Leontis Westhof Notation for RNA-Protein Complexes?

Axel Wintsche

Computational EvoDevo Leipzig

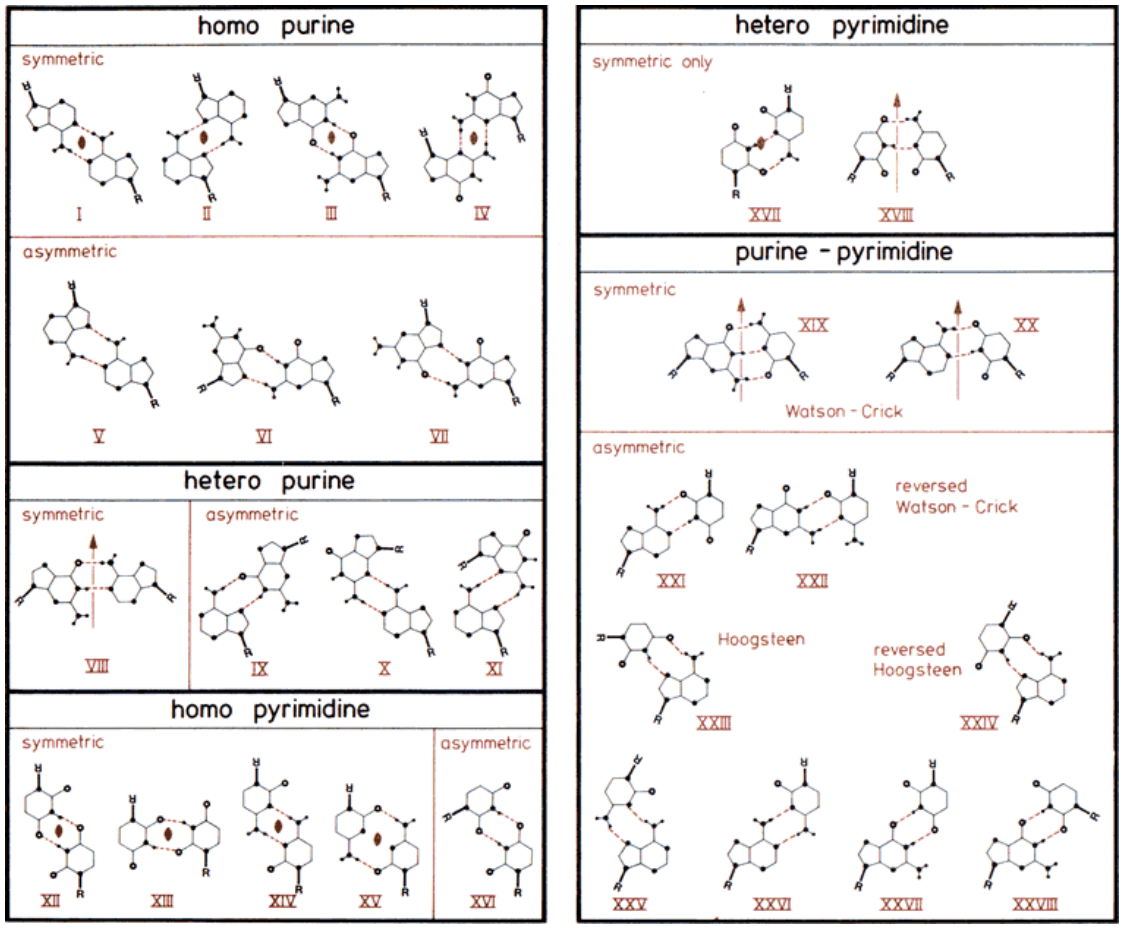
27^{th} TBI Winterseminar in Bled

Objective and Motivation

- RNA molecules present a vital part of life
- Secondary structure vs. tertiary interactions
- Growing number of 3D RNA structures enables systematic investigation of tertiary interactions
- Led to catalogs of non-WC base pairs and definition of recurring 3D modules [1, 2, 3]
- How do nucleotides interact with amino acids?

