TASK Quarterly 28 (4) 2024

Assessing AMBER and UNRES Force Fields on the Energy Landscape of Bovine Pancreatic Trypsin Inhibitor

Songsong Zhou[1], **Patryk A. Wesołowski**[1,*], **Yifei Wang**[1], **Vilmos Neuman**[2], **Krzysztof K. Bojarski**[1,3,4] and **David J. Wales**[1,†]

[1]Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, U.K.
[2]Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, U.K.
[3]Department of Physical Chemistry, Gdansk University of Technology, Narutowicza 11/12, Gdansk, 80-233, Poland
[4]Center for Functional Protein Assemblies, Technical University of Munich, Ernst-Otto-Fischer-Straße 8, Garching, Germany

[*]paw61@cam.ac.uk [†]dw34@cam.ac.uk

https://doi.org/10.34808/tq2024/28.4/b

Abstract

Biological complexity emerges from the collective behaviour of proteins, whose folding and conformational dynamics are directly governed by the underlying potential energy surface. Here, we compare the potential energy landscapes of bovine pancreatic trypsin inhibitor (BPTI) obtained with the all-atom AMBER force field and the coarse-grained UNRES potential. For the native triply disulphide-bonded state of BPTI, both models sample predominantly folded, native-like conformations with relatively small structural deviations from the crystallographic structure, though UNRES explores structures with a broader range of RMSDs, as well as radii of gyration and solvent-accessible surface areas. For comparable CPU time, the discrete path sampling approach using UNRES locates a larger number of minima and transition states, providing a more finely resolved and extensively sampled energy landscape. Together, these results highlight how all-atom and coarse-grained potentials differ in the representation of protein energy landscapes, while retaining consistent global features for a highly constrained, disulphide-rich protein such as BPTI.

Keywords:

Potential energy landscapes, AMBER force field, UNRES force field, BPTI protein

1. Introduction

BPTI is a small, single-chain globular protein of 58 amino acid residues that belongs to the serine protease inhibitor family, where it inhibits protein digestion by blocking the activity of trypsin in the bovine pancreas [1,2]. BPTI and related inhibitors contain conserved cysteine residues that form three disulphide bonds, which strongly stabilise the native fold [2]. The crystal structure of BPTI (Figure 1) reveals a compact, pear-shaped protein of approximately 30 Å along its longest axis, comprising two helices (A_1 and A_2) at the N- and C-termini, respectively, and a central antiparallel β-hairpin formed by strands B_1 and B_2 [3,4]. Three disulphide bonds partition the chain into distinct folding regions: S_1 (Cys5–Cys55) links the N- and C-termini, S_2 (Cys14–Cys38) connects the loose ends of the β-strands, and S_3 (Cys30–Cys51) wraps the helical regions around the β-sheet [2]. Owing to its simple topology, well-resolved crystal structures, and well-characterised folding intermediates and disulphide-exchange kinetics [5–7], as well as its ability to form complexes with multiple target enzymes [8], BPTI has become a paradigmatic system for theoretical studies of protein folding, stability, and molecular recognition. The folding pathway of BPTI has been extensively

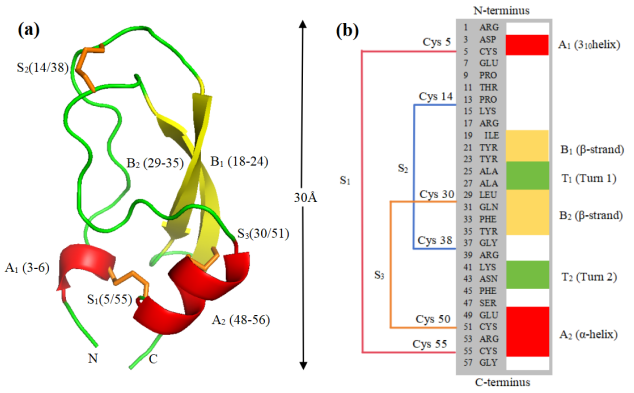

Figure 1: (a) The crystal structure and (b) sequence of BPTI (PDB ID: 5PTI [3]).

