

Divining Deamidation and Isomerization in Therapeutic Proteins: Effect of Neighboring Residue

Flaviyan Jerome Irudayanathan¹, Jonathan Zarzar², Jasper Lin² and Saeed Izadi^{2*}

¹ Human Genetics Department, Genentech Inc.

² Pharmaceutical Development Department, Genentech Inc.

Correspondence should be addressed to: * izadis@gene.com

Abstract

Deamidation of asparagine (ASN) and isomerization of aspartic acid (ASP) residues are among the most commonly observed spontaneous post-translational modifications (PTMs) in proteins. Understanding and predicting a protein sequence's propensity for such PTMs can help expedite protein therapeutic discovery and development. In this study, we utilized proton-affinity calculations with semi-empirical quantum mechanics (QM) and μ s long equilibrium molecular dynamics (MD) simulations to investigate mechanistic roles of structure and chemical environment in dictating spontaneous degradation of asparagine and aspartic acid residues in 131 clinical-stage therapeutic antibodies. Backbone secondary structure, side-chain rotamer conformation and solvent accessibility were found as three key molecular indicators of ASP isomerization and ASN deamidation. Comparative analysis of backbone dihedral angles along with N-H proton affinity calculations provides a mechanistic explanation for the strong influence of the identity of the n+1 residue on the rate of ASP/ASN degradation. With these findings, we propose a minimalistic physics-based classification model that can be leveraged to predict deamidation and isomerization propensity of therapeutic proteins.

Introduction

Protein stability is at the center of biological mechanisms spanning molecular pathways and therapeutic interventions.[1-3] Proteins are subject to various post-translational modifications (PTM) that are essential for their function, e.g. glycosylation and acylation. Some of these PTMs are spontaneous chemical reactions that lead to degradation, e.g. deamidation of asparagine (ASN→ASP/iso-ASP), isomerization of aspartic acid (ASP→iso-ASP) and oxidation of MET/TRP residues.[4, 5] These spontaneous chemical modifications often serve as a key indicator of stress that elicits a physiological response.[6] Bio-manufactured therapeutic proteins such as peptides, vaccines, monoclonal and bispecific antibodies (mAbs), and viral capsids are subject to physical and chemical stressors during manufacturing that result in accumulation of various PTMs.[7-11] Spontaneous chemical modification can negatively impact the biomolecule's therapeutic efficacy, quality and developability by altering the structure and/or molecular properties of the protein. For instance, isomerization of an ASP residue alters the backbone connectivity and deamidation of an ASN residue imparts a negative charge (Figure1), both of which can result in altered functional consequences.[12, 13] Often protein engineering strategies are employed to mitigate the negative impacts of such chemical modifications, but this process is often iterative which prolongs the development cycle of a successful therapeutic.[14-16] Understanding and predicting the propensity of a therapeutic protein to undergo chemical degradation under stress can be leveraged to expedite development timelines.[17-20]

Deamidation of an ASN residue has been experimentally well characterized.[21-23] Among the multiple pathways that are feasible, pH-dependent base-catalyzed nucleophilic attack has been shown to be the predominant pathway.[23-26] The reaction is irreversible and proceeds via (Figure 1): 1) Base-catalyzed deprotonation of the n+1 amide 2) nucleophilic attack of the anionic nitrogen on the side chain carbonyl group and a subsequent ring closure that leads to the formation of a succinimide intermediate and 3) hydrolysis of the succinimide into ASP or iso-ASP. Similarly, the isomerization reaction can proceed from an ASN or ASP residue via the same succinimide intermediate pathway. There are mechanistic parallels and kinetic dissimilarities between the isomerization and deamidation reactions.[13, 27, 28] A comprehensive investigation of both reactions has revealed that the side chain of the neighboring n+1 residue can explain fold differences in the $t_{\frac{1}{2}}$ of the reaction in peptides.[21, 29-32] Previous studies using

pentapeptides by Robinson et al.[33] have shown a higher likelihood of chemical degradation if the n+1 neighboring residue of the ASN residue is a GLY, ALA, SER or THR. Based on this, the sequences NX and DX (where X could be G/A/S/T) are often considered as sequence hotspots in proteins. However, these peptide hotspots do not sufficiently predict the reaction propensity in 3D folded proteins; a recent experimental survey of 131 clinical stage therapeutic (CST) antibodies (henceforth referenced as the Adimab dataset) by Lu et al. [34] revealed vast discrepancies in the range of deamidation and isomerization propensities among antibodies that do not follow the hotspot rules. For instance, of the twenty-seven NG hotspots only fourteen underwent chemical degradation. Similarly, of the forty-four DG cases only sixteen were isomerized. In addition, chemical degradation was observed at the non-hotspot residue locations such as NY, NW, DR and DY. The discrepancy reported in Lu et al. indicates a gap in our knowledge of how neighboring residues contribute to the chemical degradation propensity.

Multiple computational strategies, from physics-based approaches (e.g. quantum mechanics/molecular mechanics (QM/MM) and Molecular Dynamics (MD) simulations) [29, 30, 35, 36] to machine-learning models,[17, 20, 37] have been developed over the past two decades to predict deamidation and isomerization in peptides and proteins. Several detailed computational models explored site specific effects using QM/MM calculations at the peptide level. For example, ab initio calculations on model compounds have revealed the contribution of backbone secondary structure affecting the proton-affinity (Figure 1a) of the n+1 amide.[38] However, it is as of yet unclear if these insights are generalizable to ASP and ASN residues with varied neighboring residues on a large set of proteins. Another noteworthy study[36] used MD and QM/MM to compute conformational and chemical free energy barriers along the deamidation reaction pathway. While this study provided detailed insights into the free energy changes, the method was only tested on a small number of degradation sites on a single molecule and thus would be computationally prohibitive for high throughput screening. Recently, multiple machine learning models have been proposed that were trained with a number of structure-based features such as secondary structure, local flexibility, size of the neighboring residue, and solvent exposure.[17, 20, 37, 38] While computationally facile and practical in nature, these models often ignore important mechanistic and structural details (such as conformational dynamics) that are important for the reaction's feasibility. Besides, it remains unclear how broadly generalizable these models are, given that they were trained on a limited set of data points.

Here we have utilized microsecond long equilibrium molecular dynamics (MD) simulations and proton-affinity calculations to understand the impact of neighboring residues in dictating the chemical degradation at ASP/ASN site of proteins. Based on the analysis of a large set of 131 CST antibodies, we propose a physics-based classifier model that can accurately predict the propensity of deamidation and isomerization of a protein sequence. The model relies on decomposing the major steps of the reaction (deprotonation, nucleophilic attack and hydrolysis) as conformational degrees of freedom and inferring the reaction propensity based on the thermodynamic feasibility of exploring those conformations.

Results and Discussion

Conformational Determinants of Deprotonation.

QM calculations and crystal structure analysis. Deprotonation of the amide nitrogen has been experimentally characterized to proceed via a base-catalyzed reaction that is quite spontaneous even at acidic conditions.[31, 39-41] Several studies have investigated the dependence of backbone amide acidity on the conformation and chemical environment using QM/QMMM calculations.[29, 30, 35-37] In one of the pioneering works, Radkiewicz et al. [26, 27] used N-formyl-glycinamide as a model compound representing a peptide bond to calculate the relative proton affinity as a function of the ϕ and ψ dihedral angles (using HF/6-31+G**/HF/3-21G). This analysis showed that the acidity of backbone amide is largely impacted by of the ψ dihedral angle; ϕ angle was found to have a small effect. Here, we carried out DFT calculation to estimate proton affinity for the same model compound at the M062X/6-311++G(d,p) level of theory in both gas phase and aqueous phase continuum SMD (Solvation model based on density) solvent environment. Our results are in agreement with the earlier studies: proton affinity of backbone amide deviates significantly with respect to changes in the backbone conformation, Figure 2b and the SI. Specifically, when ϕ and ψ are at angles observed within α -helices, the proton bonded to the amide experiences a repulsive force from the adjacent carbonyl carbon as a result of a di-electric instability with the amide nitrogen (Figure 2c) i.e. proton affinity of the model compound is lower in the α -helical conformation centered around $-60^\circ \leq \psi \leq 60^\circ$, Figure 2b. The gas phase semi-empirical QM calculations using MOPAC with PM6-D3H4,47 PM7 and RM1 Hamiltonians (see methods and the SI) are also consistent with DFT results.