Classification of base pairs

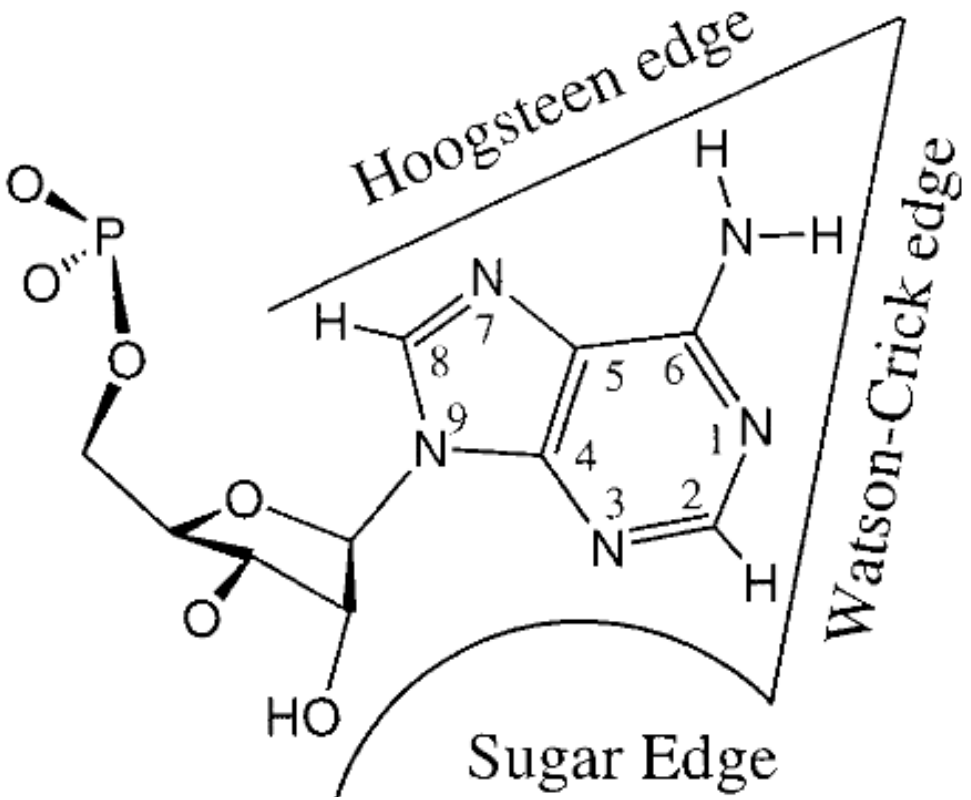
- Base pair classification by Saenger in '84 [4]



taken from http://ndbserver.rutgers.edu/atlas/legends/saenger.html

Classification of base pairs (cont.)

- Leontis Westhof (LW) notation defines 3 edges [2]



- Other bp classifications exists like in Lee and Gutell, 2004 [3]

Hydrogen bonding

- Base pairs are established via hydrogen bonding (H-bonds)
- H-bonds contribute to stability of 3D fold
- Electronegative atoms form donor/acceptor pairs

Hydrogen bonding

Period (horizontal)																	
1	H 2.20																He
2	Li 0.98	Be 1.57										B 2.04	C 2.55	N 3.04	O 3.44	F 3.98	Ne
3	Na 0.93	Mg 1.31										Al 1.61	Si 1.90	P 2.19	S 2.58	Cl 3.16	Ar
4	K 0.82	Ca 1.00	Sc 1.36	Ti 1.54	V 1.63	Cr 1.66	Mn 1.55	Fe 1.83	Co 1.88	Ni 1.91	Cu 1.90	Zn 1.65	Ga 1.81	Ge 2.01	As 2.18	Se 2.55	Br 2.96	Kr 3.00
5	Rb 0.82	Sr 0.95	Y 1.22	Zr 1.33	Nb 1.6	Mo 2.16	Tc 1.9	Ru 2.2	Rh 2.28	Pd 2.20	Ag 1.93	Cd 1.69	In 1.78	Sn 1.96	Sb 2.05	Te 2.1	I 2.66	Xe 2.60
6	Cs 0.79	Ba 0.89	*	Hf 1.3	Ta 1.5	W 2.36	Re 1.9	Os 2.2	Ir 2.20	Pt 2.28	Au 2.54	Hg 2.00	Tl 1.62	Pb 2.33	Bi 2.02	Po 2.0	At 2.2	Rn 2.2
7	Fr 0.7	Ra 0.9	**	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Uuq	Uup	Uuh	Uus	Uuo