characterised, revealing dynamic equilibrium between the native state, disulphide-bonded intermediates, and the underlying network of thiol–disulphide exchange reactions [5, 9–13]. Disulphide formation proceeds via glutathione (GSSG)-mediated oxidation followed by rapid intramolecular disulphide rearrangements whose rates are largely independent of glutathione concentration [9]. Under oxidising conditions, only about half of the molecules reach the native three-disulphide structure, while the remainder accumulate in compact, near-native intermediates in which two native disulphide bonds are formed and the remaining cysteines are buried and resistant to further oxidation [5, 10, 11, 13]. The population of specific intermediates and their conversion to the native state depend sensitively on pH, with neutral and mildly alkaline conditions favouring distinct but convergent disulphide-bonding patterns that share a common, intramolecular rearrangement step that limits the overall folding rate [5, 9, 12]. NMR studies further show that native BPTI exhibits rich internal dynamics, including aromatic side-chain rotations, internal water exchange, and disulphide isomerisation on nanosecond–millisecond timescales [14–16]. Long timescale atomistic molecular dynamics (MD) simulations on the Anton supercomputer have further shown that folded BPTI populates at least five interconverting conformational states with microsecond lifetimes, in which the backbone undergoes rare hops between distinct basins while side chains and internal waters display faster, largely state-independent fluctuations with the dominant MD state closely matches the 5PTI crystal structure [17–19]. More recently, Wesołowski et al. characterised the thermodynamic stability of folded and unfolded BPTI using a three-layer MC-ONIOM3 scheme, in which the whole protein was treated with the GFN-FF force field, selected multicentre fragments were refined with the semi-empirical GFN2-xTB method, and the three disulphide bonds were further refined using the composite DFT method r2SCAN-3c [20–23]. These calculations provided relative and Gibbs free energies for the folded and unfolded states and emphasised the crucial role of the three disulphide bonds and non-covalent interactions in stabilising the native structure. In the following, we compare the ability of the AMBER and UNRES force fields to characterise the conformational behaviour of BPTI within the Cambridge energy landscape framework.

2. Methods

The potential energy landscape of a given system is defined by its potential energy function on the nuclear configuration space, forming a high-dimensional hypersurface in $\mathbb{R}^{3N}$, where N is the number of atoms (or pseudoatoms, in the case of coarse-grained force fields). Because a continuous representation of the PES is almost always impractical, potential energy landscape theory represents the hypersurface in terms of local minima, where the Hessian (the matrix of second derivatives of the potential energy) has only non-negative eigenvalues, and the transition states that connect them, which are first-order saddle points with exactly one negative Hessian eigenvalue [24]. A key feature of any landscape is its global minimum, which corresponds to the most stable structure in the low-temperature limit and often coincides with the native structure of a protein [25]. The Cambridge energy landscape framework allows different potentials to be interfaced as energy functions. In this work, we com-

pare AMBER20 interfaced with the OPTIM [26] program (A20OPTIM) and UNRES interfaced with the OPTIM program (UNOPTIM) [27].

AMBER is a widely used suite for all-atom MD simulations of biomolecules and provides a family of empirical force fields, including ff14SB, for modelling proteins with improved backbone and side-chain dihedral parameters relative to earlier variants such as ff99SB [28–31]. In this work we employ the ff14SB protein force field, whose potential energy function has the standard AMBER form

$$\begin{aligned} E_{\text{potential}} = & \sum_{\text{bonds}} k_l (l - l_{\text{eq}})^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_{\text{eq}})^2 \\ & + \sum_{\text{dihedrals}} \frac{V_n}{2} \left[1 + \cos(n\phi - \gamma)\right] \\ & + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon R_{ij}} \right], \end{aligned} \tag{1}$$

where the first three terms describe bonded contributions (bond stretching, angle bending, and torsions) and the final term accounts for non-bonded van der Waals and electrostatic interactions [27, 32]. The ff14SB parametrisation refines backbone and side-chain torsional profiles using quantum mechanical calculations and NMR data, yielding markedly improved secondary-structure and conformational energetics across diverse peptides and proteins [31, 33, 34].

