History of the Wolynes Hydrogen Bond Potential

Finley Clark

September 20, 2025

Purpose

This document discusses the hydrogen bonding potential as presented in Wolynes group journal articles. The implementation is addressed elsewhere (see the code and user guide).

Background

The hydrogen bonding potential energy term used by the Wolynes group has a long history and has sometimes been printed and implemented incorrectly. The first known use of a hydrogen bonding potential is in the 1999 paper introducing the off-lattice backbone.[1] It was then elaborated in a paper focused on alpha/beta proteins[2] into basically the modern beta hydrogen bonding potential.[3,4] A 2006 structure prediction paper[5] updated the parameters of the beta hydrogen bonding term and added the "liquid crystal" (P-AP) hydrogen bonding term. Finally, an alpha-helical hydrogen bonding term was introduced in 2010.[6] The total potential is therefore

$$V_{HB} = V_{\alpha} + V_{\beta} + V_{P-AP} \, .$$

The beta term, liquid crystal term, and a simplified (density-independent) version of the alpha-helical hydrogen bonding term were combined in the LAMMPS AWSEM-MD potential.[4] The alpha-helical term was removed from the default OpenAWSEM potential,[3] probably because helix formation is aided by local-in-sequence interactions that can be captured by the fragment memory term.

Alpha-Helical Term

The alpha-helical term can stabilize or destabilize helical configurations of a pair of residues at indices *i* and *i+4*.

$$V_\alpha = \sum_{i=1}^{N-4} V_\alpha^i$$

$$V_\alpha^i = -\big(f(a_i) + f(a_{i+4})\big)\theta_{i,i+4}\big(\gamma_{prot}\sigma_{i,i+4}^{prot} + \gamma_{wat}\sigma_{i,i+4}^{wat}\big)$$

$$\theta_{i,i+4} = \exp\left(-\frac{\left(r_{i,i+4}^{ON} - \langle r^{ON}\rangle\right)^2}{2\sigma_{NO}^2} - \frac{\left(r_{i,i+4}^{OH} - \langle r^{OH}\rangle\right)^2}{2\sigma_{HO}^2}\right)$$

Original Presentation

To my knowledge, the alpha-helical potential was first used in 2010.[6] This subsection describes the potential used in this reference.

A pair of residues, *i* and *i+4*, is considered to be in an alpha-helical configuration when the following geometric indicator function is near its maximum

$$\exp\left(-\frac{\left(r_{i,i+4}^{ON} - \langle r^{ON}\rangle\right)^2}{2\sigma_{NO}^2} - \frac{\left(r_{i,i+4}^{OH} - \langle r^{OH}\rangle\right)^2}{2\sigma_{HO}^2}\right)$$

where $r_{i,i+4}^{AB}$ is the distance between atoms of types *A* and *B* belonging to residues *i* and *i+4*, respectively; $\langle r^{AB}\rangle$ is the center of the gaussian potential governing the bond distance; and σ_{AB}^2 sets the width of the gaussian potential governing the bond distance between atom types *A* and *B* belonging to residues *i* and *i+4*. The $\langle r^{AB}\rangle$ and (σ_{AB}^2) are constants (identical for all *i*) that were presumably chosen to be appropriate for the ideal peptide bond geometry used in the model and the resolution of the model. Note that, in general, $r_{i,i+4}^{ON} \neq r_{i,i+4}^{NO}$ and $r_{i,i+4}^{OH} \neq r_{i,i+4}^{HO}$.

Additionally, the atom type "N" stands for the backbone nitrogen, "O" for the backbone oxygen, and "H" for the backbone hydrogen, if it exists. The right-handed chirality of the alpha helix dictates that residue *i* provides the O atom and residue *i+4* provides the N and H atoms, so we see that the backbone H atom does not exist if and only if residue *i+4* is proline. The paper does not explain how this case should be handled. While we do not have access to the source code, I believe that the potential takes the coordinates of a dummy hydrogen atom calculated deterministically following the ideal peptide bond geometry in the same way as the hydrogen atom coordinates are calculated for other residues (see LAMMPS discussion).