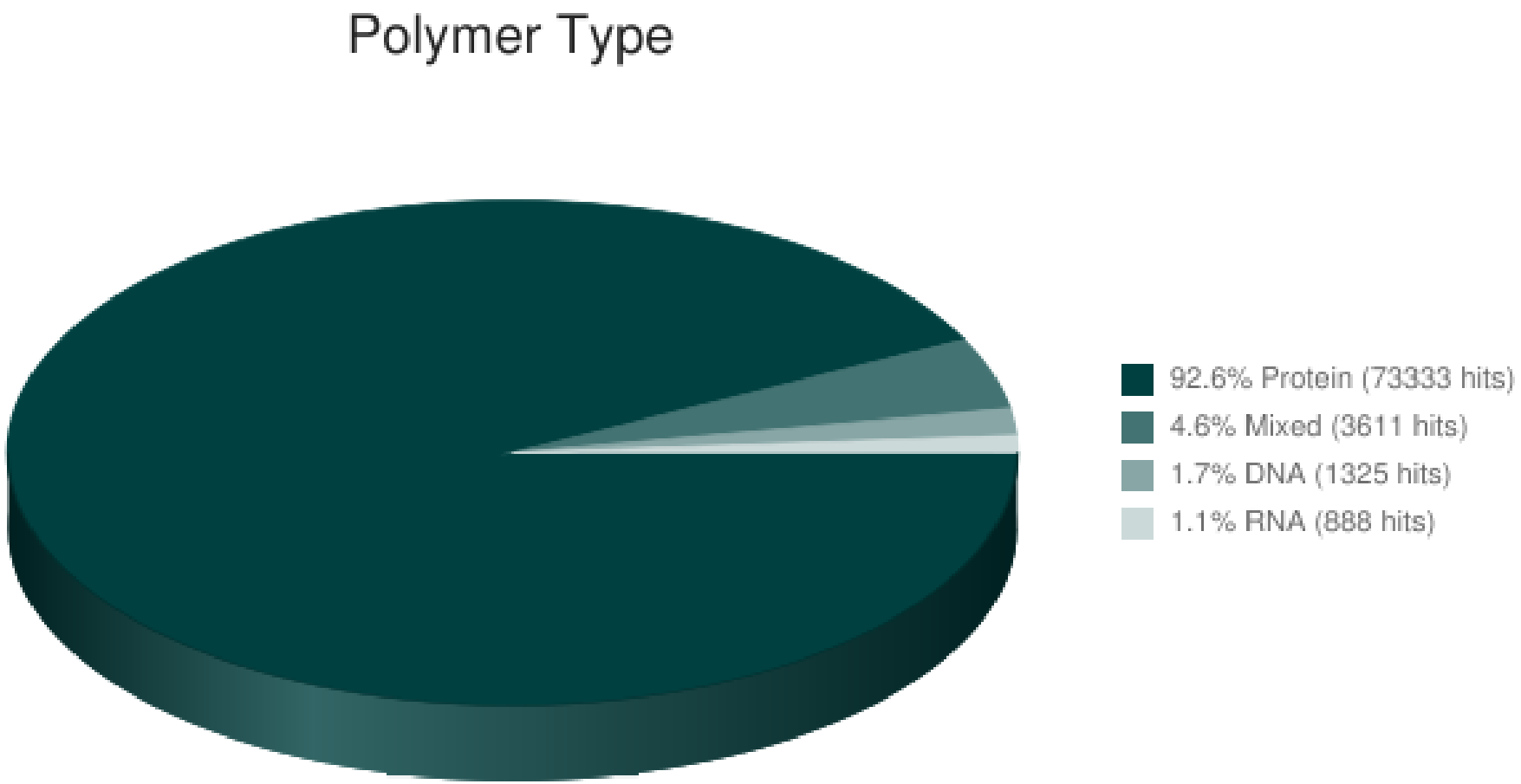
Hydrogen bonding

- Base pairs are established via hydrogen bonding (H-bonds)
- H-bonds contribute to stability of 3D fold
- Electronegative atoms form donor/acceptor pairs

