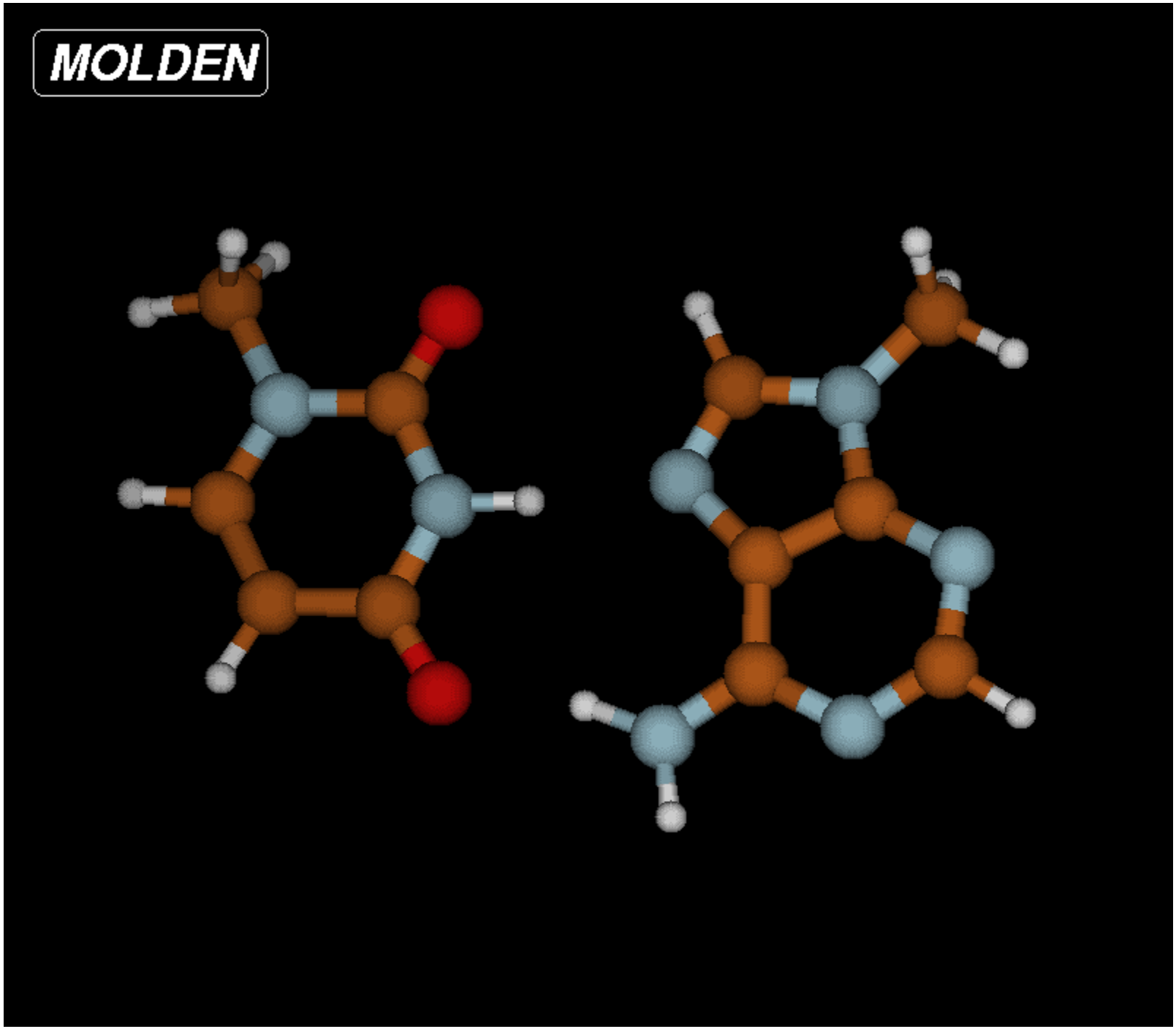
$$m\frac{d^2x}{dt^2} + kx = 0$$

$$T_{ij}\ddot{\eta}_j + V_{ij}\eta_j = 0$$

$$V_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j}$$

V is Total (QM) Potential Energy

$$\begin{vmatrix} V_{11} - \omega^2 T_{11} & V_{12} - \omega^2 T_{12} & .. & V_{1n} - \omega^2 T_{1n} \\ V_{21} - \omega^2 T_{21} & .. & .. & V_{2n} - \omega^2 T_{2n} \\ .. & .. & .. & .. \\ V_{n1} - \omega^2 T_{n1} & .. & .. & V_{nn} - \omega^2 T_{nn} \end{vmatrix} = 0$$