We then sought to find the generalizability of this conformation-dependent amide acidity rules in the context of peptides. In order to do so we extended the proton affinity calculations to DX and NX dipeptides as a function of their backbone ϕ & ψ dihedral angles, Figure 3 and the SI. The goal was to investigate if proton affinity calculations can explain the fold differences of reaction half-lives observed in pentapeptides due to the variations in the n+1 residue. For instance, an interesting comparison can be made between Gly and Ala as n+1 residue. In solution, GGNGG and GGNAG peptides show high propensity to adopt an α -helical conformation[41, 42] yet their deamidation half-lives are an order of magnitude different (1.03 and 21.1 days respectively).[33] Our dipeptide proton affinity analysis suggests that the amides in both NG and NA are acidic dependent on the ψ angle, $-60^\circ \leq \psi \leq 60^\circ$ being acidic (Figure 3a). However, contrary to the model compound data, the proton affinity of the amide in the NA peptide significantly changes as a function of the ϕ angle. Specifically, the backbone amide of the NA residue is more acidic in conformations when $\phi > 0$ compared to when $\phi < 0$ within the same boundary of $-60^\circ \leq \psi \leq 60^\circ$. This indicates that the proton affinity NA peptide is also dependent on the type of the helical conformation (right-handed α -helix vs left-handed α -helix). A right-handed α -helix conformation places the methyl sidechain group in close proximity to the backbone amide that offers protection from exposure to the base and/or hydrolysis (Figure 3c, conformation 2). Similarly it can be inferred from the proton affinity plots (see SI) that: 1) the sidechain mediated protection increases with increasing the size of non-polar, hydrophobic groups (i.e. Phe > Ile > Leu > Val > Ala), and 2) the presence of a polar hydroxyl or carbonyl groups in the n+1 side-chain (as in Asp, Ser, Thr, Asn, Tyr) can facilitate the deprotonation of the amide by solvent mediated electrostatic interactions. In the case of GLY or in any left-handed α -helix conformation the protection from the n+1 residue is completely lost (Figure 3c, conformation 3).

It must be noted that a large fraction of combinations of ϕ & ψ dihedral angles explored in the dipeptide models are energetically inaccessible in folded proteins. To identify the relevant secondary structures for the NX and DX motifs in proteins, we collected ϕ & ψ angle distribution from all the antibody crystal structures available from the PDB database. These distributions were analyzed using kernel density estimation (KDE) to arrive at the probability density function (PDF) of conformational preference for [N/D]G and [N/D]A residues in ϕ & ψ space (Figure 3b and SI). Comparing the proton affinity calculations with the conformational PDF reveals that [N/D]G sites mainly populate the acidic left-handed α -helical

conformations, whereas [N/D]A sites largely favor the basic right-handed helix ($\phi < 0$ region) and a beta sheet. The latter is in agreement with a previous work showing that a steric interaction between the methyl group and the terminal carbonyl oxygen destabilizes all structures with $\phi \sim 120^\circ$, relative to glycine dipeptides.[43] These results clearly explain why [N/D]A sites are often orders of magnitude less prone to degradation than [N/D]G sites.

A comprehensive analysis of the proton affinity calculations and conformational PDF in ϕ & ψ space for all the [N/D]X sites (SI) revealed that the frequency of observing a residue in the acidic region from the PDB is strongly correlated with the likelihood of the residues to undergo degradation. In particular, we observed that the most frequently modified [N/D]X sequences based on various publications (i.e. DG,[44-46] DS, [14, 47, 48] DD,[48] DT,[20] DH,[49] NG, NS, NN, and NT[12, 20, 33, 44]) mainly favor the right-handed helix in the Ramachandran plot, i.e. the more acidic region. Interestingly, the strongest right-handed helix population is seen for DG and DS, followed by DD, DN and DT. Similarly, for deamidation motifs, a decreasing population of right-handed helix is observed in the order of NG, NS, NN and NT. These results correlate with the approximate hierarchy of these hotspots based on published studies.[34]

Overall, these observations indicate that the ϕ & ψ backbone dihedral angle of the peptide bond at an [N/D]'s n+1 position can be used as a metric to identify sites that have higher propensity to deprotonate: 1) β -sheet conformation is less deprotonation prone, 2) the left-handed α -helix regions ($-60^\circ \leq \psi \leq 60^\circ$, $\phi > 0^\circ$) is acidic for all amino acids that have a non-polar (A, V, L, I) or bulkier side-chain groups (F, W), 3) On the contrary for small polar amino acids (G, S, T, D) both the left and right-handed α -helix ($-60^\circ \leq \psi \leq 60^\circ$) conformations are acidic.

Generalizability of the conformation dependent backbone acidity in proteins. To test the generalizability of the proton affinity data at the peptide-level, we explored the relationship between backbone conformations of all of the [N/D]X sites in the CDR region of 131 therapeutic monoclonal antibodies and their reported degradation rates. To obtain the backbone conformation preference, we calculated the conformational free energy surface (FES) in the ϕ & ψ space for the n+1 residue from microsecond long equilibrium molecular dynamics (MD) simulations of the Fab structures. We arbitrarily chose a value of 2KbT energy barrier or ~ 1 kcal/mol to define energetically favorable conformations, i.e. in the FES a region whose energy values are

< 1 kcal/mol are likely to be sampled under equilibrium conditions. This dataset corresponds to 1039 [N/D]X sites (498 ASP and 541 ASN) of which 30 DX and 39 NX sites were prone to degradation (reactive).

Figure 4 shows that the calculated conformational distribution (PDF) of reactive and non-reactive residues correspond very well with the expected conformational preference seen in proton affinity calculations. The experimentally reported non-reactive residues show preference to the right-handed α -helix or the β -sheet conformation while the reactive residues show large preference to the left-handed helical conformation. Figure 6 (D1) reveals a correlation between the rate of degradation and the probability of the backbone amide being found in the acidic regions. With an exception of two sites, the degradation rates fall below the desirable 5% cutoff for all the [N/D]X sites that are in a basic backbone conformation (i.e. Free Energy > 1kcal, which corresponds to either a beta-sheet or a right-handed helix when the n+1 residue is non-polar). However, the results in Figure 4 suggest that when the backbone is in an acidic conformation (Free Energy < 1kcal), the [D/N]X site is not necessarily guaranteed to undergo degradation. In other words, a high propensity for a conformation with a basic backbone hydrogen can prevent the site from degradation but lack thereof is not sufficient for degradation.

Overall, the backbone conformational preference of NX and DX residues under equilibrium conditions as a single descriptor was found to classify the reactive vs non-reactive residues at 65% accuracy (Table 1). This shows that proton affinity inferred from the secondary structure of n+1 residue is a good descriptor to predict degradation propensity of NX and DX sites.

Conformational determinants of cyclization.

Once deprotonation is enabled by the acidity of the backbone, the next step in the reaction is ring closure leading to succinimide formation via nucleophilic attack between N- atom of the n+1 residue and C γ of the ASN/ASP residue (Figure 1). In order for the nucleophilic attack to be feasible the N- and C γ should come closer than distances of 3.0 Å (near-attack conformation).[29, 30] In folded proteins sidechain interaction including hydrogen bond formation can limit the conformational flexibility of ASN/ASP residues to visit near attack conformations. This in turn can disfavor succinimide formation. Multiple sidechain orientations can

result in the near attack conformation (Figure 5a). Conformational degrees of freedom affecting the distance between C_γ and N_{n+1} can be defined as a function of the two dihedral angles ψ & χ_1 .

To identify the relevant conformations that enable a nucleophilic attack distance, we sampled the free energy space of ψ & χ angles in Ace-GGNAG-Nme pentapeptide using Metadynamics simulations. By applying an additive historical bias, Metadynamics pushes the system to explore conformations that are kinetically limited in an equilibrium simulation. There are six low energy states in the ψ & χ space (Figure 5b), of which two regions are coincident with the near attack conformation distance (Figure 5c).

With this information, we looked at the correlation between the near attack side-chain conformation and the experimental rates reported in the adimab dataset. As with the backbone conformation, we looked at the FES from the equilibrium MD simulation. Figure 6, D2 shows that when the side-chain is not in a near-attack conformation (i.e. Free Energy > 1kcal/mol), the reaction rates largely remain below the 5% cutoff, with a few exceptions. Otherwise, the residue may or may not be reactive, as suggested by the large number of data points below and above 5% for Free Energy < 1kcal/mol. Therefore, the side chain conformation indeed proved to be a useful descriptor in identifying reactive residues (Figure 6, D2). The overall accuracy of the prediction was at 40%, but when used in combination with the backbone conformation rule, i.e. both the acidic conformation and the near attack conformations are energetically feasible, the accuracy of the classification model improves to 75% (Table 1).

Solvent Accessibility

Water plays an important role in the deamidation and isomerization reaction as a proton donor/acceptor. Both the base catalyzed deprotonation and hydrolysis of succinimide intermediate to ASP/iso-ASP is only feasible in the presence of water molecules (Figure 1). In previous studies, solvent accessibility weakly predicted the reaction propensity.[19, 20, 33] These previous calculations were done on static protein structures that lacked proper structural relaxations that may affect solvent accessibility calculations. Our solvent accessibility calculations utilize a dynamic approach in combination with intuitions about the reaction. Here we calculated the solvent accessibility at a residue level. In order to capture the dynamic aspect of the solvent occupancy, we represented solvent accessibility as a fraction of time in which the

ASN/ASP residue remained accessible vs inaccessible (see methods). Using this value, we once again sought to classify the ASN/ASP residues in the adimab set as reactive vs non-reactive. Figure 6, D3 shows that solvent inaccessible side-chains (time fraction of solvent exposure < 75%) are indeed protected from the reaction. However, a large number of both reactive and non-reactive cases are seen in the solvent accessible region (time fraction of solvent exposure > 75%). The overall accuracy in predicting the reactivity using SASA as a single descriptor is 39%, which is no better than the performance of the near attack conformation metric.