- H-atoms missing in most 3D structures

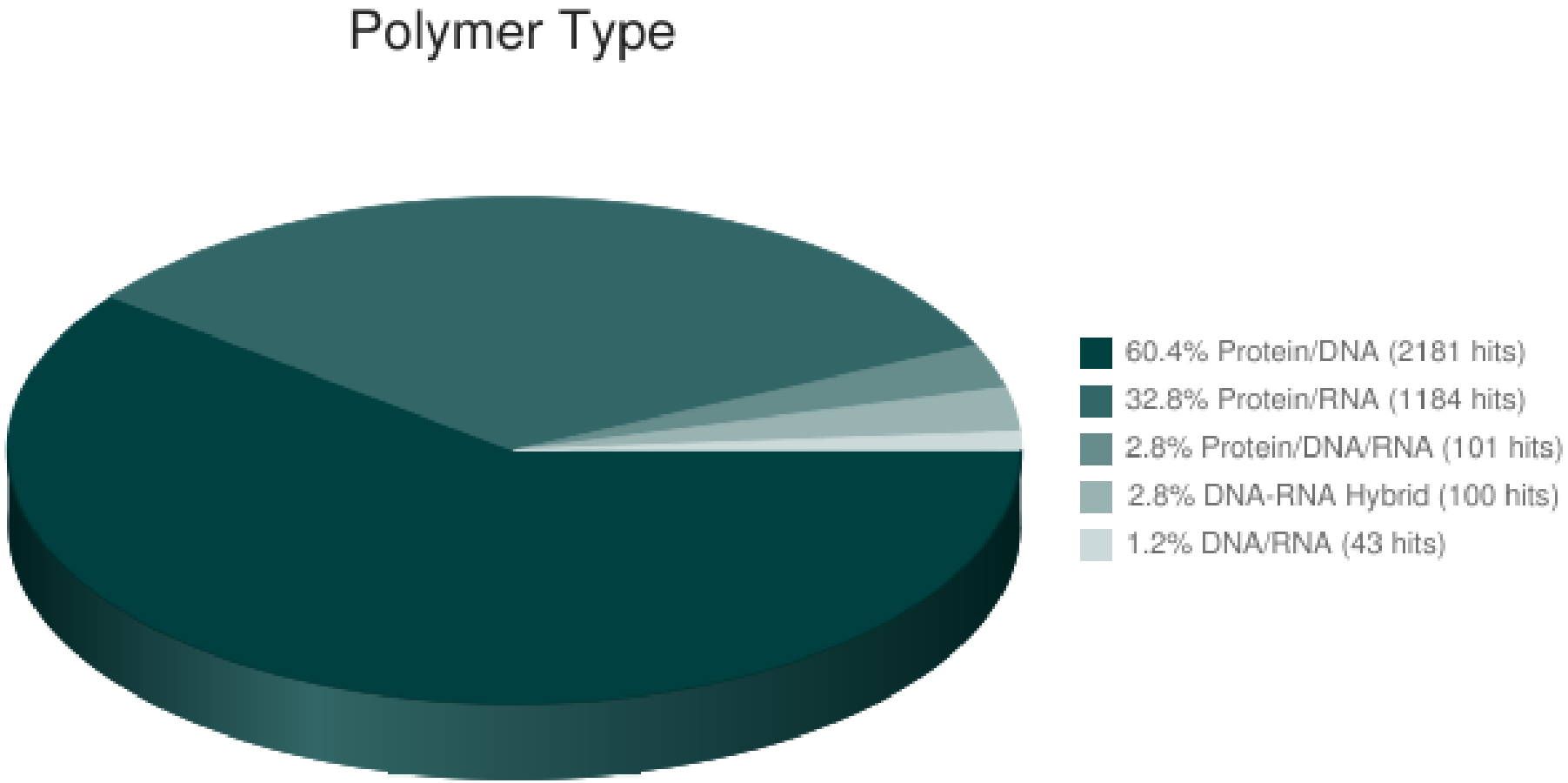
Dataset

- 3D Structures from Protein Data Bank



Dataset

- 3D Structures from Protein Data Bank



Predicting non-canonical base pairs

Computational identification of base pairs

- Reference structures give empirical values for
1. distances between (donor/acceptor) atoms
2. angles between planes of bases

Characterizing the geometry of nucleic acids interacting with proteins, obviously, brings up a whole new host of geometrical issues.