The UNRES (UNited RESidue) coarse-grained model [35] provides a carefully designed compromise between reduced representation and physicality by coarse graining proteins at the residue level while retaining the essential interaction sites. This level of description enables efficient simulations of large, complex systems over long timescales, as demonstrated in CASP assessments [36, 37] and in a broad range of recent applications [38, 39]. UNRES has been used successfully for systems ranging from small globular proteins to large assemblies [36, 37, 40, 41], and ongoing developments continue to extend its capabilities and accuracy [42, 43]. The model has also been incorporated into the Cambridge energy-landscape framework [27, 44], forming the basis of the UNOPTIM program employed here.

In the UNRES representation, coarse graining is achieved by mapping each amino-acid residue onto a C^α site accompanied by two interaction centres: a united side chain, SC, and a united peptide group, p. The chain geometry is described by virtual-bond lengths, angular degrees of freedom (θ, γ), and the side-chain orientation angles $(\alpha_{\text{SC}}, \beta_{\text{SC}})$. The NEWCT-9P force field [45], version of UNRES implemented in UNOPTIM program, expresses the total potential energy as a sum of physically motivated contributions, U_x, summarised in Equation 2. The long-range nonbonded terms involve interactions between side chains ($U_{\text{SC}_i\text{SC}_j}$), between side chains and peptide groups ($U_{\text{SC}_i\text{p}_j}$), and between peptide groups ($U^{\text{VDW}}_{\text{p}_i\text{p}_j}$ and $U^{\text{el}}_{\text{p}_i\text{p}_j}$). Local terms include virtual-bond deformation (U_{bond}), virtual-bond-angle potentials (U_b), torsional contributions along the backbone (U_{tor}), side-chain rotameric preferences (U_{rot}), and the three-body correction and turn-forming terms, $U^{(3)}_{\text{corr}}$ and $U^{(3)}_{\text{turn}}$. These components collectively account for coupling between short-range geometric preferences and backbone electrostatics [45]. The virtual-bond length d_i corresponds either to a $C^\alpha \cdots C^\alpha$ or a $C^\alpha \cdots \text{SC}$ connection. Disulphide bridges are described by U_{ssbond}, with n_{ss} denoting the number of such cross-links and d_{ss} their bond lengths. The full potential takes the form:

$$\begin{aligned} U = & w_{\text{SC}} \sum_{i<j} U_{\text{SC}_i\text{SC}_j} + w_{\text{SCp}} \sum_{i\neq j} U_{\text{SC}_i\text{p}_j} \\ & + w^{\text{VDW}}_{\text{pp}} \sum_{i<j-1} U^{\text{VDW}}_{\text{p}_i\text{p}_j} + w^{\text{el}}_{\text{pp}} f_2(T) \sum_{i<j-1} U^{\text{el}}_{\text{p}_i\text{p}_j} \\ & + w_{\text{tor}} f_2(T) \sum_i U_{\text{tor}}(\gamma_i, \theta_i, \theta_{i+1}) \\ & + w_{\text{b}} \sum_i U_{\text{b}}(\theta_i) + w_{\text{rot}} \sum_i U_{\text{rot}}(\theta_i, \alpha_{\text{SC}_i}, \beta_{\text{SC}_i}) \\ & + w_{\text{bond}} \sum_i U_{\text{bond}}(d_i) + w_{\text{ssbond}} \sum_{n_{\text{ss}}} U_{\text{ssbond}}(d_{\text{ss}}) \\ & + w^{(3)}_{\text{corr}} f_3(T)\, U^{(3)}_{\text{corr}} + w^{(3)}_{\text{turn}} f_3(T)\, U^{(3)}_{\text{turn}}. \end{aligned} \tag{2}$$

Each contribution is multiplied by a corresponding weight w_x, and those associated with higher-order cumulant terms include additional temperature factors $f_n(T)$ [46]. These coefficients, defined by the first cumulant correction [47], incorporate the reference temperature $T_0 = 300$ K [46, 48] and are given by