Once the geometric helical indicator function has been evaluated, it must be weighted based on the identities of the amino acids at positions *i* and *i+4*. A negative weight gives a favorable

interaction, while a positive weight makes it unfavorable for the pair to adopt a helical geometry. For simplicity, it is assumed that the true weighting function can be represented as a sum of single-residue helical propensities, and that the propensity of a residue to be at position *i+4* (hbond donor) is equal to its propensity to be at position *i* (hbond acceptor). The weighting function is therefore

$$-\big(f(a_i) + f(a_{i+4})\big)$$

where a_j is the amino acid type of residue *j* and *f* is the single-residue helical propensity. The paper does not directly give helical propensities but says that they come from an old paper.[7]

Finally, the propensity-geometry product is multiplied by density-based indicators:

$$\gamma_{prot}\sigma_{i,i+4}^{prot} + \gamma_{wat}\sigma_{i,i+4}^{wat}$$

where $\gamma_{prot} = 2$ (favorable, according to the paper's sign convention), $\gamma_{wat} = -1$ (unfavorable), and σ^{wat} and $\sigma^{prot} = 1 - \sigma^{wat}$ have the same meaning as in the burial potential. This term therefore interpolates between one weight for a completely solvent-exposed environment and another for a completely buried environment. The paper states that the helical hydrogen bonding in the solvent-exposed environment is made unfavorable because amino acid residues have many opportunities to form hydrogen bonds with the solvent.

AWSEM-MD Presentation

The SI of the paper[4] introducing the LAMMPS implementation of the "AWSEM-MD" potential acknowledges the issue of having a proline residue at an *i+4* position. It states that a strongly unfavorable propensity $f_{i+4}(\mathrm{PRO}) = -3$ is used when proline is at the *i+4* position, compared to a moderate weight of $f_i(\mathrm{PRO}) = 0.4$ for proline at the *i* position. The LAMMPS code clearly uses a dummy H atom for proline at the *i+4* position (see the Hbond Implementation document). These findings support the interpretation of the original potential as somehow including the proline at the *i+4* position by way of a dummy atom.

The AWSEM-MD paper gives a table of helical propensities that seem to differ from the paper[7] that provides helical propensities for the original potential. I don't know where this table of propensities comes from.

Otherwise, I believe that the AWSEM-MD helical potential is identical to the potential used in the original hydrogen bonding paper.

OpenAWSEM Presentation

The OpenAWSEM paper[3] does not mention the helical hydrogen bonding term, but the code offers a density-independent potential, where $\gamma_{prot}\sigma_{i,i+4}^{prot} + \gamma_{wat}\sigma_{i,i+4}^{wat}$ is dropped.

Beta Term

The beta term is the most complex of the three hydrogen bonding terms, but it has a similar form to the alpha term: some indicator functions to quantify the similarity between a given configuration and an ideal hydrogen bond, multiplied by a numerical weight that depends on the amino acid sequence. For this term, the geometric indicator function has two parts: θ and ν. The potential is mostly described by the following equations, but see "Knowledge-Based Adjustments."

$$V_{\beta} = \sum_{i} \sum_{j \neq i} \nu_i^I \nu_j^{II} \left(V_{\beta 1}^{ij} + V_{\beta 2}^{ij} + V_{\beta 3}^{ij} \right)$$

$$\nu_k^{\mu} = \begin{cases} \frac{1}{2}\left(1 + \tanh\left(\eta^{\mu} \times \left(r_{k-2,k+2}^{C_\alpha C_\alpha} - 12\text{Å}\right)\right)\right), & \text{k} - 2 \text{ and k} + 2 \text{ exist} \\ 1, & \text{k} - 2 \text{ or k} + 2 \text{ does not exist} \end{cases}$$

