Digital Biology, Spring 2015
Project Suggestions and Reading List

April 22, 2015

The course will require

- presentation in class of one significant paper,

- the execution of a research project, and

- presentation in class of your results.

To promote coherence of the class, it is suggested that you choose a project from the following list, as well as a paper for the presentation that is closely related to the project. Students are encouraged to form groups to work on a single project. Students may choose other projects subject to approval of the instructor.

1 Solvent-centric view of dehydrons

These projects will explore a solvent-centric view of dehydrons as opposed to the accepted structure-centric view that has been the primary view of dehydrons in the past. The initial phase of the class will cover the basics of protein biophysics and present the ideas behind dehydrons and their role in structural interactions. The material will include basic bioinformatics techniques for examining protein structure. After this introductory period, the class will be based primarily on student presentations.

The dehydron is a structural defect in gene products (especially proteins) and represents a key concept in the understanding of biological interfaces. Its unique physico-chemical properties have inspired a multi-scale theory of

biological water while making the dehydron a pivotal factor in protein associations. In this course, we intend to examine dehydrons from a solvent-centric perspective, rather than adopting the structure-centric view that most people are familiar with. This shift in view point entails several advantages as one strives to understand the functional properties of dehydrons. In particular, a solvent-centric view leads to the striking conjecture that dehydrons may behave as catalytic elements, often stimulating, enabling and even promoting enzyme function. We want to study the involvement of the dehydron in biochemical events after we have already delineated its involvement in protein associations, the hallmarks of biological activity.

The new solvent-centric view of dehydrons also yields an interesting interpretation of solvent-as-material in biological associations, with dehydrons playing a significant role in causing material defects. This could have significant implications for materials engineering involving solvents in other biomaterials and beyond. These ideas and projects have been developed in collaboration with Ariel Fernández Stigliano.

1.1   Overview of Projects

The following are suggested projects for the course. The first six projects relate to a possible role of dehydrons in phosphorylation, and the following two are closely related to one of the six.

Direct inspection of protein structure shows that dehydrons tend to cluster around residues implicated in catalytic function and protein association. The latter property comes as no surprise due to the dehydration propensity of dehydrons and has been widely studied, as will be explained at the beginning of the class. However, the possibility of a dehydron-based catalyst or a dehydron-based stimulator/enabler of protein function has only recently been explored [1]. Some questions you can investigate include the following.

- To what extent are phosphorylation sites surrounded with dehydrons?

- To what extent are dehydrons over-represented at the interfaces of transciption factors?

These are the types of questions that, when properly addressed may shed light on the unique functional properties of dehydrons, i.e., on their participation in chemical events. In essence, we want to explore such functional properties

from a solvent-centric perspective. That is we want to understand what is the dehydron doing to the surrounding solvent in order to prepare it for catalysis. Because we seek to delineate the participation of dehydrons in chemical events, we may ultimately need to resort to a quantum mechanical treatment [2, 3, 4] of the dehydron-decorated catalytically active site.

1.2   Phosphorylation and wrapping

One way to examine the role of dehydrons and phosphorylation is by datamining in the PDB of the sort we have done in the course. The objective is to determine the extent to which there is a correlation between phosphorylation sites and nearby dehydrons. The tasks are as follows.

First analyze explicitly phosphorylated residues PTR etc. in PDB files to see how frequently their phosphate groups are in desolvation spheres of underwrapped mainchain hydrogen bonds.

- Keep a count of how many dehydrons are near a given PTR etc.

- Further, plot the frequency of appearance of PTR etc. in desolvation domains as a function of wrapping.

- Compare this with nonphosphorylated residues Tyr, Ser, Thr.

Use wrappa to determine desolvation domains, amount of wrapping, and candidate dehydrons.

Next, if time permits, survey the various software tools available on the web to predict phosphorylation sites, e.g., http://www.phosphosite.org/ and www.phosida.com. Pick one or more of them to predict phosphorylation sites as needed. Using the software to predict phosphorylation sites, consider wrapping around all predicted phosphorylation sites versus wrapping around all other Tyr, Ser, Thr. Count the number of dehydrons near a given site (for a given PTR etc. how many desolvation domains is it in?).

1.3   Baseline statistics

In the projects in sections 1.2 and 2 investigate the proximity of particular motifs (e.g., phosphorylation sites) to dehydrons. As a baseline, determine the number of dehydrons that would be expected based on a purely random model based on the statistical definition of dehydon. That is, for any atom in

a protein, what is the probability of finding a nearby dehydron. That is, what is the probability that an arbitrary atom in a protein is in the desolvation shell of some dehyrdon? Do this in two ways. First of all, the concept of dehydron is based on the amount of wrapping being below some level. Based on this, estimate the probability of finding a dehydron near any part of a proteins if they were randomly (uniformly) distributed. This depends on the particular radius of desolvation domain used. From this, estimate the number of atoms found within and without desolvation domains of dehydrons. Next, evaluate this by looking at the atoms found within desolvation domains of dehydrons in typical proteins. For example, take a representative set of proteins for which the hydrogens are included in the associated PDB files.

