Is the H4 histone tail intrinsically disordered or intrinsically multifunctional? Supplementary Material

Konstantin Röder[1,2]

(1) Robinson College, Grange Road, CB3 9AN, Cambridge, UK
(2) Department of Chemistry, University of Cambridge,
Lensfield Road, CB2 1EW, Cambridge, UK

Correspondence should be addressed to K. Röder (kr366@cam.ac.uk)

Detailed description of the structural ensembles

As the structures are very similar for the isolated and bound H4 tails, the structural ensembles are discussed in general, and deviations in the unbound or bound set are only highlighted if they are significant. A number of structural features can be employed to distinguish the ensembles, and three of them, the radius of gyration, the percentage of the solvent-accessible surface area (SASA) that is occupied by the basic patch, and the distance between the lysyl nitrogen atoms in residues 16 and 20 are given in Tables S1 and S2 for the isolated and bound histone tail, respectively.

Set A shows structures with a hairpin that is pulled into a more compact form due to strong contacts formed by Arg19 and Lys20. Key interactions are formed between Leu10 and His18, and a number of additional hydrogen bonds are formed by neighbouring residues (Gly9-O and Arg19-NH, Arg17-O and GLy11-NH). This basic structure is then pulled into a more compact

form, bending the hairpin, and leading to a hydrogen bond formed between Leu22-NH and His18-O. While these structural motives are pressed, deprotonation reduces the interactions of Lys20 and instead interactions are formed between the N-terminal residues and Gly6 and Gly7. This change leads to a somewhat more compact structure after deprotonation, but the set is the most compact as measured by the radius of gyration, the distance between lysyl nitrogen is large, and the lysines in position 16 and 20 are not easily accessible.

Set B exhibits in the isolated and bound form two subsets, as indicated by the two subfunnels observed in the disconnectivity graphs. The entire set is based on, in some cases distorted, antiparallel β-sheets formed by Gly11-Ly12-Gly13 and Val21-Leu22-Arg23. In one subset, B(1) the N terminus points away from the centre of the molecule, which can lead to extension of the β-sheets to Lys5-Gly4. In B(2), the N-terminus forms more interactions, curling up above the β-sheets. As a consequence the β-sheets are somewhat twisted, but DSSP still identifies them as β-sheets. In both cases, the change of protonation does not alter the structures significantly, likely due to the fact that the basic patch, and with it Lys16 and Lys20, are exposed to the surrounding at one end of the molecule. As a result the percentage of SASA formed by the patch is the highest observed across the landscape.

Set C is the only ensemble that exhibits helices in the histone tail structure. Two helical regions are formed around residues Gly7 to Leu10 and Val21 to Asp24. The resulting structure brings the N and C termini close together, and, through additional interactions with Lys16, a stable scaffold is formed. Furthermore the sidechains of Lys16 and Arg17 both interact with the first helical region. As a result the first part of the basic patch is fundamentally involved in the intramolecular structure stabilisation, and only Arg19 and Lys20 are solvent-exposed. This feature can also be identified from the low percentage of the SASA that is formed by the basic patch.

The structures in set D exhibit hairpins with a free N-terminus. A difference between the bound and isolated set is observed based on the length of the free terminus. In the isolated form the hairpin involves fewer residues, with Gly6-Asp24 being the first interacting pair of residues, while for the bound form the bottom of the hairpin is formed by Gly4 and Gln26. In addition this lower section of the hairpin for the bound histone tail is twisted compared to the isolated structure. Nonetheless, the key interactions around the basic patch are conserved upon protonation, namely the sidechain of Lys20 interacts with the backbones of residues 15 and 17, the sidechain of Arg17 with the backbone of residues 20 and 22, and finally the sidechain of Arg19 with residues 7 and 24. Again, the SASA percentage by the basic patch is relatively low, as Arg17, Arg19 and Lys20 are involved in key intramolecular interactions, but Lys16 is accessible as it protrudes into free space.