$$\eta^I = 1 \text{ Å}^{-1}$$

$$\eta^{II} = 0.5 \text{ Å}^{-1}$$

$$V_{\beta 1}^{ij} = \begin{cases} 0, & \text{i at chain end or j at chain start or j is proline} \\ \Lambda_1(|i-j|)\theta_{ij}, & \text{otherwise} \end{cases}$$

$$V_{\beta 2}^{ij} = \begin{cases} 0, & (i \text{ or } j \text{ at chain end or chain start}) \text{ or } (i \text{ or } j \text{ is proline}) \\ \Lambda_2\left(|i-j|, a_i, a_j, a_{i-1}, a_{j-1}, a_{i+1}, a_{j+1}\right)\theta_{ij}\theta_{ji}, & \text{otherwise} \end{cases}$$

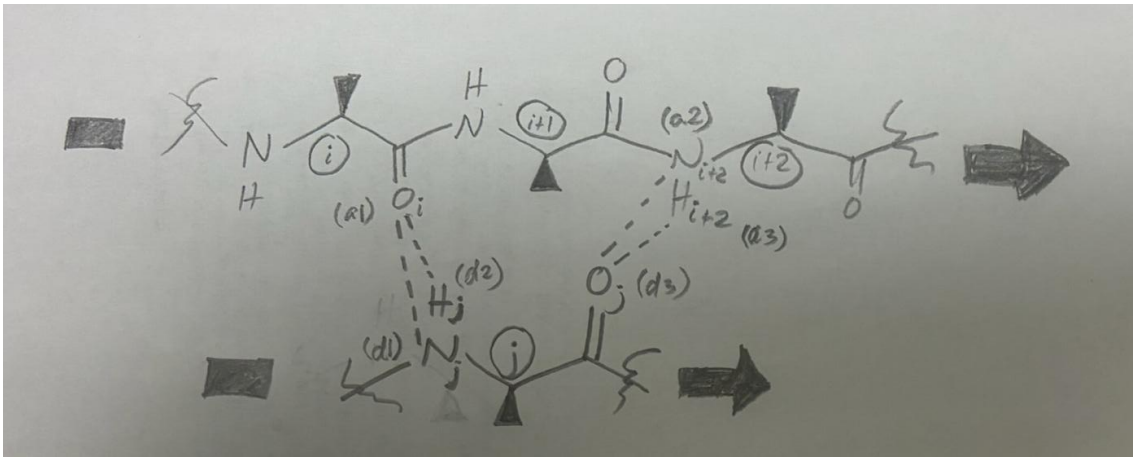
$$V_{\beta 3}^{ij} = \begin{cases} 0, (\text{i and i} + 2 \text{ not in same chain}) \text{ or } (\text{j at chain start or end}) \text{ or } (\text{j or i} + 2 \text{ is proline}) \\ 0, \quad \text{i} + 2 \text{ does not exist} \\ \Lambda_3\left(|i-j|, a_j, a_{i+1}\right)\theta_{ij}\theta_{j,i+2}, \quad \text{otherwise} \end{cases}$$