Combined Model

As discussed above, deamination and isomerization reactions can be described in the three consecutive steps: deprotonation, cyclization and hydrolysis. Utilizing a multiscale computational approach, we were able to test the propensity of degradation in the adimab dataset using these physics-based metrics. Since all three steps are essential for the reaction, we sought to check if a combination of the three metrics would improve the classification accuracy: D1) deprotonation – as represented by the free-energy profile of ϕ , ψ angles of the n+1 residue recovered from the MD simulation, D2) cyclization – as represented by the free-energy profile of χ , ψ angles also recovered from the MD simulation, and D3) hydrolysis – as represented by the solvent accessibility calculated as the fraction of solvent occupied state in MD trajectory. The decision tree that we used to classify if an ASN/ASP residue is likely to degrade is shown in Figure 7. The adimab dataset discussed herein is disproportionately distributed toward true negatives (non-reactive) vs true positives (reactive), so this presented a challenge to interpreting common statistical measures such as accuracy. We instead chose to use balanced accuracy, Matthew's correlation coefficient (MCC) and Cohen's Kappa score (CKS) as the statistic to compare the three individual descriptors as well as combinations. We also created a naïve predictor that would predict all the residues as non-reactive; this naïve predictor will score 0 on both the MCC and CSK statistic. As can be seen from Table 1, the physics-based classifier model based on the three descriptors offers an overall balanced accuracy of 83% in predicting the isomerization and deamidation sites. All three descriptors have a better MCC and CKS score compared to the naïve predictor. Similarly, any combination of two descriptors performed moderately better than a single predictor. The combined model using all three descriptors is better at predicting non-reactive

residues over reactive residues, which could be useful in a therapeutic manufacturing setting where the cost of predicting a reactive residue as non-reactive is higher. This combined mechanistic model acts as a good screening criterion in identifying antibodies that are prone to chemical degradation via deamidation or isomerization.

Conclusion

Spontaneous chemical degradation at the ASN/ASP residues is an important problem that affects the developability of the protein therapeutics. Typically, the issues of chemical degradation are encountered at later stages in the development at which point iterative protein-engineering efforts are employed to mitigate these modifications. Such iterative efforts prolong development timelines and delays lifesaving therapeutics reaching patients. Multiple computational strategies ranging from QM calculation to machine learning models have been proposed to predict degradation in ASP/ASN. Very few such tools (if any) generalize for high-throughput screening. We leveraged prior theoretical and experimental findings on the ASN/ASP degradation reactions to carry out multiscale analysis to comprehensively evaluate the chemical degradation in 1039 residues (498 ASP and 541 ASN) spanning 131 therapeutic mAbs. In doing so we have made some interesting findings: 1) sidechain steric interactions of the n+1 residue affected the proton affinity of backbone amide in the right-handed α -helix conformation, 2) NG and DG sites as well as the reactive N/DX sites are preferentially observed in the left-handed α -helix conformation that correlates with low proton affinity, and 3) there are specific regions of the sidechain conformational space are both energetically favorable and in a near-attack conformation. Utilizing an experimentally validated truth set we have been able to connect molecular intuition from the reaction mechanism to a real-life scenario. Utilizing just three descriptors that are fundamental to the protein structure one can heuristically classify the likelihood of the protein to undergo chemical degradation at the ASN/ASP sites. Although we have exclusively looked at mAbs in this study, we believe that the structural and chemical insights described herein are broadly applicable and generalizable to all proteins and protein-based therapeutics.

Methods

Experimental dataset

The validation data contained 498 ASP and 541 ASN residues in the CDR loops of 131 therapeutic mAbs published by Lu et al.[34] The isomerization events were analyzed at low pH stress (pH 5.5) for 2 weeks at 40 °C. A total of 31 isomerization sites were identified with total modification measurements ranging from 2.1 to 44.0%. The deamidation data was obtained at high pH stress (pH 8.5) for 1 week at 40 °C. A total of 39 deamidation sites were reported with degradation levels ranging from 2.0% to 67.7%. We used a threshold cutoff of 5.0% for both isomerization and deamidation events to classify reactive versus non-reactive.

QM calculations

Semi-empirical QM (SQM) calculations to calculate proton affinity were performed using GAUSSIAN09[50] and MOPAC2016[51] software packages. To rule out any uncertainties in the proton affinity calculations, we carried out DFT calculation of proton affinity for N-formyl-glycinamide at M062X/6-311++G(d,p) level of theory in both gas phase and aqueous phase continuum SMD solvent environment.[52] The gas phase calculations were then repeated in MOPAC with PM6-D3H4,[53] PM7 and RM1 semi-empirical methods (SI).[52] The regions of proton affinity qualitatively matched across DFT and SQM levels of theory albeit with variations in the energy values as expected due to variations in the theory and the Hamiltonian. Throughout the manuscript we tried to make only qualitative inferences from the QM calculations and we refrained from interpreting the exact energy as that is beyond the scope of the current study.

Dipeptides of the Ace-NX-Nme and Ace-DX-Nme as well as their corresponding deprotonated anions (where X can be single letter code of all 20 amino acids) were generated using openable.[54] The generated structures were subject to geometry optimization under PM6 level of theory to arrive at an initial structure. A 2D PES scan at PM6-D3H4, PM7 and RM1 levels of theory. The scan was performed at 10° intervals spanning the entire Ramachandran region. A total of 1369 data points were scanned for each dipeptide and its corresponding anion. At each step geometry optimization was requested to adhere to TIGHT convergence criterion while the angles being scanned would be kept constant. The resulting PES were used to arrive at the proton affinity expressed in the heat of formation for the forward reaction as

described in Figure 2. The discrete energy points were interpolated and plotted using linear spline interpolation as implemented in the matplotlib[55] python package.

PDB Library

We downloaded the pdb file repository of all the antibody (and antibody like Fab, scFV, ect...) that have crystal structure deposited in the PDB from SAbDab (The Structural Antibody Database <http://opig.stats.ox.ac.uk/webapps/newsabdab/sabdab/>).[56, 57] We arrived at 3493 structures which were then cleaned, energy minimized and isolated from any other ligands. Using these structures, we calculated backbone dihedral angles of all Nx and Dx residues from the Fab domain. All the PDB handling was performed using MDanalysis[58] and PyTraj[59] python tools. The backbone dihedral angle distribution from this data was used for calculating 2D kernel density estimation using the SciKit-Learn python package.[60]

KDE calculation

In order to compare the KDEs that were calculated from the PDB database and the backbone dihedral angles observed in the MD simulations, we first classified each N/Dx site as reactive or non-reactive based on the deprotonation metric. We collected the ϕ & ψ angle distribution for all these sites across all MD trajectories. This data was used to calculate the second kernel density estimation plot in Figure 4, using the same SciKit-Learn package as before.

Molecular Dynamics Simulation

Molecular Operating Environment (MOE)[61] software from Chemical Computing Group was used to homology model each of the fragment antigen-binding (Fab) domain of all the antibodies accessed in the adimab set. The Fab structures were energy minimized to remove any steric clashes. The energy minimization was performed using SANDER program in Amber2015.[62] The protonation states at pH 5.5 and 8.5 were assigned using the PDB2PQR tool. The two pH values mimic the buffer conditions in which the reactivity was estimated in the adimab experiment.

Energy minimized Fab structures were then prepared for equilibrium MD simulations by solvating the structure in a cubic box of TIP3P explicit water (model).[63] It was ensured such that the Fab was positioned at a minimum 10Å distance from the edge of the box. The system was then adjusted to have a net neutral charge by adding Na⁺ and Cl⁻ counter ions. Hydrogen Mass Repartitioning was performed on the solute atoms to enable long simulation time steps of 4fs.[64]

The GPU implementation of Amber 2015 MD software package with the SPFP precision model was used for the MD simulation. The simulations proceeded by relaxing the system with 2000 steps of conjugate-gradient energy minimization. Harmonic restraining potentials with the force constant of 10 kcal/mol/Å² were imposed to restrain the solute to its initial structure. Post which the system was brought to an NPT ensemble with the pressure maintained at 1 atm and the thermostat set to 300K over the course of 200 ps, while the protein was restrained using a harmonic positional restraint of 10 kcal/mol/Å². The system was then equilibrated for 1 ns with a restraint force constant of 1 kcal/mol/Å² on the protein structure maintaining the pressure and temperature at the same levels. The system was then subject to equilibrium MD simulation at NPT ensemble without any restraints at a time step of 4 fs. Long range interactions were cut off at 9Å and restricted range limited interactions. Coulombic interaction was calculated using the Particle Mesh Ewald algorithm for long-range electrostatics. The thermostat was maintained at 300K with Langevin dynamics and collision frequency set to 1 ps⁻¹. The simulation was carried out for 500 ns during which the SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms. Two other replicates of the 500 ns simulation were performed instantiated with a different random number to bring the combined sampling time to 1.5 μs. Trajectory snapshots were saved at 10 ps intervals and used for downstream analysis. The combined trajectories from three simulations were used for the analysis.

CPPTRAJ software in AmberTools was used to analyze the trajectories. The free energy surface (FES) in the space of dihedral angles were calculated from bin populations using $G_i = -k_B T \ln(N_i/N_{max})$, where k_B is Boltzmann's constant, T is the temperature, N_i is the population of bin i and N_{max} is the population of the most populated bin. Bins with no population were given an artificial barrier equivalent to a population of 0.5. Solvent accessible surface area (SASA) was calculated using the MSMS package[65] using a probe radius of 1.4 and triangle density of 3.