The last common set, E, is again hairpin like, with β-sheet like structure formed by Lys8 and Gly9 with Leu22 and Arg23. The turn of the hairpin like structure is located at residues 14 to 16, leading to a mismatch in length on either side of the turn. This strain is released by a bend occurring around GLy11, shortening the distance between the β-sheet like region and the turn such that another turn, involving residues 19 to 21 can form. In both cases, the upper portion is stabilised by Arg17 interaction with the backbone of the residues 13 and 14. In the isolated form, additional interaction between the sidechain of Arg19 and the C-terminus is observed. Overall this leads to a U-shaped structure, with a reasonably high exposure of the basic batch, and importantly the shortest lysyl distance observed. In the bound form, this distance can be as short as 4 Å. The average observed in the isolated form, likely due to Coulombic repulsion between the charged sidechain is around 8 Å.

The changes in relative energies between the landscapes are relatively small, and likely down to the reduced Coulombic repulsion. The interactions within the structural sets are mostly preserved, but we observe changes in the distance between the sidechains of Lys16 and Lys20

for sets D and E, going along with a lower relative energy as for example compared to set B after binding. Similar set A is relatively lowered in potential energy, likely due to lower repulsion that would affect its compact structure more.

The two remaining sets are unique, but related to other ensembles. For the isolated histone tail, the additional funnel is similar to E, as it exhibits β-sheet like structures towards the bottom of a hairpin like structure with a distorted top section. Compared to E, the interactions are shifted, with the main interactions formed between Gly6 to Lys8 and Lys20 to Arg23. This pattern corresponds to a shift by two residues relative to E, and is complemented by interactions between Leu10-NH and Lys16-O and by the sidechain of Lys16 to Gly11-O and Gly13-O. These interactions significantly distort the structure and result in a relatively high potential energy.

The additional subset in the bound case stems from a distortion of A-like conformations. As the structures in set A form hydrogen bonds between the backbone atoms of Gly13 and Lys16, a sharp turn is formed at Ala15. This turn leads to the compact structure and a large number of interactions between the basic patch and residue in the C-terminal region. In the higher energy subset, the turn is less sharp, and as a result, instead of backbone interactions, Arg17 and his18 only interact via their sidechains with Gly7-O and Gly9-O, respectively. This change leads to fewer favourable interactions, i.e. a higher potential energy, but exposes more of the basic patch. The new interactions are formed by a rotation in the backbone orientation in the basic patch, which further brings the lysyl groups of residues 16 and 20 closer together (11 Å compared to around 20 Å for A). As this would increase the repulsion in the isolated case, it is likely that such structures would be even higher in energy in the protonated molecule.

Table S1: Overview of key properties for the isolated H4 tail. Here, the radius of gyration, the percentage of the solvent-accessible surface area (SASA) formed by the basic patch (residues 16 to 20) and the distance between lysyl nitrogen atoms in residues 16 and 20 are provided. The values are the averages for all local minima in the respective funnels. For set B, we distinguish into the two subfunnels with distinct structures. B (1) contains the structures with curled up N terminus, while B (2) has an extended, free N terminus.

Set	R_{gyr} / Å	SASA basic patch / %	d(NZ16-NZ20) / Å
A	8.3 ± 0.2	21.5 ± 1.6	21.0 ± 2.9
B (1)	9.0 ± 0.2	30.4 ± 1.5	13.4 ± 0.9
B (2)	8.6 ± 0.2	34.6 ± 1.2	15.5 ± 1.9
C	8.8 ± 0.3	22.6 ± 3.3	15.9 ± 1.6
D	8.8 ± 0.4	24.3 ± 3.4	13.4 ± 4.2
E	8.7 ± 0.4	29.4 ± 2.5	8.2 ± 2.8

Table S2: Overview of key properties for the bound H4 tail. Here, the radius of gyration, the percentage of the solvent-accessible surface area (SASA) formed by the basic patch (residues 16 to 20) and the distance between lysyl nitrogen atoms in residues 16 and 20 are provided. The values are the averages for all local minima in the respective funnels. For set B, we distinguish into the two subfunnels with distinct structures. B (1) contains the structures with curled up N terminus, while B (2) has an extended, free N terminus.