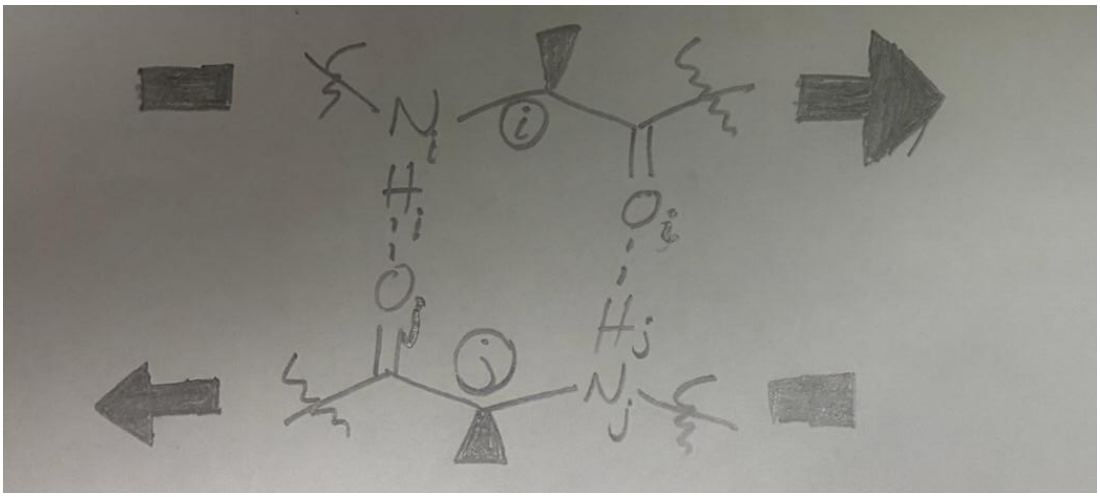
Geometric Indicator Functions

The θ function here has the same form as it does in the alpha hydrogen bonding term, with the same $\langle r^{AB} \rangle$ and σ_{AB}^2 parameters, except the gaussian widths are written as σ_{NO}^2 and σ_{HO}^2 instead of σ_{ON}^2 and σ_{OH}^2. Due to the double summation, however, the θ function is, given a residue *i*, evaluated for many potential hydrogen bonding partners *j*, not the just the *i+4* residue. While the double sum runs over *all* pairs of residues (*i*, *j*), the conditional definitions of $V_{\beta 1}^{ij}$, $V_{\beta 2}^{ij}$, and $V_{\beta 3}^{ij}$ cause the functions θ_{ij}, θ_{ji}, and $\theta_{j,i+2}$ to not be evaluated in some cases (i.e., $V_{\beta N}^{ij} = 0$ so it doesn't matter what θ_{ij} is). One reason for using these conditional definitions is that one or more of the θ functions cannot be defined for some pairs (*i*, *j*).

There are two potential issues resulting in undefined θ functions: (1) missing atoms and (2) missing residues. Missing atoms occur in θ_{ij}, θ_{ji}, and $\theta_{j,i+2}$ when residues *j*, *i*, and *i*+2, respectively, are prolines, which chemically lack a backbone hydrogen atom (unlike what has sometimes been for the alpha helical potential, a dummy H atom is never used for proline in the beta potential). Missing N and H atoms occur in the first residue of each chain because defining their coordinates requires the coordinates of atoms in the previous residue in a chain, which does not exist (this is not an issue for the alpha helical potential because the first residue in a chain can never be located at the *i+4* position). Missing residues occur when we try to evaluate $\theta_{j,i+2}$ but residue *i* is the last or second-to-last residue in its chain.

The functions $V_{\beta 1}^{ij}$, $V_{\beta 2}^{ij}$, and $V_{\beta 3}^{ij}$ involve different combinations of the θ terms. The lone θ_{ij} works in a pairwise manner, determining whether the O acceptor group from residue *i* and the N-H donor group from residue *j* are roughly colinear at the correct distance to engage in a hydrogen bond. This is a valuable geometric indicator that contributes to the overall potential through $V_{\beta 1}^{ij}$. However, we also want to sharpen the folding transition through cooperative antiparallel beta sheet and parallel beta sheet interactions. We get antiparallel cooperativity by computing $\theta_{ij}\theta_{ji}$, which rewards residues *j* and *i* making two hydrogen bonds with each other (each with one donor and one acceptor group), which requires an antiparallel conformation of the two residues. We get parallel cooperativity by computing $\theta_{ij}\theta_{j,i+2}$, which rewards hydrogen bonding of residue *j* with both residue *i* and residue *i*+2, which requires a parallel alignment of the two residues.