Olson et al., 2001 [5]

Predicting H-bonds

Scanning for putative H-bonds

- Donor/acceptor atoms: N, O, [C, S, Se]
- Minimum euklidean distance to find putative atom pair
- Distance cutoff ($d \leq 3$ angstrom)
- Penalty for atoms from adjacent nucleotides

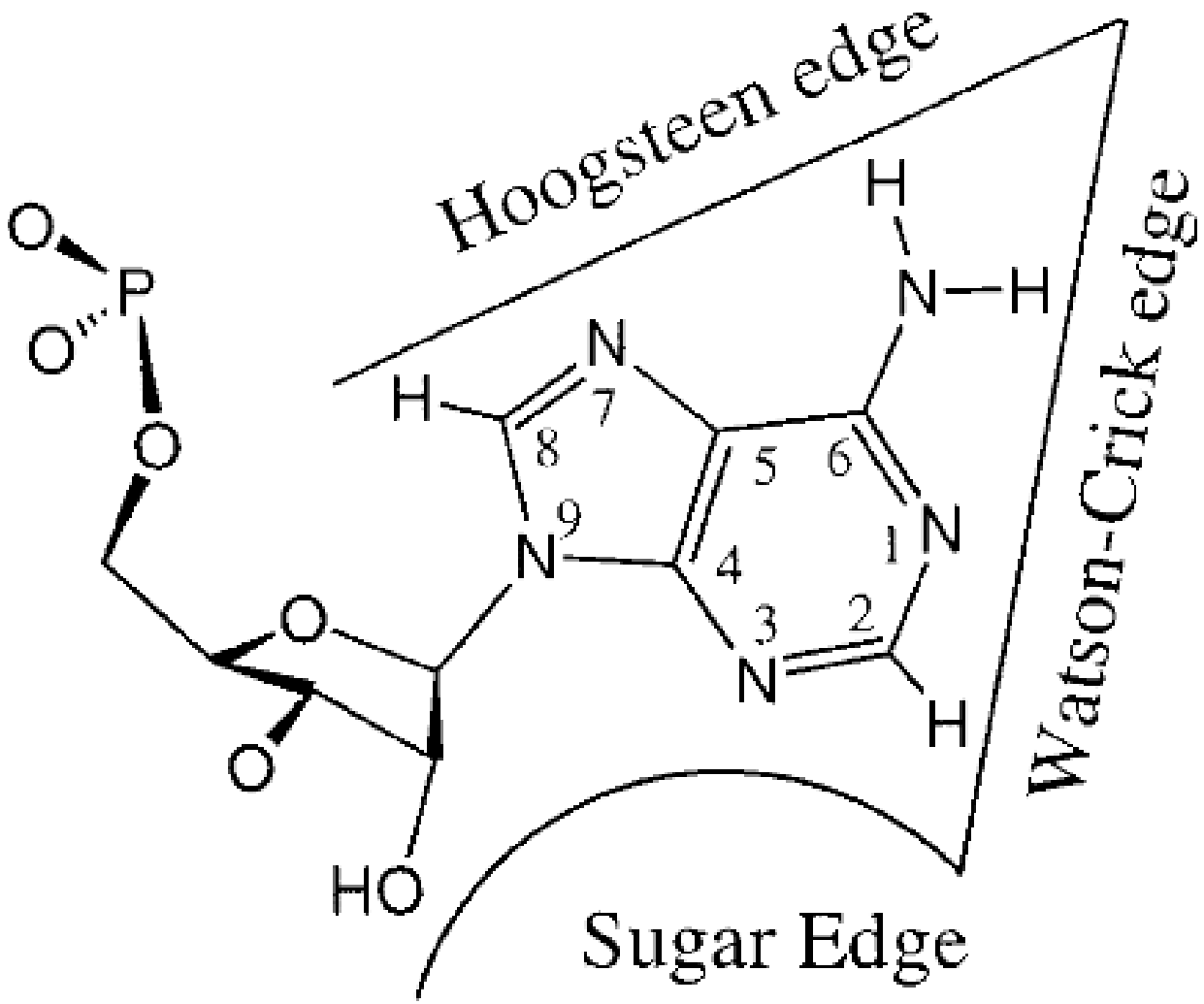