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^{n-1}] + \exp[-(T/T_0)^{n-1}]\}}. \tag{3}$$

In potential energy landscape theory, stationary points, defined by vanishing gradients of the potential energy, provide the fundamental structural features of the surface. Of particular interest are local minima, characterised by Hessians with non-negative eigenvalues, and the transition states that link them, which correspond to first-order saddle points with a single negative eigenvalue [24]. Within the discrete path sampling (DPS) framework [25, 49, 50], these stationary points are generated and organised into an evolving database. Here, we use OPTIM [26] to locate and characterise individual stationary points, while PATHSAMPLE [51] distributes successive OPTIM calculations to expand the database systematically. Local minima are obtained via limited-memory BFGS optimisation [52, 53], and candidate connections between selected minima are constructed using doubly-nudged [54, 55] elastic-band [56–58]

(DNEB), with the chain of images relaxed by LBFGS. Transition-state candidates are subsequently refined with hybrid eigenvector-following [59–62], and connectivity is established using steepest-descent pathways to adjacent minima. The resulting collection of minima and transition states, combined with unimolecular rate theory [63], constitutes a kinetic transition network [64–67]. Because this procedure relies on extensive geometry optimisation, the computational cost grows rapidly with system size, reflecting the exponential increase in the number of stationary points expected for complex systems [68,69].

Visualising energy landscapes is challenging because they are naturally embedded in high-dimensional spaces. A common approach is to project the landscape onto a small number of collective variables, but such reductions inevitably obscure aspects of the underlying organisation, for example by merging structurally distinct minima separated by significant barriers [65,70–74]. Disconnectivity graphs [75–78] offer an alternative representation that preserves the essential high-dimensional topology. In these graphs, energy increases vertically and each branch tip denotes the energy (or free energy) of a local minimum; minima are arranged horizontally so that the lowest members occupy the centres of funnel-like features [75,76]. Minima belong to the same superbasin when they can be connected by a pathway whose highest transition state lies below the branching threshold. In practice, minima are arranged horizontally so that the lowest members of local funnels lie near the centre of those funnel features. This representation highlights the global organisation of the landscape: self-assembling systems typically exhibit single-funnel topography [76], whereas glassy or frustrated systems show multifunnel organisation with many competing low-energy states separated by substantial barriers [44, 79, 80]. It is also possible to visualise the underlying energy landscape directly from molecular dynamics simulations, an approach that will be employed in future work and has previously been benchmarked for both UNRES and AMBER simulations [81].

3. Results and Discussion

We begin by discussing the AMBER and UNRES energy landscapes, respectively, followed by their comparison. For each landscape, we performed structural analysis on low-lying structures from each major funnel, which we assume to be representative conformational states for that funnel. For both potentials, a DPS calculation was initiated from five structures obtained from the long-timescale MD simulations of Shaw *et al.* [18]. The resulting kinetic transition network contains 7447 minima and 9134 transition states. The disconnectivity graph constructed reveals four dominant funnels on the landscape (Fig. 2). From these, we selected eight low-lying minima (two per funnel, including the global minimum), labeled A for the global minimum and B–H in alphabetical order with increasing energy, as representative structures for further analysis. These minima provide a compact set of conformations that capture the organisation of the AMBER landscape into distinct funnels.

To characterise differences in surface exposure among these alternative conformations, we computed the solvent-accessible surface area (SASA) for the experimental BPTI structure (PDB ID: 5PTI) and for each of the eight representative minima, decomposing the total SASA into polar and hydrophobic contributions. The PDB structure has a total SASA of 3946 Å^2, with polar and hydrophobic components of 2529 and 1417 Å^2, respectively, corresponding to ≈64% polar exposure, as expected for a compact globular protein. Most AMBER minima show only modest increases in total SASA and preserve a similar polar fraction, indicating that they remain native-like in both compactness and surface polarity distribution. In contrast, some higher-energy funnel representatives (e.g. minima G and H) exhibit a noticeably larger hydrophobic SASA fraction, consistent with more open conformations in which hydrophobic side chains are partially exposed to solvent.