MOLDEN

Assignment of Vibration modes (frequency) to type of Motion:

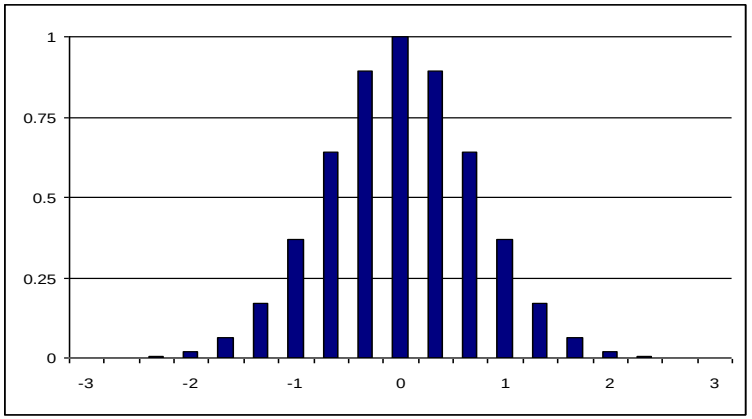
Generated two sets of coordinates of all the atoms, X^i_{max} & X^i_{min}

Ran NUPARM on both to find major differences in parameters

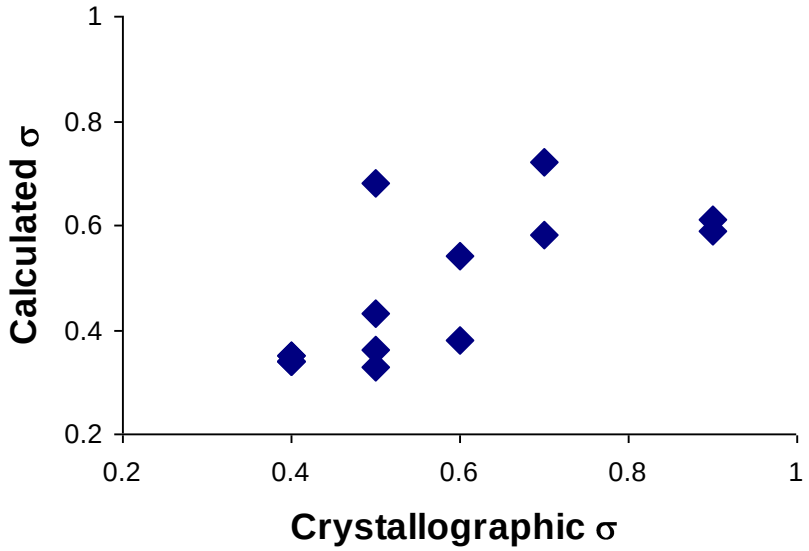
$$X^i_{max} = X^i_o\{1 + A\sin(90)\} \quad \text{and } X^i_{min} = X^i_o\{1 + A\sin(-90)\}$$

$$\frac{d^2x}{dt^2} + \omega^2 x = 0,$$

$$where, \omega^2 = \sqrt{k/\mu}$$



$$\rho(\alpha) = e^{-\frac{1}{2}k(\theta-\theta_o)^2/k_BT}$$



$$\sigma_{calc} = \sqrt{2ln2.k_BT/k}$$

Roy, Panigrahi, Bhattacharyya & Bhattacharyya, *J. Phys. Chem.* **B (2008) B112,** 3786

Sen, K.; Basu, S.; Bhattacharyya, D. *Int. J. Quant. Chem.* (**2006**) ***106***, 913