Predicting H-bonds

Scanning for putative H-bonds

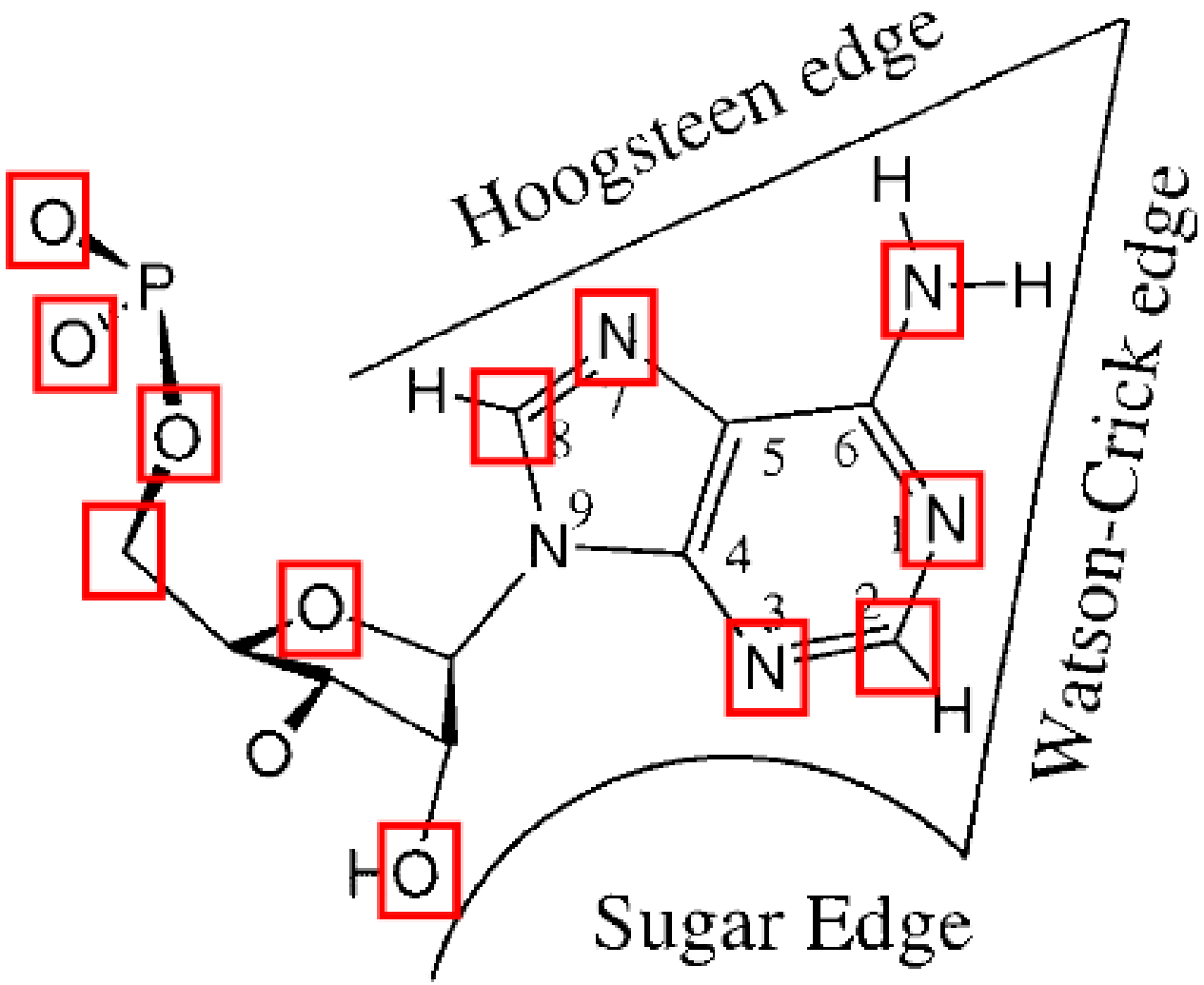
- Donor/acceptor atoms: N, O, [C, S, Se]
- Minimum euklidean distance to find putative atom pair
- Distance cutoff ($d \leq 3$ angstrom)
- Penalty for atoms from adjacent nucleotides