Overall, the SASA analysis suggests that the different funnels on the AMBER landscape correspond to subtly reorganised but broadly native-like folded states, with backbone RMSDs relative to the crystallographic structure not exceeding 4.1 Å (table 1), and with variations in surface exposure that may modulate stability and interactions.

For the UNRES potential, the DPS calculation produced a substantially larger kinetic transition network than for AMBER, comprising 62179 minima and 43725 transition states for comparable CPU time. A disconnectivity graph reveals nine main funnels, although these are less clearly defined than in the case of the AMBER potential, along with several low-lying kinetic traps that lie outside these funnels (Fig. 4). From this landscape we selected one representative minimum per funnel, along with a small set of prominent kinetic traps, as candidate conformational states for further analysis.

Because UNRES is a coarse-grained model in which residues are represented by pseudoatoms, all selected minima were first reconstructed to all-atom structures prior to comparison with the experimental BPTI structure and the AMBER results. Coordinates for chosen minima were extracted from the UNRES database and reconstructed using the PULCHRA software [82] and then side chain refined with SCWRL4 [83] for the disulphide

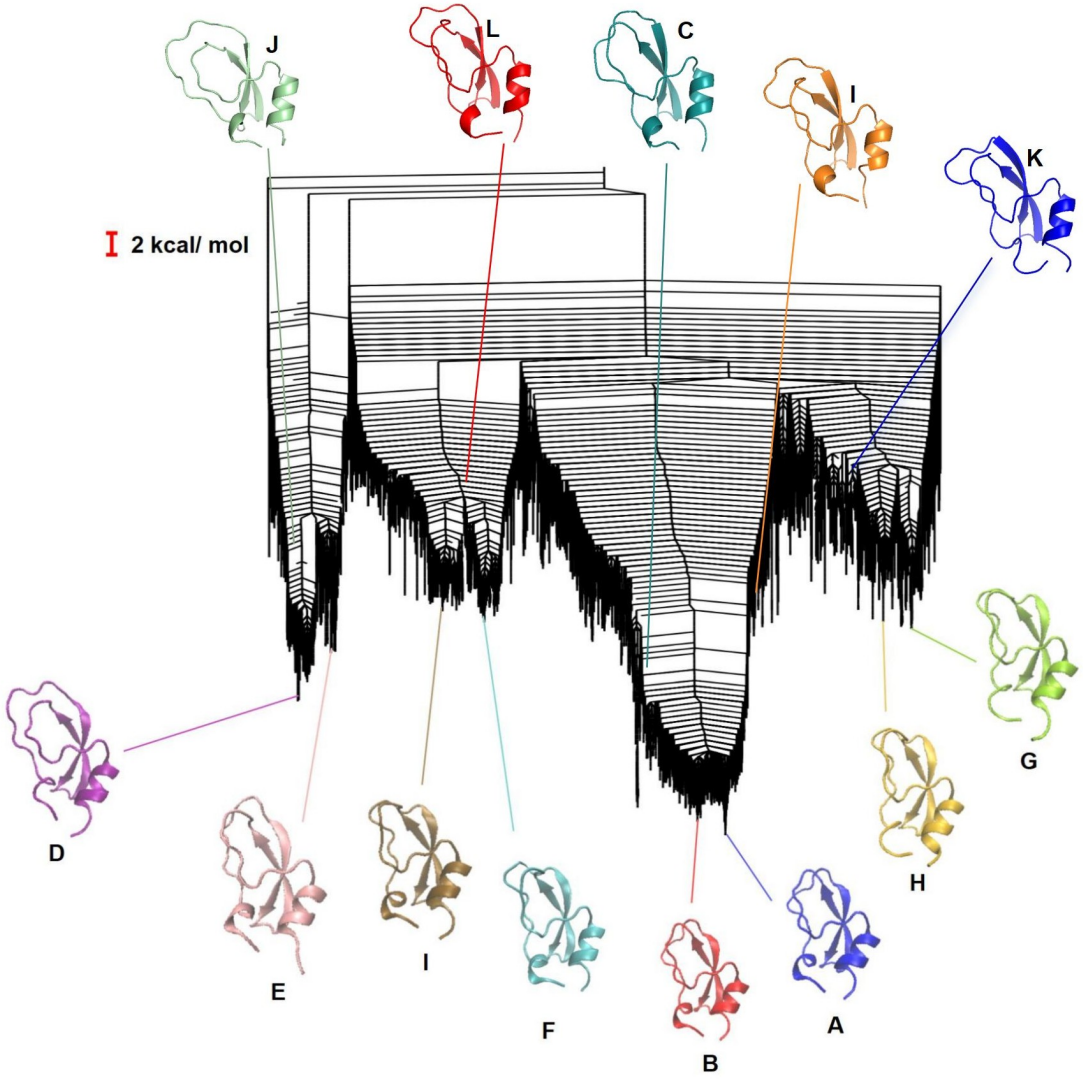

Figure 2: Disconnectivity graph for the BPTI potential energy landscape obtained with the AMBER potential. The four main funnels are indicated, with explicitly displayed minima, starting from minimum A as the global minimum and continuing in alphabetical order with increasing energy. Initial structures are labelled I, C, J, K, L.

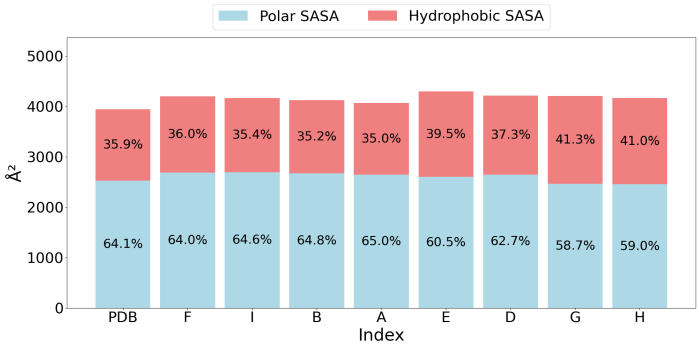

Figure 3: Total, polar and hydrophobic SASA for the experimental BPTI structure (PDB) and the eight representative minima on the AMBER energy landscape.