1.4 Transcription factors and wrapping

Transcription factors are proteins that bind to specific DNA sequences. The key side chains in such proteins are the DNA-intercalating Arg and Lys. Arg and Lys have significantly different hydration demands, but in additon dehydrons can alter local solvation environments. In this way, dehydrons are often involved in lowering the dehydration cost in biomolecular associations.

- Examine the distribution of dehydrons around DNA-intercalating Arg and Lys sidechains by surveying PDB files that display proteins binding to DNA.

- As a baseline, give the distribution of dehydrons around all Arg and Lys sidechains in PDB files.

1.5 Kinases and wrapping

Kinases are proteins that are often implicated in cancer. By searching in the literature, examine constitutively active kinases that are often carcinogenic. Many oncogenic mutations in such kinases are in the phosphorylation sites and some occur in their periphery, yet the latter are seldom studied since it is not clear how to interpret them by currently accepted standards. Sidechain mutations often cause significant changes in wrapping without causing changes to three-dimensional strucutre, and thus they can have significant changes in function without raising red flags. A mutation that would lead to misfolding (or at least nonfolding) might cause the protein

4

to be immediately degraded. Thus sidechain mutations that lead to changes in wrapping can often be the most dangerous.

Many oncogenic mutations in kinases are known, and these provide a rich data set to explore.

- See if perhaps there are mutations that create or destroy dehydrons in the periphery of catalytic sites.

- See if corresponding changes in phosphorylation rates correlate with nearby dehydrons.

1.6   Proton exchange

What is the rate of proton exchange for surface residues Ser or Thr or Tyr when decorated by dehydrons versus when they are not? We suspect it is much higher in the first case, if the dehydrons turn surrounding water into a chemical base (proton acceptor). Explore the NMR and other literature on this matter. Start with the web site

`http://nmrwiki.org/wiki/index.php?title=Amide_Hydrogen_Exchange_(HX)`

or other sources on the web and then

- do a literature search to see what data exists that could address this question

- compare this data with wrapping data.

1.7   Dual roles

Dehydrons are known for their role as dehydration promoters. If there is a role of dehydrons as proton acceptors, as suggested in the above projects,

- what is the significance of the dual role as proton acceptors and dehydration promoters for enzyme catalysis?

- Are their examples of catalytic performance compromised by mutations that remove nearby dehydrons, and vice-versa?

The above projects suggest that a dehydron can be viewed as a chemical base, in the sense that it causes the surrounding waters to act as a chemical base, in addition to its accepted role as a promoter of its own dehydration. But each role is exerted at a different time. When the dehydron nanocavity is filled up with water, the dehydron can be viewed as a base, or rather, the water that envelops it is a base. When a binding partner approaches, the dehydron reveals itself as promoter of self-dehydration, as this property translates into a mechanical attractive drag known as dehydronic field. As water is expelled upon association, the dehydron ceases to act as a base.

The basic nature is used to titrate local charges, disguising them upon protein associations. For example, the charge on amonium (Lys) or guanidinium (Arg) in transcription factors is "disguised" through deprotonations induced by dehydrons. On the other hand, activation of nucleophilic groups for catalysis also requires deprotonation by dehydrons nearby.

As the dehydron also helps with the anchoring of the binding partner, it ceases to act as a base. Titrating the local charge is very important: it activates the enzymatic site or disguises the charges mitigating the dehydration cost that results upon association.

This dual role, if correct, would say that dehydrons contribute to functionalize sites both for protein recognition (binding) and for enzymatic activity. Mechanistically, it would say that water wants to leave the dehydron and in so doing it can take a proton with it.

The same dual role may also affect the concentration of dehydrons around Arg or Lys in protein-nucleic acid recognitions (i. e., transcription factors), so we expand on this topic in project 1.4.

1.8 Deprotonation dynamics

Phosphorylation of tyrosine kinases is a key step in the cascade of events involved in cancer. A first step in the phosphorylation of sidechains is deprotonation, that is, the removal of the proton of the terminal hydrogen on sidechains Tyr, Ser, and Thr, leaving an excess electron and thus a negative charge at that site [5, 1]. For example, Figure 1 shows how this might work with the removal of the proton from the terminal OH group on threonine. Deprotonation of the tyrosine terminal (phenolic) hydroxyl enables the latter to perform the nucleophilic attack on the terminal phosphoester linkage of ATP, the step that enables the Tyr phosphorylation. So, the question becomes: which nearby chemical or quasi-chemical entity promotes that de-

protonation? What has this deprotonation process to do with the fact that there are dehydrons near a functionally competent Tyr? In what way may it be that dehydrons contribute to promote this deprotonation? To begin to answer some of these questions, we propose some simple computational experiments.

The extent of wrapping changes the nature of hydrogen bonds [6] and the structure of nearby water [7]. Hydrogen bonds that are not protected from water have different dynamics [6]. Figure 5 of [6] shows the striking difference of water residence times for well wrapped versus underwrapped hydrogen bonds.