Advanced Sampling Using Metadynamics

Metadynamics simulations were carried out using the GROMACS2018[66] MD engine patched to interface with PLUMED.[67] Metadynamics was performed in the well-tempered[68] scheme along the conformational space of ψ & χ angles of the pentapeptide Ace-GGNAG-Nme. The simulations utilized the same forcefield and MD parameters with 2 fs time steps. Bias was added at every 500 steps with a sigma of 0.1 and bias of 1.0 kcal/mol. The simulation was carried out for 1 μ s and the trajectory was analyzed after unbiasing the simulation using the standard reweighting[69] procedure implemented in PLUMED. The free energy surface and corresponding distance distribution was plotted using matplotlib.

Acknowledgements

The authors would like to thank Robert Kelly, Vikas Sharma and Sreedhara Alavattam for their engaging discussions. We would like to thank Benjamin Sellers for his valuable comments and suggestions on the manuscript.

Competing interest

FJI and SD are co-inventors on the provisional patent application describing the computational method discussed in the manuscript

Supporting Information

The supporting information provides the proton affinity calculations and the corresponding KDE plots for other dipeptides.

Contact

All correspondence and data requests should be directed to the corresponding author.

Figure 1. Deamidation (1a) and Isomerization (1b) reaction mechanisms showing deprotonation (1) of the n+1 amide leading to the formation of succinimide intermediate (2) and hydrolysis of succinimide (3) into aspartic acid (3a) or iso-aspartic acid (3b).

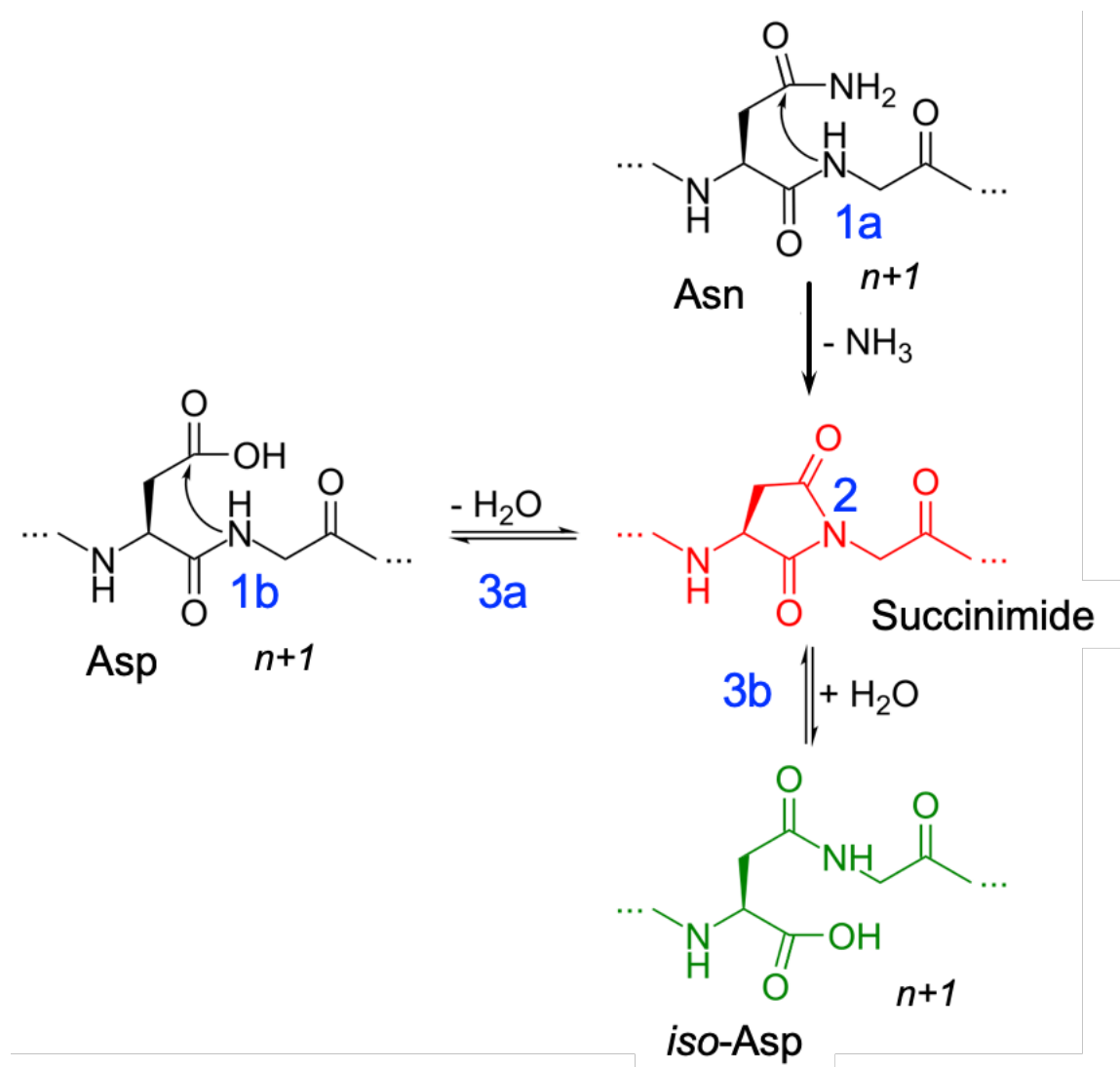


Figure 2. Backbone conformation dependence for proton affinity. (a) the deprotonation reaction and the calculation of proton affinity as the heat of formation of the forward reaction, (b) the proton affinity distribution of the small molecule N-formyl-glycinamide as a function of the backbone dihedral angles, and (c) the most acidic and basic conformations based on the proton affinity distribution.

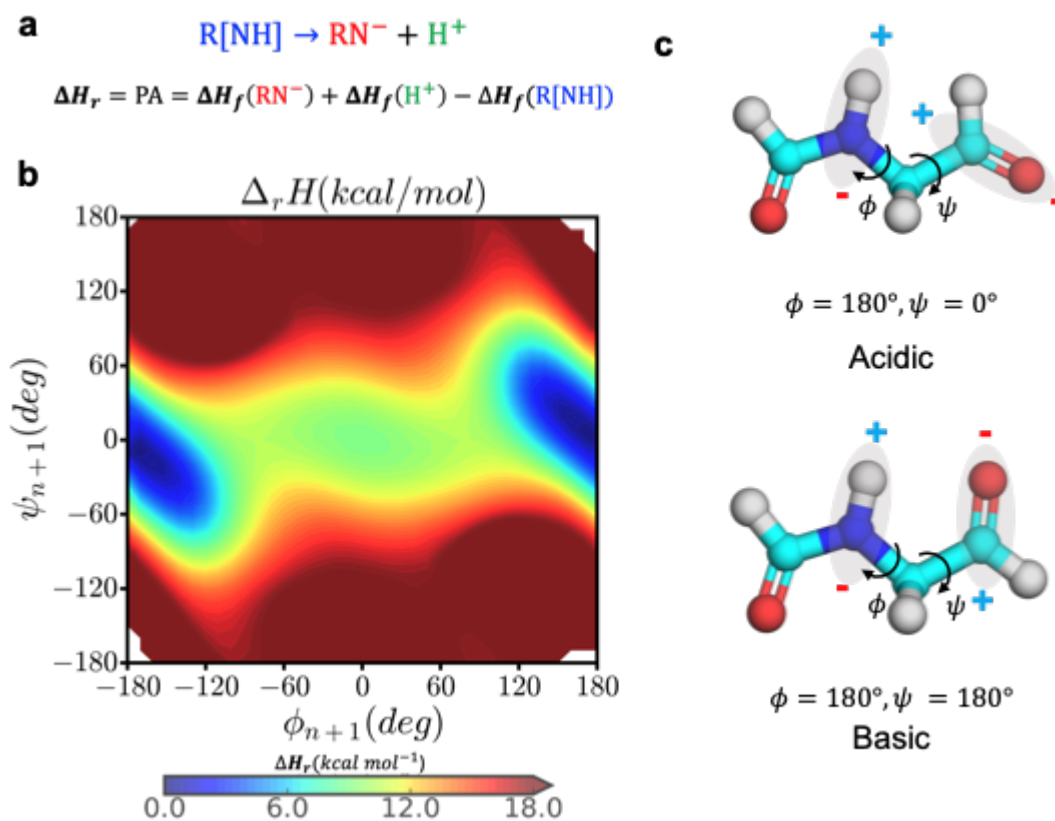


Figure 3. Conformational determinants of proton affinity in NG and NA dipeptides. a) the gas-phase proton-affinity calculated with the semi-empirical RM1 Hamiltonian as a function of ϕ & ψ angles. The colors scale from 0→18+ kcal/mol from blue to red. b) the corresponding kernel density estimation

plot of the respective dipeptides observed in the PDB database of crystallized antibodies. The color gradient scales from low→high from yellow to

purple. c) representative basic and acidic conformations beta sheet, right-handed helix and left-handed helix.