- ✓Base pairs vibrate mostly along five (instead of six) directions
- ✓Vibrations by breaking H-bonds are often prohibited
- ✓Vibrations are in the time scale of pico second
- ✓Force constants can be used for CG simulations

TABLE 5: Calculation of Dynamics of the Base Pairs along the Five Intra-Base-Pair Parameter Directions[a]

no.	base pair	buckle	open angle	propeller	stagger	Shear
1	G:C W:W C	43.19; 82.5		31.00; 26.2	73.36; 3.18	
2	A:U W:W C	29.84; 38.0		38.74; 30.6	73.38; 3.0	100.8; 6.0
3	G:U W:W C	32.04; 43.7	60.09; 22.3	31.55; 32.5	73.57; 3.0	
4	A:G H:S T	20.49; 28.9	38.22; 119.5	38.22; 40.2	64.26; 2.5	
5	A:G s:s T	12.47; 12.7	24.81; 56.2		44.3; 0.97	
6	A:U H:W T	25.0; 24.4	48.24; 120.0	33.72; 23.2	67.39; 3.0	93.57; 5.5
7	A:A H:H T	20.06; 24.4	44.34; 139.9	29.45; 23.1	75.3; 2.8	60.79; 2.0
8	G:A W:W C	16.88; 21.4		25.27; 17.6	59.74; 1.9	74.96; 3.0
9	G:A S:W T	14.59;16.8	37.72; 133.0	26.37; 11.8	37.72; 0.78	101.49; 5.9
10	A:A W:W T	19.33; 28.6	59.6; 306.5	28.16; 15.8	42.07; 1.1	97.32; 5.6
12	A:U W:W T	19.71; 27.3	64.76; 335.7	35.65; 22.7	69.77; 2.9	107.3; 7.5
13	A:A H:W T	22.86; 35.1	48.04; 181.6	30.13; 22.9	48.04; 1.2	100.8; 6.6
14	A:U H:W C	24.63; 23.6	65.49; 221.5	36.61; 27.3	74.4; 3.2	
15	A:G w:s C	18.78; 27.0	30.87; 94.4	23.22; 16.0	41.5; 1.0	61.43; 2.1
18	G:G S:S T		36.28; 131.1	24.2; 19.7	36.28; 0.7	102.5; 5.5
21	A:C w:w C	26.98; 30.2		27.51; 29.6	39.01; 0.8	65.92; 2.38
22	AU s:w C	40.6; 71.6	110.2; 672.4	31.71; 24.7	73.17; 3.0	
24	G:G H:W T	13.94; 15.1	40.9; 157.2	33.67; 45.6	52.96; 1.8	70.2; 3.3
25	A:A w:w C	20.07; 29.9	34.97; 111.2	15.89; 10.9	44.08; 1.1	82.00; 3.7
26	U:U h:w T	45.75; 83.3	93.39; 488.1	30.83; 15.3	67.86; 2.9	75.89; 3.5
30	A:C W:W T	53.32; 127.2	68.3; 288.3	34.56; 24.6	53.32; 1.6	
31	C:U W:W T	37.90; 45.7	119.9; 592.4	22.2; 9.5	68.67; 2.8	95.81; 6.1
33	C:C w:h C	47.71; 99.3	57.61; 177.5	39.41; 41.4	65.53; 2.1	

Conclusions / Appeals

- Non canonical base pairs are important for RNA structure prediction
- Many of these are sufficiently strong
- Estimated Force-constants can be used for CG modeling
- Their stacking interactions (combination of p-p interactions and hydrophobic effect) needs to be estimated.

Major contributors:

Malyasri Bhattacharyya

Shayantani Mukherjee

Jhuma Das

Swati Panigrahi

Ashim Roy

Sukanya Halder

Collaborators:

Prof. Manju Bansal

Prof. Abhijit Mitra

Prof. Jaydeb Chakrabarti

Prof. Jiri Sponer

Supporters:

CSIR and DBT (Govt. of India)

CAMCS and CBAUNP (DAE, SINP)

CDAC (Pune)