bridges. Disulphide bonds, hydrogen atoms, solvent molecules and counterions were added, and the resulting systems were relaxed using AMBER. A short energy minimisation was first used to remove unfavourable contacts, after which the systems were gradually heated to 300 K and subsequently equilibrated at 300 K to ensure structural stability prior to analysis. The same refinement protocol has been employed before [44]. SASAs were computed for the reconstructed UNRES conformations and compared with the experimental PDB structure (Fig. 5). Compared to the crystal structure, all UNRES-derived conformations exhibit systematically larger total SASA values in the range 4134–4802 Å^2. The polar fraction varies between ∼60% and 67%, while the hydrophobic contribution is generally increased relative to the experimental structure.

These trends indicate that the UNRES funnels sample more open, solvent-exposed conformations than AMBER, with enhanced exposure of hydrophobic groups in some funnels. Nevertheless, the polar–hydrophobic balance remains broadly comparable to the experimental structure, suggesting that UNRES still favours compact,

folded-like states, albeit with greater conformational diversity and surface rearrangement. We also note that the reconstructed UNRES minima analysed showed higher RMSDs with respect to the crystallographic structure, with the maximal value reaching 5 Å (Table 2).

Disulphide bonds heavily constrain the conformational space explored in both AMBER and UNRES, so our analysis focuses on alternative folded conformations with all three native disulphide bonds present, rather than on the unfolded to folded transition. The disconnectivity graphs obtained for AMBER and UNRES exhibit broadly similar global topologies. This can be seen by the branching topography of the disconnectivity graphs and is further discussed below. In both cases, several low-lying funnels are separated by substantial barriers, with the leftmost funnels being kinetically isolated from the rest.