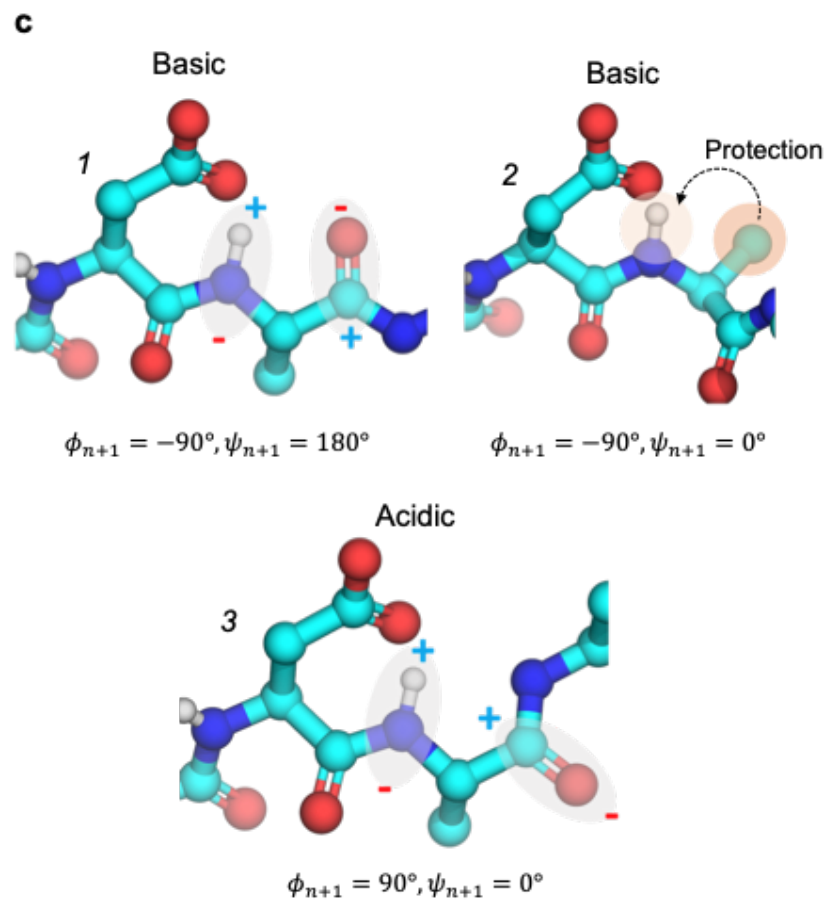
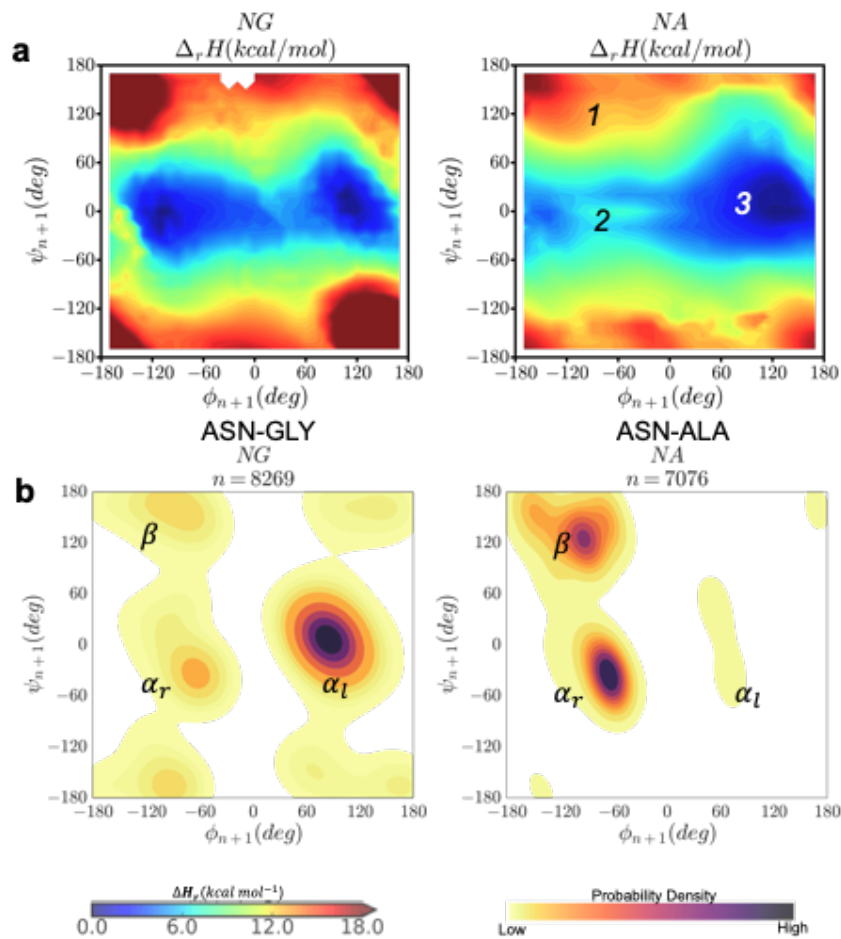


Figure 4. Kernel density estimation plots for the backbone conformation of all 1039 $n+1$ residues that neighbor an ASP (top) or ASN (bottom) in the adimab dataset. The ϕ & ψ angle distribution was collected from μ s long equilibrium MD simulations and classified as reactive (left) and non-reactive (right) based on the report from the adimab set. A clear pattern emerges around the left-handed α -helix conformation in reactive cases that is distinct from the non-reactive cases on the right. The right most panel shows the Ramachandran plot boundaries labelled with the α -helix and β -sheet conformations.

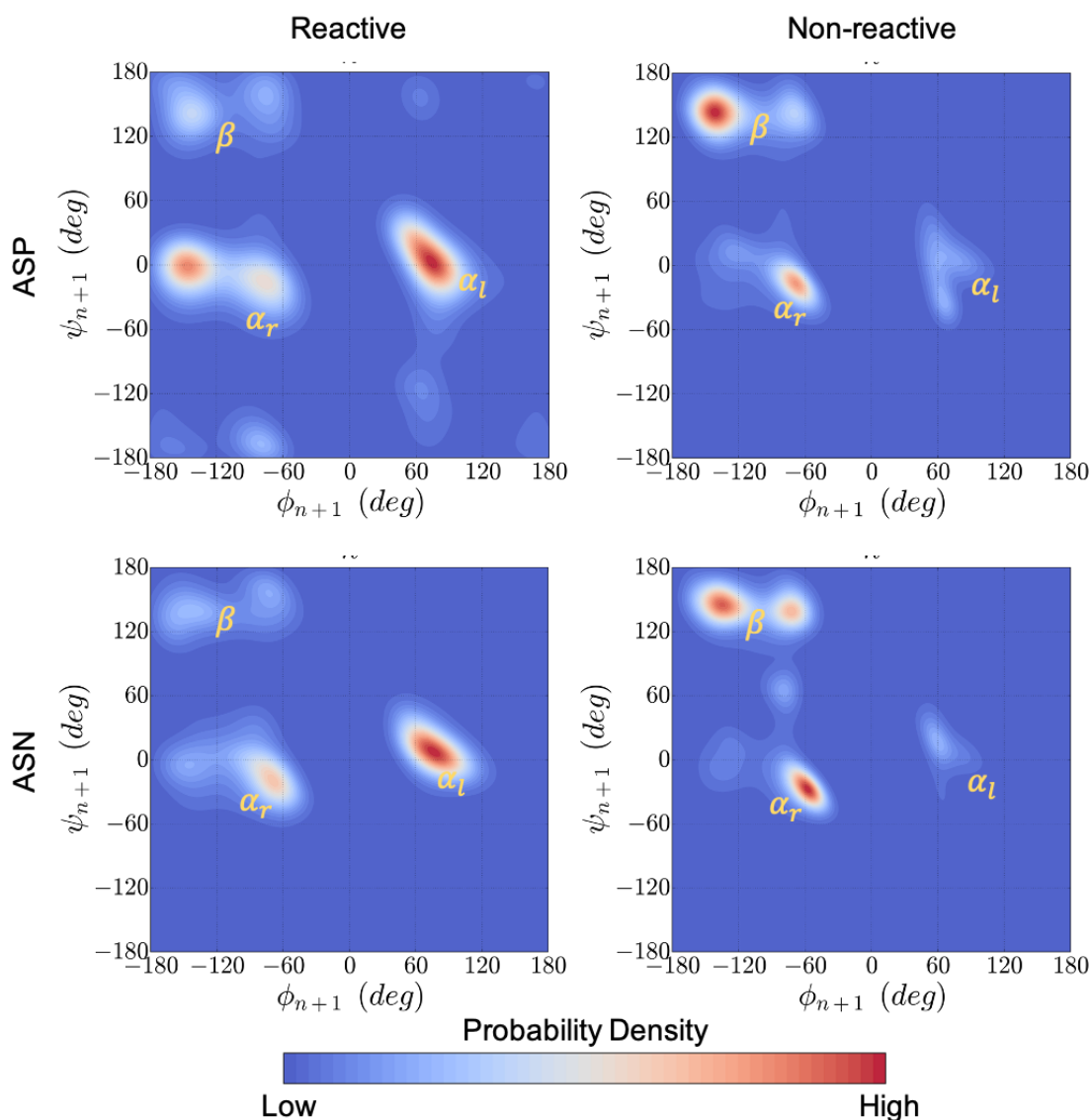


Figure 5. Conformational space for the near-attack conformation that enables succinimide formation. a) The structural orientation of the reactive (1a and 2a) and non-reactive sidechain conformations (2b and 3) of an aspartic acid residue is shown in ball and stick representation. b) Free-energy surface of the sidechain conformation in ψ & χ dihedral angle space and c) the corresponding distance between C_γ and the $n+1$ amide nitrogen calculated from metadynamics simulation of Ace-GGNAG-Nme pentapeptide.