Lambda Functions- Overview

The values Λ are constants that formally depend on the sequence separation, $|j - i|$, of the residues; however, energy landscape-based optimization of the parameters inside Λ_1, Λ_2, and Λ_3 found only small numerical differences between the optimal values for some parameters, while available data were too limited to optimize other parameters in a sequence separation-dependent manner.[5] It seems likely that there is only a weak, if any, dependence of these parameters upon sequence separation. While certain physical effects, such as the backbone dihedral energy landscape and backbone loop entropy, might be expected to affect hydrogen bonding behavior in a manner dependent upon sequence separation, these effects should be captured by other terms in the potential. Residues belonging to different chains are treated as residues of the largest intrachain sequence separation class.

The Λ terms paired with cooperative indicator functions, Λ_2 (with $\theta_{ij}\theta_{ji}$) and Λ_3 (with $\theta_{ij}\theta_{j,i+2}$), also depend on the propensities of different amino acid types to be in an antiparallel or parallel beta sheet, respectively. Specifically, Λ_2 depends on the amino acid identities of residues *i*-1, *i*, *i*+1, *j*-1, *j*, and *j*+1, written as a_{i-1}, a_i, a_{i+1}, a_{j-1}, a_j, and a_{j+1}, respectively, while Λ_3 depends on a_{i+1} and a_j. There are both per-residue propensities (quantitatively, probabilities in a structural dataset[2]) for parallel and antiparallel sheets, P_{anti} and P_{par}, and couplings between pairs of amino acids across from each other on adjacent beta strands (see the pair (*i*, *j*) in the antiparallel example in the figure above and (*i*+1, *j*) in the parallel example). In adjacent parallel beta strands, a pair of residues directly across from each other always contains one residue that makes two hydrogen bonds with the opposite strand and one residue that makes zero hydrogen bonds with the opposite strand, while, in adjacent antiparallel strands, two residues across from each other may be hydrogen bonded (HB) to each other or not hydrogen bonded to any residue in the opposite strand (NHB).[2,8] HB and NHB pairs have different amino acid type-dependent energetics.[9] Therefore, two sets of coupling parameters, P_{HB} and P_{NHB}, are used for antiparallel pairs and only one set of parameters, called P_{parHB} or $P_{par}(a_{i+1}, a_j)$ (to emphasize that it takes two arguments, in contrast to the per-residue $P_{par}(a_{i+1})$), is used for parallel pairs.[2]

There are some pairs (*i*, *j*) for which one or more of the Λ expressions is undefined. For many such pairs, one or more of θ_{ij}, $\theta_{ij}\theta_{ji}$, and $\theta_{ij}\theta_{j,i+2}$ is also undefined. The only case that I am aware of where a Λ function cannot be defined but its corresponding geometric indicator function can be defined is Λ_2 in the case that *j* is at the end of a chain and *i* is not proline or at the start or end of a chain. In this case, there is no issue calculating $\theta_{ij}\theta_{ji}$, but a_{j+1} does not exist, so Λ_2 cannot be evaluated by the usual expression. It is instead set to 0 by the conditional in the function definition. See also the following "Note on the Conditionals…"

Note on the Conditionals in $V_{\beta 1}^{ij}$, $V_{\beta 2}^{ij}$, and $V_{\beta 3}^{ij}$

There are some pairs (*i*, *j*) for which all θ and Λ functions can be defined that still result in $V_{\beta 1}^{ij}$, $V_{\beta 2}^{ij}$, and $V_{\beta 3}^{ij}$ being set to 0 by the conditionals, without evaluating the θ functions. I don't know why this is done. See github.com/cabb99/openawsem/issues/58 for more information.

Lambda Functions - Initial Definition

The first paper to introduce V_β in its modern form[2] defines the Λ functions in the following way, where the λ and α values are the sequence separation-dependent optimized parameters:

$$\Lambda_1 = \lambda_1$$

$$\Lambda_2 = \lambda_2 - \alpha_1 \ln P_{anti}(a_i) + \alpha_1 \ln P_{anti}(a_j) + 0.5\alpha_2 \ln P_{HB}(a_i, a_j) - 0.25\alpha_3\left(\ln P_{NHB}(a_{i+1}, a_{j-1}) + \ln P_{NHB}(a_{i-1}, a_{j+1})\right)$$

$$\Lambda_3 = \lambda_3 - \alpha_4\left(\ln P_{par}(a_{i+1}) + \ln P_{par}(a_j)\right) + \alpha_5 \ln P_{par}(a_{i+1}, a_j)$$

While it is not clear how to optimally integrate the structural database information contained in the P values into our hydrogen bonding potential, we can wave our hands and say that taking the logarithm turns it into an energy whose proper weight relative to the other terms can be optimized by energy landscape theory.