For AMBER, the global minimum (index A) resides in a central funnel bordered by two clearly separated neighbouring funnels. In UNRES, the corresponding side funnels appear merged into a larger central basin, reflecting the greater smoothing and coarse-graining of this potential. The initial structures used in the DPS calculations occupy similar regions of the landscape in both models, and some starting conformations can be directly identified across the two potentials (e.g. AMBER minimum C and UNRES minimum O represent the same structure). In both disconnectivity graphs this conformation lies within the funnel containing the global minimum, and is among the lowest in energy and entropy of the starting set, indicating that UNRES is able to capture key microscopic features of the folded conformational ensemble despite its reduced representation.

Because AMBER and UNRES employ different functional forms and parametrisations, their absolute energies are not directly comparable. Only relative trends, such as barrier heights between selected minima, can be meaningfully contrasted. Even then, the numerical values differ by factors of a few: for example, barriers between analogous initial structures differ by roughly factors of three to six between the two force fields. These discrepancies reflect the distinct ways in which the two models represent the same underlying states, and mean that the disconnectivity graphs should be compared qualitatively rather than quantitatively in terms of energy scales.

We calculated RMSDs across the representative structures from the energy landscapes of both force fields before and after the AMBER optimisation protocol for the UNRES minima. Figure 6 highlights the importance of the full reconstruction routine in eliminating the geometric artefacts imposed by the reconstruction of a coarse-grained representation. The overall structural mapping between AMBER and UNRES is not one-to-one. Notably, the global minima of the two potentials (minimum A in both disconnectivity graphs) do not correspond to the closest-matching structures in terms of RMSD, and analogous discrepancies are observed for the lowest minima in the leftmost funnels (D (AMBER) vs. K (UNRES)). These observations suggest that, while the two landscapes share similar funnel topologies, the detailed placement of individual minima within those funnels differs between the all-atom and coarse-grained descriptions.

Table 1: The RMSD values of eight local minima, simulated using the AMBER potential, after alignment with experimental PDB structure.

Index	RMSD [Å]
A	2.9
B	2.8
D	4.1
E	3.4
I	2.8
F	2.6
H	3.2
G	3.1

Table 2: The RMSD values of fourteen local minima, simulated using the UNRES potential, after alignment with experimental PDB structure.

Index	RMSD [Å]
A	4.4
B	4.7
C	3.9
D	4.8
E	5.0
F	3.7
G	4.3
H	4.6
I	5.0
J	3.6
K	4.0
L	3.3
M	4.1
N	3.8

In line with the larger and more diverse UNRES database, representative structures from the UNRES landscape generally show larger deviations from the crystallographic BPTI structure than those obtained with AMBER. By contrast, the AMBER representatives remain closer to the experimental structure. Nevertheless, given the constraints imposed by the three disulphide bonds, both force fields predominantly sample folded, native-like