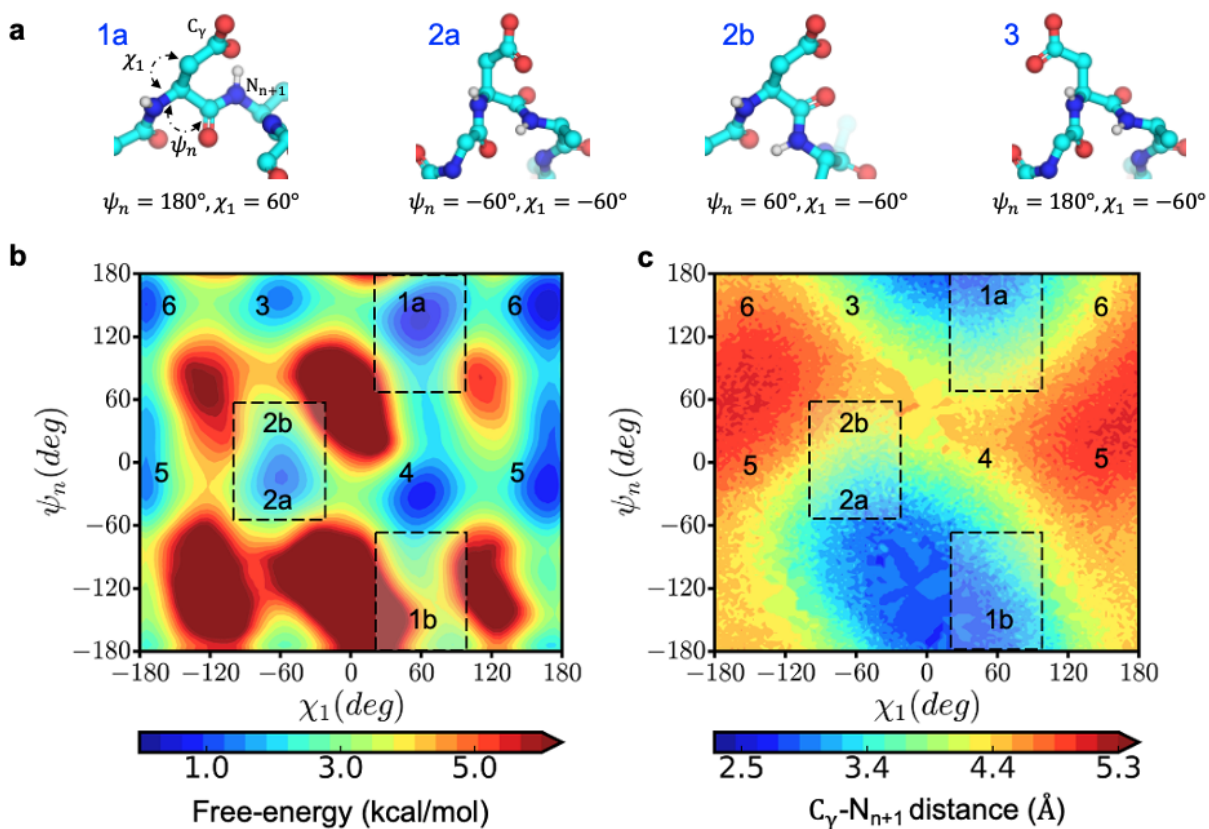


Figure 6. Performance of individual descriptors in discriminating degraded ASP (top) and ASN (bottom) residues from the adimab dataset. The percentage degradation values reported in the adimab study are represented on the X-axis. The Y-axis in the leftmost panel (D1) is the free-energy barrier at the deprotonation prone backbone helical conformation of the n+1 residue. The Y-axis in the middle panel (D2) is the free-energy barrier at the near-attack conformation defined by the ψ & χ angles. The Y-axis in the rightmost panel (D3) is the time fraction when the ASP/ASN residue remains solvent accessible. The red cut-off markers indicate the values that determine if a residue is prone to degradation or not. The blue shaded regions indicate the false negative predictions (experimentally observed to degrade but classified as non-reactive) and the grey shaded region indicates the false positives (predicted reactive but not evidenced experimentally)

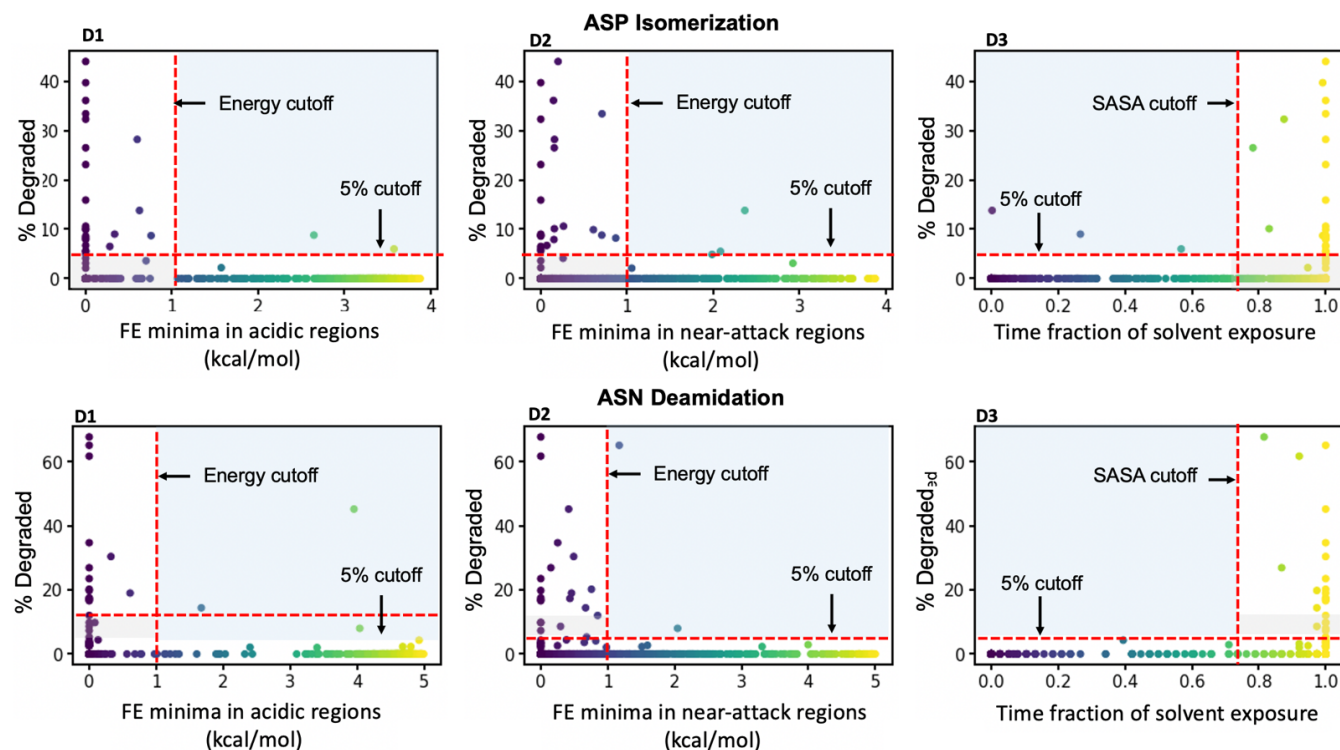


Figure 7. Flowchart showing the decision scheme used to determine if a given ASN/ASP residue is likely to degrade in folded proteins.