The way the total potential is written, larger (more positive) Λ values imply a more favorable interaction. More favorable interactions should also be implied by higher probabilities P. Therefore, any reasonable definition of Λ_k requires

$$\frac{d\Lambda_k}{dP_{\text{type}}} > 0$$

for all applicable types. Given that $\alpha_1 < 0$ and all other constants are positive, we conclude that the former equation has several sign errors. The sign errors are corrected below. Note that α_1 has been replaced with its absolute value:

$$\Lambda_1 = \lambda_1$$

$$\Lambda_2 = \lambda_2 + |\alpha_1| \ln P_{anti}(a_i) + |\alpha_1| \ln P_{anti}(a_j) + 0.5\alpha_2 \ln P_{HB}(a_i, a_j) + 0.25\alpha_3 \left(\ln P_{NHB}(a_{i+1}, a_{j-1}) + \ln P_{NHB}(a_{i-1}, a_{j+1})\right)$$

$$\Lambda_3 = \lambda_3 + \alpha_4 \left(\ln P_{par}(a_{i+1}) + \ln P_{par}(a_j)\right) + \alpha_5 \ln P_{par}(a_{i+1}, a_j)$$

But we're not done yet. The indices of the α parameters in the above equation are incorrect. They are fixed below, giving the completely correct equations:

$$\Lambda_1 = \lambda_1$$

$$\Lambda_2 = \lambda_2 + \alpha_3 \ln P_{anti}(a_i) + \alpha_3 \ln P_{anti}(a_j) + 0.5|\alpha_1| \ln P_{HB}(a_i, a_j) + 0.25\alpha_2 \left(\ln P_{NHB}(a_{i+1}, a_{j-1}) + \ln P_{NHB}(a_{i-1}, a_{j+1})\right)$$

$$\Lambda_3 = \lambda_3 + \alpha_5 \left(\ln P_{par}(a_{i+1}) + \ln P_{par}(a_j)\right) + \alpha_4 \ln P_{par}(a_{i+1}, a_j)$$

Lambda Functions - Retraining

In the next paper to discuss the hydrogen bond term in detail,[5] the same incorrect equations from the original paper[2] were copied over. Because α_1 was changed from negative to positive in the more recent paper, we can drop the absolute value sign from the equations:

$$\Lambda_1 = \lambda_1$$

$$\Lambda_2 = \lambda_2 + \alpha_3 \ln P_{anti}(a_i) + \alpha_3 \ln P_{anti}(a_j) + 0.5\alpha_1 \ln P_{HB}(a_i, a_j) + 0.25\alpha_2 \left(\ln P_{NHB}(a_{i+1}, a_{j-1}) + \ln P_{NHB}(a_{i-1}, a_{j+1})\right)$$

$$\Lambda_3 = \lambda_3 + \alpha_5 \left(\ln P_{par}(a_{i+1}) + \ln P_{par}(a_j)\right) + \alpha_4 \ln P_{par}(a_{i+1}, a_j)$$

Lambda Functions - AWSEM-MD Paper

The paper introducing the "AWSEM-MD" potential implemented in LAMMPS[4] also makes sign errors and has an incorrect index on one of the α terms. The correct equations are the same as those in the previous subsection.

Lambda Functions- OpenAWSEM Paper

The OpenAWSEM paper[3] does not write out the Λ function definitions, so there's nothing to talk about here.