To deprotonate Tyr, Ser, or Thr, a water molecule has to approach the OH group

(1) close enough to the terminal hydrogen and

(2) in the correct orientation, with the water oxygen adjacent to the terminal hydrogen.

In a typical PDB file, you would not expect to see such water molecules since to be in a PDB file they have to be immobilized in some way. The project is to examine waters in PDB files and characterize them. In particular

- determine the correlation between wrapping and **distances** of waters in PDB files that are within a given distance of a hydrogen bond

- determine the distribution of **distances** and **angles** for waters near threonine sidechain hydrogens and see if there is a difference between well wrapped THR and underwrapped THR.

- Do the same analysis for other sidechains that get phosphorylated.

2   Metalloproteins

Many proteins interact in critcal ways with metals, such as iron in the heme group in myoglobin. Look at http://en.wikipedia.org/wiki/Metalloprotein for a list of such proteins. The role of dehydrons as interaction sites for other proteins and small molecules is well established. Now we want to see if dehydrons are critical interaction sites for single atoms, specifically for metal ions.
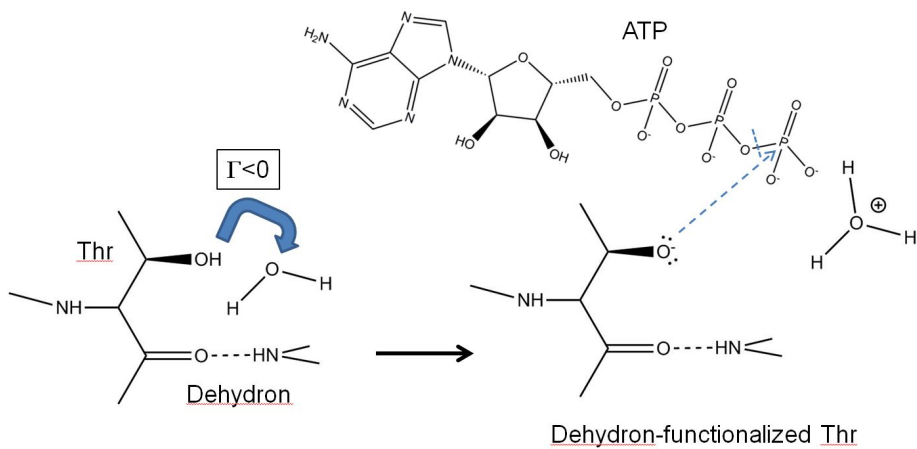
Figure 1: Phosphorylation process for threonine. Figure due to Ariel Fernández Stigliano [private communication].

Quantify the extent to which there are dehydrons in the vicinity of the metal ions in proteins, such as iron in hemeglobin. Include in your study the cadherin proteins. More precisely,

- make a list of all PDB files containing such metal ions, and

- list for each such protein, and each such metal engaged with the protein, the number of dehydrons having the ions within a desolvation shell of that dehydron.

3 Wrapping and ω-angle

Recall that the ω-angle is defined as the dihedral angle

$$[C_\alpha(x_i), C(x_i), N(x_{i+1}), C_\alpha(x_{i+1})].$$

See how the distribution of ω-angle correlates with wrapping. More precisely, for each i, determine the number ρ of non-polar carbonaceous groups within a radius R (use $R = 5, 5.5, 6, 6.5$ in your experiments) of the midpoint between $C(x_i)$ and $N(x_{i+1})$. For each ρ, plot the distribution of ω-angles. More precisely, look for patterns in the plots, e.g., plot the distributions for small ρ and large ρ. How does the choice of R affect the overall conclusions?

References

[1] Ariel Fernández. Packing defects functionalize soluble proteins. *FEBS letters*, 2015.

[2] Albert P Bartók, Michael J Gillan, Frederick R Manby, and Gábor Csányi. Machine-learning approach for one-and two-body corrections to density functional theory: Applications to molecular and condensed water. *Physical Review B*, 88(5):054104, 2013.

[3] Markus Bachmayr and Wolfgang Dahmen. Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. *Foundations of Computational Mathematics*, pages 1–60, 2013.

[4] Markus Bachmayr. *Adaptive low-rank wavelet methods and applications to two-electron Schrödinger equations*. PhD thesis, Universitätsbibliothek, 2012.

[5] Ariel Fernández. Communication: Chemical functionality of interfacial water enveloping nanoscale structural defects in proteins. *JChemPhys*, 140(22):221102, 2014.

[6] Alfonso De Simone, Guy G. Dodson, Chandra S. Verma, Adriana Zagari, and Franca Fraternali. Prion and water: Tight and dynamical hydration sites have a key role in structural stability. *Proceedings of the National Academy of Sciences, USA*, 102:7535–7540, 2005.

[7] Ariel Fernández, Jianping Chen, and Alejandro Crespo. Solvent-exposed backbone loosens the hydration shell of soluble folded proteins. *Journal of Chemical Physics*, 126(24):245103, 2007.