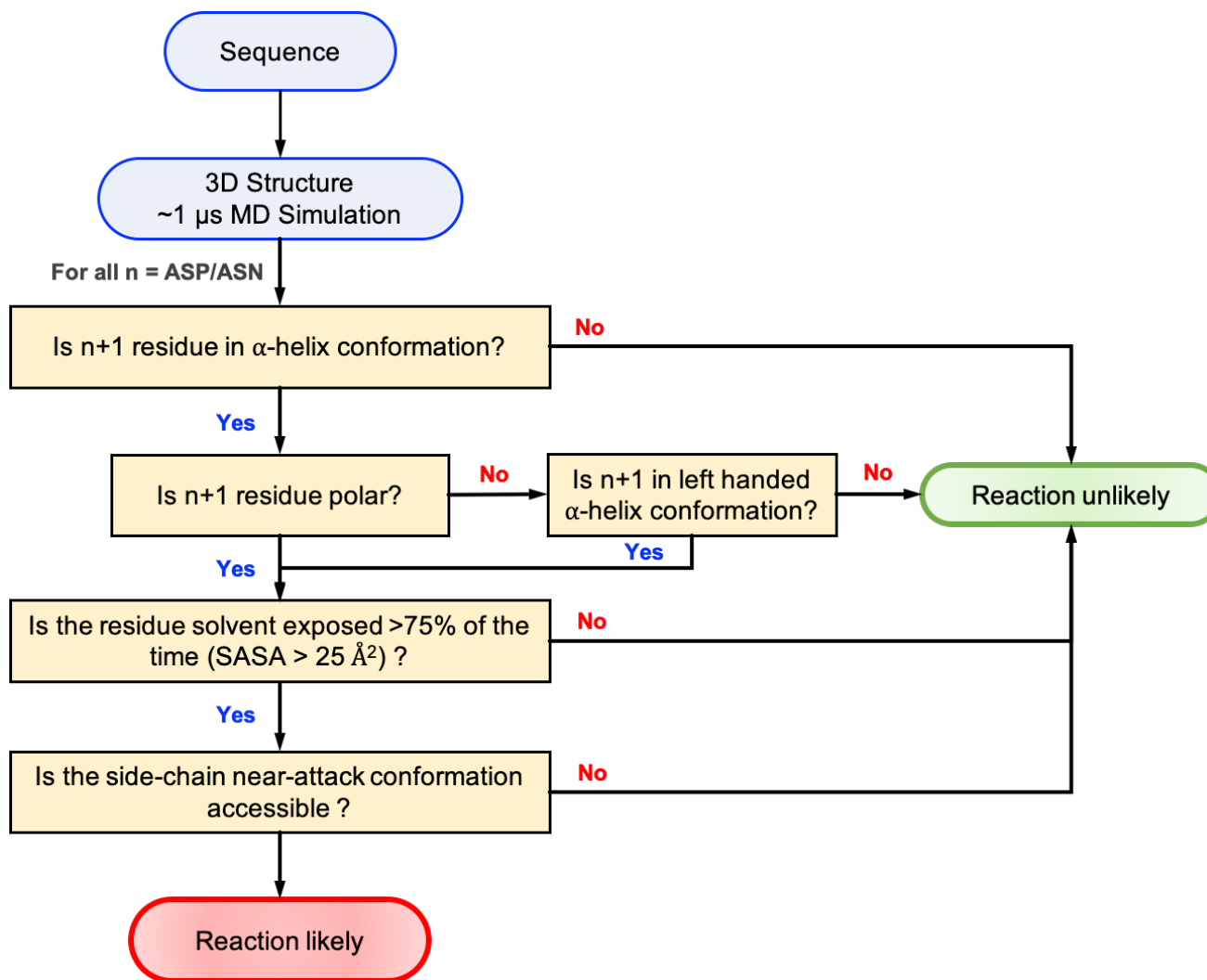


Table 1. Performance statistic for the physics-based classifier model when different descriptors are used: D1- backbone dihedral conformation of the n+1 residue, D2- side-chain dihedral conformation of ASP/ASN residue, D3- fraction of time the ASN/ASP residue remains solvent accessible. Combinations are represented by DXX where the X represents the descriptors. The true distribution of the adimab set is 47 TP and 992 TN.

<i>Type</i>	<i>TP*</i>	<i>FN*</i>	<i>FP*</i>	<i>TN*</i>	<i>Accuracy</i>	<i>Accuracy[†]</i>	<i>Precision</i>	<i>Recall</i>	<i>F₁ Score</i>	<i>F_{0.5} Score</i>	<i>MCC[‡]</i>	<i>CKS[§]</i>
ASN-D1	20	3	173	345	0.67	0.77	0.95	0.67	0.77	0.87	0.23	0.12
ASN-D2	21	2	301	217	0.44	0.67	0.95	0.44	0.57	0.75	0.14	0.05
ASN-D3	23	0	316	202	0.42	0.69	0.96	0.42	0.54	0.73	0.16	0.05
ASN-D12	19	4	130	388	0.75	0.79	0.95	0.75	0.83	0.90	0.26	0.16
ASN-D13	20	3	92	426	0.82	0.85	0.96	0.82	0.87	0.92	0.34	0.24
ASN-D23	21	2	183	335	0.66	0.78	0.96	0.66	0.76	0.86	0.23	0.12
ASN-naïve	0	23	0	518	0.96	0.50	0.92	0.96	0.94	0.92	0.00	0.00
ASN-D123	19	4	62	456	0.88	0.85	0.96	0.88	0.91	0.94	0.40	0.32
ASP-D1	22	2	186	288	0.62	0.76	0.95	0.62	0.73	0.84	0.23	0.11
ASP-D2	22	2	318	156	0.36	0.62	0.94	0.36	0.48	0.68	0.11	0.03
ASP-D3	21	3	312	162	0.37	0.61	0.94	0.37	0.49	0.68	0.10	0.03
ASP-D12	20	4	118	356	0.76	0.79	0.95	0.76	0.82	0.89	0.28	0.18
ASP-D13	20	4	133	341	0.72	0.78	0.95	0.72	0.80	0.88	0.26	0.16
ASP-D23	20	4	205	269	0.58	0.70	0.94	0.58	0.69	0.82	0.17	0.08
ASP-naïve	0	24	0	474	0.95	0.50	0.91	0.95	0.93	0.91	0.00	0.00
ASP-D123	19	5	77	397	0.84	0.81	0.95	0.84	0.88	0.92	0.34	0.26
ALL-D1	42	5	359	633	0.65	0.77	0.95	0.65	0.75	0.86	0.23	0.12
ALL-D2	43	4	619	373	0.40	0.65	0.95	0.40	0.53	0.72	0.13	0.04
ALL-D3	44	3	628	364	0.39	0.65	0.95	0.39	0.52	0.71	0.13	0.04
ALL-D12	39	8	248	744	0.75	0.79	0.95	0.75	0.83	0.90	0.27	0.17
ALL-D13	40	7	225	767	0.78	0.81	0.95	0.78	0.84	0.90	0.30	0.19
ALL-D23	41	6	388	604	0.62	0.74	0.95	0.62	0.73	0.85	0.20	0.10
ALL-naïve	0	47	0	992	0.95	0.50	0.91	0.95	0.93	0.92	0.00	0.00
ALL-D123	38	9	139	853	0.86	0.83	0.95	0.86	0.89	0.93	0.37	0.29

* TP – True positive, FN – False Negative, FP – False positive and TN – True Negative

† Balanced accuracy calculated with weighted distribution of positives and negatives.

‡ Matthew's Correlation Coefficient ranges from -1.0 to +1.0, where the negative values indicate a bad predictor.

§ Cohen Kappa Score ranges from -1.0 to +1.0, similar to the MCC statistic but a better indicator to compare an imbalanced dataset.

References

1. Jain, T., et al., *Biophysical properties of the clinical-stage antibody landscape*. Proc Natl Acad Sci U S A, 2017. **114**(5): p. 944-949.
2. Raybould, M.I.J., et al., *Five computational developability guidelines for therapeutic antibody profiling*. Proc Natl Acad Sci U S A, 2019. **116**(10): p. 4025-4030.
3. Xu, Y., et al., *Structure, heterogeneity and developability assessment of therapeutic antibodies*. MAbs, 2019. **11**(2): p. 239-264.
4. Robinson, N.E. and A.B. Robinson, *Deamidation of human proteins*. Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12409-13.
5. Walsh, G. and R. Jefferis, *Post-translational modifications in the context of therapeutic proteins*. Nat Biotechnol, 2006. **24**(10): p. 1241-52.
6. Klimova, N., A. Long, and T. Kristian, *Significance of Mitochondrial Protein Post-translational Modifications in Pathophysiology of Brain Injury*. Transl Stroke Res, 2018. **9**(3): p. 223-237.
7. Giles, A.R., et al., *Deamidation of Amino Acids on the Surface of Adeno-Associated Virus Capsids Leads to Charge Heterogeneity and Altered Vector Function*. Mol Ther, 2018. **26**(12): p. 2848-2862.
8. Jiskoot, W., et al., *Protein instability and immunogenicity: roadblocks to clinical application of injectable protein delivery systems for sustained release*. J Pharm Sci, 2012. **101**(3): p. 946-54.
9. Krause, M.E. and E. Sahin, *Chemical and physical instabilities in manufacturing and storage of therapeutic proteins*. Curr Opin Biotechnol, 2019. **60**: p. 159-167.
10. Strader, M.B., et al., *Post-translational modification as a response to cellular stress induced by hemoglobin oxidation in sickle cell disease*. Sci Rep, 2020. **10**(1): p. 14218.
11. Zhu, M.M., et al., *Industrial Production of Therapeutic Proteins: Cell Lines, Cell Culture, and Purification*, in *Handbook of Industrial Chemistry and Biotechnology*, J.A. Kent, T.V. Bommaraju, and S.D. Barnicki, Editors. 2017, Springer International Publishing: Cham. p. 1639-1669.
12. Harris, R.J., et al., *Identification of multiple sources of charge heterogeneity in a recombinant antibody*. J Chromatogr B Biomed Sci Appl, 2001. **752**(2): p. 233-45.
13. Reissner, K.J. and D.W. Aswad, *Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals?* Cell Mol Life Sci, 2003. **60**(7): p. 1281-95.
14. Patel, C.N., et al., *N+1 Engineering of an Aspartate Isomerization Hotspot in the Complementarity-Determining Region of a Monoclonal Antibody*. J Pharm Sci, 2016. **105**(2): p. 512-518.
15. Salazar-Fontana, L.I., et al., *Approaches to Mitigate the Unwanted Immunogenicity of Therapeutic Proteins during Drug Development*. AAPS J, 2017. **19**(2): p. 377-385.
16. Verma, A. and D.L. Burns, *Improving the stability of recombinant anthrax protective antigen vaccine*. Vaccine, 2018. **36**(43): p. 6379-6382.
17. Jia, L. and Y. Sun, *Protein asparagine deamidation prediction based on structures with machine learning methods*. PLoS One, 2017. **12**(7): p. e0181347.