Recap: H atoms missing

H-bond donor/acceptor atoms



H-bond donor/acceptor atoms



Effect of missing H atoms

- Use all structures that contain H atoms
- Treat them as:
  a) containing no H atoms
  b) containing H atoms
- Compare predicted H-bonds

Effect of missing H atoms

- Use all structures that contain H atoms
- Treat them as:
  a) containing no H atoms
  b) containing H atoms
- Compare predicted H-bonds

Calculate number of *equal* and *different* predictions *b*

- **X-H** type: add 1 to $\begin{cases} \textit{equal} & \textbf{if} \quad b(\mathrm{X}) = b(\mathrm{H}) \\ \textit{different} & \textbf{if} \quad b(\mathrm{X}) \neq b(\mathrm{H}) \end{cases}$

Effect of missing H atoms

- Use all structures that contain H atoms
- Treat them as:
  a) containing no H atoms
  b) containing H atoms
- Compare predicted H-bonds

Calculate number of *equal* and *different* predictions *b*

- **X-H** type: add 1 to $\begin{cases} \textit{equal} & \textbf{if} \quad b(\mathrm{X}) = b(\mathrm{H}) \\ \textit{different} & \textbf{if} \quad b(\mathrm{X}) \neq b(\mathrm{H}) \end{cases}$
- $\mathbf{H_1}$-**X**-$\mathbf{H_2}$ type: add 1 to $\begin{cases} \textit{equal} & \textbf{if} \quad b(\mathrm{X}) = b(\mathrm{H}_i) \\ \textit{different} & \textbf{if} \quad b(\mathrm{X}) \neq b(\mathrm{H}_i) \end{cases}$

4 non-exclusive rules!

Quality of H bond prediction

Number of *equal* and *different* predictions

type	all	*equal*	%	*different*	%
X-H	4021	3499	0.87	522	0.13
$\mathbf{H_1}$-**X**-$\mathbf{H_2}$	2758	1600	0.58	1158	0.42