Set	R_{gyr} / Å	SASA patch / %	d(NZ16-NZ20) / Å
A	8.0 ± 0.3	22.6 ± 1.9	21.0 ± 2.9
B (1)	8.9 ± 0.2	30.8 ± 1.3	13.4 ± 0.6
B (2)	8.6 ± 0.1	37.1 ± 1.0	14.4 ± 1.2
C	8.4 ± 0.3	20.4 ± 3.6	17.7 ± 0.9
D	8.7 ± 0.2	20.7 ± 2.0	8.8 ± 0.6
E	8.6 ± 0.3	27.3 ± 2.7	4.5 ± 2.4

Comparison of frustration indices

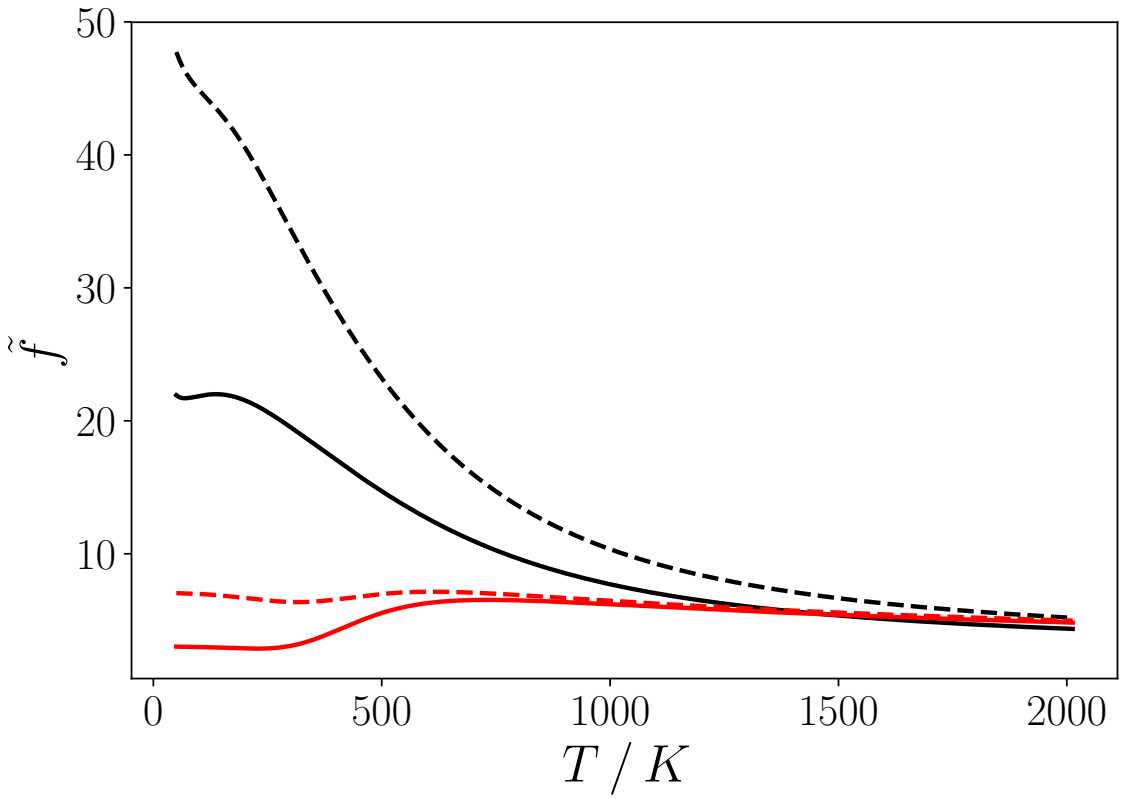

Figure S1: Renormalised frustration indices for the H4 histone tail (red) and the Aβ monomer. Two different energy intervals for the superbasin analysis are shown, 1.0 kcal/mol (solid line) and 10.0 kcal/mol (dashed line). At low temperatures, the histone tail shows a smaller level fo frustration.