Knowledge-Based Adjustments

Some adjustments are typically made to V_β:

1. Parallel beta sheets forming over short sequence separations are rare. Since the retraining paper[5] (or maybe only in the retraining paper), the parallel hydrogen bonding potential has for short sequence separations been removed by setting $\Lambda_3 = 0$. This is probably unnecessary because it is very difficult to get the backbone into a conformation that would result in parallel beta sheets at a short sequence separation.
2. The retraining paper[5] says that Λ_2 and Λ_3 were set to 0 for pairs (*i*, *j*) unless both residues were predicted to be in a beta sheet. The AWSEM-MD[4] and OpenAWSEM[3] paper do not mention this behavior, but the LAMMPS code actually does something similar: it sets Λ_1, Λ_2, and Λ_3 equal to 0 for short sequence separations unless both residues are believed to be in a beta sheet, but will not set any of the Λ functions to 0 at longer sequence separations.

Nu Functions

Recall that the sum of the pairwise, antiparallel, and parallel beta terms for a pair of residues is multiplied by "nu" terms:

$$\nu_i^I \nu_j^{II} \left(V_{\beta 1}^{ij} + V_{\beta 2}^{ij} + V_{\beta 3}^{ij} \right)$$

Each of the ν terms switches between 1, when the protein chain is highly extended in the vicinity of the residue indicated in the subscript, and 0, when the chain is highly collapsed. The superscript Roman numeral sets the sharpness of the transition. The two terms work together to make interactions less favorable when either residue *i* or residue *j* is sterically surrounded by its own chain and unable to fully incorporate into a beta sheet.

Sometimes, an atom/residue needed to calculate a ν term does not exist. In this case, the ν term is set to 1.

The ν terms were introduced in 2006 in what I've been calling the "retraining" paper[5] as part of θ_{ij}. In the AWSEM-MD paper, their form was modified and they were written separately from the θ terms. They were then removed in OpenAWSEM by default to aid efficiency.[3] Removing these terms allows us to implement the potential as a CustomHbondForce (http://docs.openmm.org/latest/api-python/generated/openmm.openmm.CustomHbondForce.html) instead of a CustomCompoundBondForce (http://docs.openmm.org/latest/api-python/generated/openmm.openmm.CustomHbondForce.html).

Liquid Crystal Term

The liquid crystal, or P-AP (parallel-antiparallel), term is assembled from terms that are similar to but not identical to the ν terms used in the beta potential.

$$V_{P-AP} = \sum_{i}^{N-4} \sum_{j>i} \nu_i^{EO} * \left[-\gamma_{AP} \left(\text{CONDITION}_{\text{AP}} * v_{ij}^{P-AP} \nu_{i+4,j-4} \right) - \gamma_P (\text{CONDITION}_{\text{P}} * v_{ij}^{P-AP} \quad \nu_{i+4,j+4}) \right]$$

$\text{CONDITION}_{\text{AP}}$: ($j \geq i + 13$ OR *i*, *j* in different chains)

AND (*i*, *i*+4 in same chain) AND (*j*, *j*-4 in same chain*)

$\text{CONDITION}_{\text{P}}$: $j \geq i + 9$ OR *i*, *j* in different chains

AND (*i*, *i*+4 in same chain) AND (*j*, *j*+4 in same chain)

$$\nu_i^{EO} = \frac{1}{2}(1 + \tanh(7\text{Å}^{-1} \times (r_{i,i+4}^{C_\alpha C_\alpha} - 12\text{Å})))$$

$$\nu_{ij} = \frac{1}{2}\left(1 + \tanh\left(7\text{Å}^{-1} \times \left(8\text{Å} - r_{ij}^{C_\alpha C_\alpha}\right)\right)\right)$$

*Formally, residue *j*-4 does not exist for *j*<5. This has the same physical significance as *j* and *j-4* being in different chains: *j* cannot participate in an antiparallel interaction.

Coordination with Efficiency-Optimized Beta Term

Normally, v_i^{EO} should actually be set to 1 for all *i*. The functional form above only comes into play when we want to run the "efficiency-optimized" HB potential, which eliminates $\nu_i^I \nu_j^{II}$ from the Beta potential (sets it to 1).