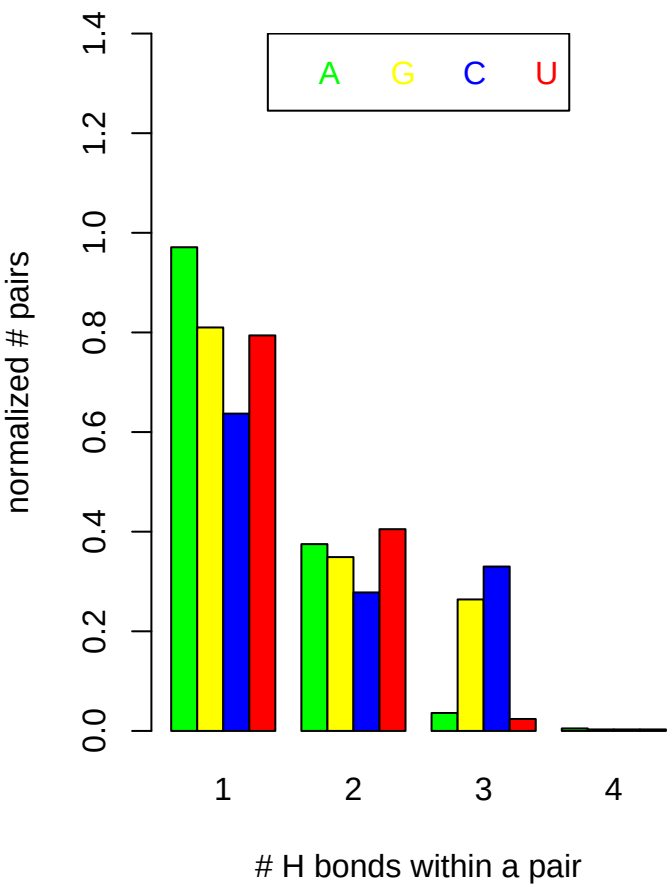
H-bond distribution

Relative frequency of having an H-bond with another nuclceotide or amino acid

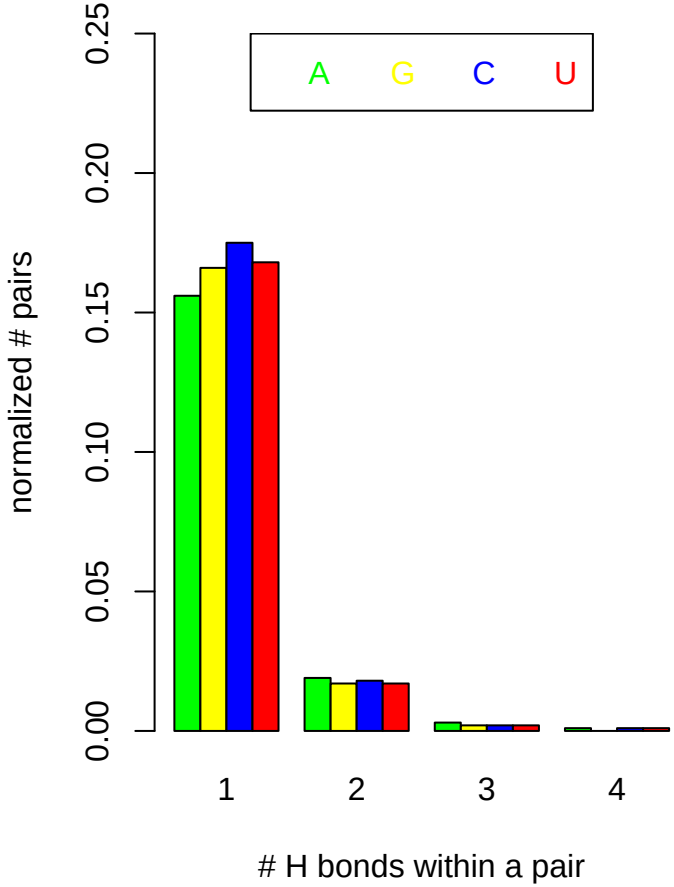
	A		G		C		U		
N1	0.39	0.01	0.59	0.00	0.66	0.02	0.35	0.02	O2
C2/N2	0.10	0.00	0.65	0.01	0.59	0.01	0.53	0.01	N3
N6/O6	0.43	0.01	0.52	0.01	0.53	0.01	0.30	0.01	N4/O4
N7	0.19	0.01	0.05	0.01	0.01	0.00	0.02	0.00	C5
C8	0.04	0.00	0.02	0.00					
O2'	0.26	0.04	0.25	0.03	0.21	0.04	0.21	0.04	O2'
O5'	0.02	0.00	0.01	0.01	0.01	0.00	0.01	0.00	O5'
O	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	O
OP1	0.07	0.06	0.05	0.07	0.05	0.07	0.07	0.06	OP1
OP2	0.12	0.03	0.04	0.03	0.04	0.04	0.07	0.03	OP2

H-bond co-occupation for pairs

Nucleotide-Nucleotide



Nucleotide-Amino Acid



Outlook

- More sophisticated model for H-bonds
- Redundancy/Bias of structures in PDB
- Definitions of pair and stacking interactions
- Empirical rules for nucleotide – amino acid interactions
- Do we see all possible interactions in the PDB data?

Acknowledgements

- Sonja
- Peter
- Christian Otto
- Winterseminar organizers

Thank you for your attention!

Gefördert aus Mitteln
der Europäischen Union

References I

N.B. Leontis and E. Westhof.
Analysis of rna motifs.
Current opinion in structural biology, 13(3):300–308, 2003.

N.B. Leontis and E. Westhof.
Geometric nomenclature and classification of rna base pairs.
Rna, 7(4):499–512, 2001.

J.C. Lee and R.R. Gutell.
Diversity of base-pair conformations and their occurrence in rrna structure and rna structural motifs.
Journal of molecular biology, 344(5):1225–1249, 2004.

References II


W. Saenger.
Principles of nucleic acid structure.
Springer-Verlag New York, 1984.


W.K. Olson, M. Bansal, S.K. Burley, R.E. Dickerson, M. Gerstein, S.C. Harvey, U. Heinemann, X.J. Lu, S. Neidle, Z. Shakked, et al.
A standard reference frame for the description of nucleic acid base-pair geometry.
Journal of molecular biology, 313(1):229–237, 2001.