18. Lorenzo, J.R., L.G. Alonso, and I.E. Sanchez, *Prediction of Spontaneous Protein Deamidation from Sequence-Derived Secondary Structure and Intrinsic Disorder*. PLoS One, 2015. **10**(12): p. e0145186.
19. Robinson, N.E. and A.B. Robinson, *Prediction of protein deamidation rates from primary and three-dimensional structure*. Proc Natl Acad Sci U S A, 2001. **98**(8): p. 4367-72.
20. Sydow, J.F., et al., *Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions*. PLoS One, 2014. **9**(6): p. e100736.
21. Robinson, A.B. and C.J. Rudd, *Deamidation of Glutaminyl and Asparaginyl Residues in Peptides and Proteins*, in *Current Topics in Cellular Regulation*, B.L. Horecker and E.R. Stadtman, Editors. 1974, Academic Press. p. 247-295.
22. Capasso, S., et al., *Deamidation via cyclic imide in asparaginyl peptides*. Pept Res, 1989. **2**(2): p. 195-200.
23. Patel, K. and R.T. Borchardt, *Chemical pathways of peptide degradation. II. Kinetics of deamidation of an asparaginyl residue in a model hexapeptide*. Pharm Res, 1990. **7**(7): p. 703-11.
24. Tyler-Cross, R. and V. Schirch, *Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides*. J Biol Chem, 1991. **266**(33): p. 22549-56.
25. Capasso, S., et al., *Kinetics and mechanism of succinimide ring formation in the deamidation process of asparagine residues*. Journal of the Chemical Society, Perkin Transactions 2, 1993(4): p. 679-682.
26. Radkiewicz, J.L., et al., *Accelerated Racemization of Aspartic Acid and Asparagine Residues via Succinimide Intermediates: An ab Initio Theoretical Exploration of Mechanism*. Journal of the American Chemical Society, 1996. **118**(38): p. 9148-9155.
27. Radkiewicz, J.L., et al., *Neighboring side chain effects on asparaginyl and aspartyl degradation: an ab initio study of the relationship between peptide conformation and backbone NH acidity*. J Am Chem Soc, 2001. **123**(15): p. 3499-506.
28. Ritz-Timme, S., I. Laumeier, and M.J. Collins, *Aspartic acid racemization: evidence for marked longevity of elastin in human skin*. Br J Dermatol, 2003. **149**(5): p. 951-9.
29. Catak, S., et al., *Reaction mechanism of deamidation of asparaginyl residues in peptides: effect of solvent molecules*. J Phys Chem A, 2006. **110**(27): p. 8354-65.
30. Catak, S., et al., *Deamidation of asparagine residues: direct hydrolysis versus succinimide-mediated deamidation mechanisms*. J Phys Chem A, 2009. **113**(6): p. 1111-20.
31. Fogolari, F., et al., *pKaShift Effects on Backbone Amide Base-Catalyzed Hydrogen Exchange Rates in Peptides*. Journal of the American Chemical Society, 1998. **120**(15): p. 3735-3738.
32. Gamage, C.L.D., T.S. Hageman, and D.D. Weis, *Rapid Prediction of Deamidation Rates of Proteins to Assess Their Long-Term Stability Using Hydrogen Exchange-Mass Spectrometry*. J Pharm Sci, 2019. **108**(6): p. 1964-1972.
33. Robinson, N.E., et al., *Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides*. J Pept Res, 2004. **63**(5): p. 426-36.
34. Lu, X., et al., *Deamidation and isomerization liability analysis of 131 clinical-stage antibodies*. MAbs, 2019. **11**(1): p. 45-57.

35. Oliva, A., B. Henry, and M.F. Ruiz-López, *Insights on peptide backbone N–H acidity: Structure of anions, hydration effects*. Chemical Physics Letters, 2013. **561-562**: p. 153-158.
36. Plotnikov, N.V., et al., *Quantifying the Risks of Asparagine Deamidation and Aspartate Isomerization in Biopharmaceuticals by Computing Reaction Free-Energy Surfaces*. J Phys Chem B, 2017. **121**(4): p. 719-730.
37. Ugur, I., et al., *Why does Asn71 deamidate faster than Asn15 in the enzyme triosephosphate isomerase? Answers from microsecond molecular dynamics simulation and QM/MM free energy calculations*. Biochemistry, 2015. **54**(6): p. 1429-39.
38. Delmar, J.A., et al., *Machine Learning Enables Accurate Prediction of Asparagine Deamidation Probability and Rate*. Mol Ther Methods Clin Dev, 2019. **15**: p. 264-274.
39. Hamuro, Y. and S.Y. E, *Determination of Backbone Amide Hydrogen Exchange Rates of Cytochrome c Using Partially Scrambled Electron Transfer Dissociation Data*. J Am Soc Mass Spectrom, 2018. **29**(5): p. 989-1001.
40. Huang, L., P.K. So, and Z.P. Yao, *Protein dynamics revealed by hydrogen/deuterium exchange mass spectrometry: Correlation between experiments and simulation*. Rapid Commun Mass Spectrom, 2019. **33 Suppl 3**(S3): p. 83-89.
41. Kosky, A.A., et al., *The effects of alpha-helix on the stability of Asn residues: deamidation rates in peptides of varying helicity*. Protein Sci, 1999. **8**(11): p. 2519-23.
42. Forood, B., E.J. Feliciano, and K.P. Nambiar, *Stabilization of alpha-helical structures in short peptides via end capping*. Proc Natl Acad Sci U S A, 1993. **90**(3): p. 838-42.
43. Head-Gordon, T., et al., *A theoretical study of alanine dipeptide and analogs*. International Journal of Quantum Chemistry, 2009. **36**(S16): p. 311-322.
44. Diepold, K., et al., *Simultaneous assessment of Asp isomerization and Asn deamidation in recombinant antibodies by LC-MS following incubation at elevated temperatures*. PLoS One, 2012. **7**(1): p. e30295.
45. Wakankar, A.A., et al., *Aspartate isomerization in the complementarity-determining regions of two closely related monoclonal antibodies*. Biochemistry, 2007. **46**(6): p. 1534-44.
46. Yan, B., et al., *Succinimide formation at Asn 55 in the complementarity determining region of a recombinant monoclonal antibody IgG1 heavy chain*. J Pharm Sci, 2009. **98**(10): p. 3509-21.
47. Pace, A.L., et al., *Asparagine deamidation dependence on buffer type, pH, and temperature*. J Pharm Sci, 2013. **102**(6): p. 1712-1723.
48. Sreedhara, A., et al., *Characterization of the isomerization products of aspartate residues at two different sites in a monoclonal antibody*. Pharm Res, 2012. **29**(1): p. 187-97.
49. Rehder, D.S., et al., *Isomerization of a single aspartyl residue of anti-epidermal growth factor receptor immunoglobulin gamma2 antibody highlights the role avidity plays in antibody activity*. Biochemistry, 2008. **47**(8): p. 2518-30.
50. Frisch, M., et al., GAUSSIAN09. Gaussian Inc., Wallingford, CT, USA. 2009.
51. Stewart, J., MOPAC2016; Stewart Computational Chemistry: Colorado Springs, CO, 2016. Google Scholar There is no corresponding record for this reference.

52. Hostaš, J., J. Řezáč, and P. Hobza, *On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions*. Chemical Physics Letters, 2013. **568**: p. 161-166.
53. Kromann, J.C., et al., *A third-generation dispersion and third-generation hydrogen bonding corrected PM6 method: PM6-D3H+*. PeerJ, 2014. **2**: p. e449.
54. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox*. J Cheminform, 2011. **3**(1): p. 33.
55. Ari, N. and M. Ustazhanov. *Matplotlib in python*. in *2014 11th International Conference on Electronics, Computer and Computation (ICECCO)*. 2014. IEEE.
56. Dunbar, J., et al., *SAbDab: the structural antibody database*. Nucleic Acids Res, 2014. **42**(Database issue): p. D1140-6.
57. Raybould, M.I.J., et al., *Thera-SAbDab: the Therapeutic Structural Antibody Database*. Nucleic Acids Res, 2020. **48**(D1): p. D383-D388.
58. Gowers, R.J., et al., *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations*. 2019, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
59. Nguyen, H., et al., *PYTRAJ: Interactive data analysis for molecular dynamics simulations*. New Brunswick, NJ: Rutgers University, 2016.
60. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python*. The Journal of machine Learning research, 2011. **12**: p. 2825-2830.
61. Vilar, S., G. Cozza, and S. Moro, *Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery*. Curr Top Med Chem, 2008. **8**(18): p. 1555-72.
62. Salomon-Ferrer, R., et al., *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald*. J Chem Theory Comput, 2013. **9**(9): p. 3878-88.
63. Price, D.J. and C.L. Brooks, 3rd, *A modified TIP3P water potential for simulation with Ewald summation*. J Chem Phys, 2004. **121**(20): p. 10096-103.
64. Hopkins, C.W., et al., *Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning*. J Chem Theory Comput, 2015. **11**(4): p. 1864-74.
65. Sanner, M.F., A.J. Olson, and J.C. Spohner, *Reduced surface: an efficient way to compute molecular surfaces*. Biopolymers, 1996. **38**(3): p. 305-20.
66. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1**: p. 19-25.
67. Tribello, G.A., et al., *PLUMED 2: New feathers for an old bird*. Computer Physics Communications, 2014. **185**(2): p. 604-613.
68. Barducci, A., G. Bussi, and M. Parrinello, *Well-tempered metadynamics: a smoothly converging and tunable free-energy method*. Phys Rev Lett, 2008. **100**(2): p. 020603.
69. Bonomi, M., A. Barducci, and M. Parrinello, *Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics*. J Comput Chem, 2009. **30**(11): p. 1615-21.