Clashing Variable Names

Historically, the symbol ν has been used in the Beta and P-AP equations in a few different ways. To try to be consistent with the previous ways of writing the equations, I've kept the symbol ν but added subscripts and superscripts to distinguish the different types of ν.

A single subscript with a Roman numeral superscript, as in ν_i^I and ν_i^{II}, refers to the ν in the Beta equations, which indicates the compactness of the chain around residue *i*. The superscript indicates the two different possible η (sharpness) values for the switching between the maximum and minimum values of the sigmoid-like function.

In the "efficiency-optimized" (EO) hydrogen bond potential, ν_i and ν_j are set to 1 and its effects on the overall potential are approximated by adding a similar factor to the P-AP term. This factor, called ν_i^{EO}, has a slightly different functional form. The critical distance for switching between active and inactive is slightly larger than in the Beta case because the P-AP potential is supposed to be more forgiving of geometric deviations from an ideal beta sheet. Also, it measures the distance between residues *i* and *i+4* instead of *i-2* and *i+2*. This is only for efficiency (see note in hydrogenBondTerms.py file).

Also part of the P-AP term is ν_{ij}, a switching function with a similar form to the other ν terms, except that it depends on the distance between two residues that are far in sequence and is not measuring local chain compactness.

Knowledge-Based Adjustments

The liquid crystal potential is sometimes adjusted based on the predicted or known secondary structure. Specifically, in the case that both residues *i* and *j* are viewed as beta-strand residues, the interaction strength for the pair (i, j) may be multiplied by a weight of 1.5, unless the residues are in an antiparallel alignment with $|i - j| < 17$.

References

(1) Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. Backbone Dynamics, Fast Folding, and Secondary Structure Formation in Helical Proteins and Peptides. *Proteins Struct. Funct. Genet.* **1999**, *34* (3), 281–294.

(2) Hardin, C.; Eastwood, M. P.; Prentiss, M. C.; Luthey-Schulten, Z.; Wolynes, P. G. Associative Memory Hamiltonians for Structure Prediction without Homology: α/β Proteins. *Proc. Natl. Acad. Sci.* **2003**, *100* (4), 1679–1684.

(3) Lu, W.; Bueno, C.; Schafer, N. P.; Moller, J.; Jin, S.; Chen, X.; Chen, M.; Gu, X.; Davtyan, A.; De Pablo, J. J.; Wolynes, P. G. OpenAWSEM with Open3SPN2: A Fast, Flexible, and Accessible Framework for Large-Scale Coarse-Grained Biomolecular Simulations. *PLOS Comput. Biol.* **2021**, *17* (2), e1008308.

(4) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116* (29), 8494–8503.

(5) Zong, C.; Papoian, G. A.; Ulander, J.; Wolynes, P. G. Role of Topology, Nonadditivity, and Water-Mediated Interactions in Predicting the Structures of α/β Proteins. *J. Am. Chem. Soc.* **2006**, *128* (15), 5168–5176.

(6) Oklejas, V.; Zong, C.; Papoian, G. A.; Wolynes, P. G. Protein Structure Prediction: Do Hydrogen Bonding and Water-Mediated Interactions Suffice? *Methods* **2010**, *52* (1), 84–90.

(7) Pace, C. N.; Martin Scholtz, J. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998**, *75* (1), 422–427.
(8) Merkel, J. S.; Sturtevant, J. M.; Regan, L. Sidechain Interactions in Parallel β Sheets: The Energetics of Cross-Strand Pairings. *Structure* **1999**, *7* (11), 1333–1343.
(9) Wouters, M. A.; Curmi, P. M. G. An Analysis of Side Chain Interactions and Pair Correlations within Antiparallel B-sheets: The Differences between Backbone Hydrogen-bonded and Non-hydrogen-bonded Residue Pairs. *Proteins Struct. Funct. Bioinforma.* **1995**, *22* (2), 119–131.