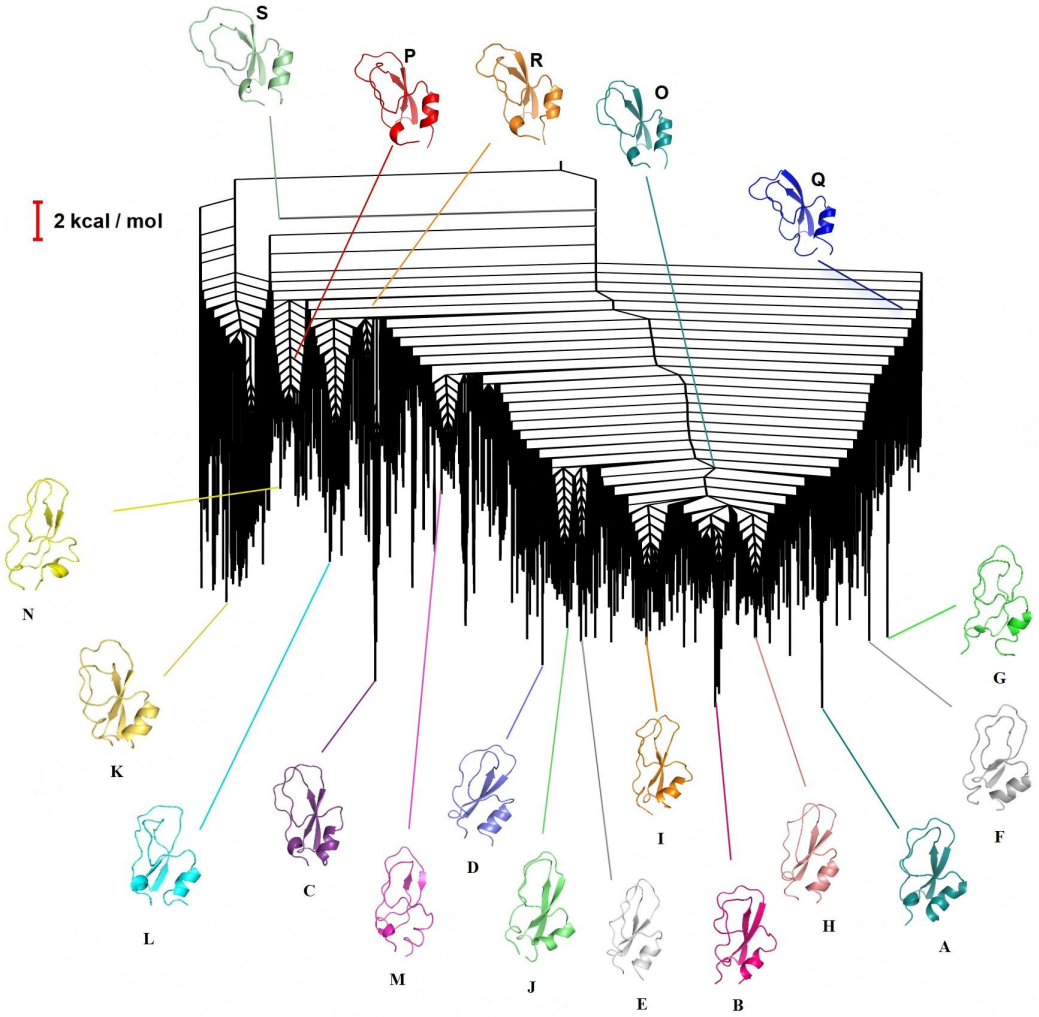

Figure 4: Disconnectivity graph for the BPTI energy landscape obtained with the UNRES potential, showing nine principal funnels and selected representative minima and kinetic traps used for structural analysis, starting from minimum A as the global minimum and continuing in alphabetical order with increasing energy. Initial structures are labelled R, O, S, Q, P.

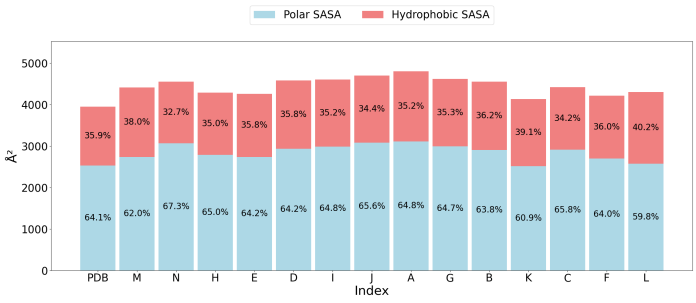

Figure 5: Total, polar, and hydrophobic SASA for the experimental BPTI structure (PDB) and selected representative minima and kinetic traps from the UNRES energy landscape.

conformations, with UNRES providing a more flexible and structurally heterogeneous view of the underlying energy landscape. These findings are summarised in Figures 7 and 8 which show representative structures from both AMBER and UNRES energy landscapes aligned with the crystallographic PDB structure. In AMBER, low-lying minima remain close to the experimental conformation, with RMSDs of roughly 2.5–4.1 Å (table 1) and average per-atom deviations near 1 Å, while in UNRES the corresponding structures (after reconstruction and equilibration) are somewhat more dispersed (RMSDs ∼3.3–5.0 Å) (table 2) but still clearly folded and native-like. In both cases, the α-helix and β-strand regions align well with the PDB structure, and the largest RMSD values, as expected, are predominantly localised in the most flexible loop segments. The grey regions indicate residues for which the per-residue RMSD with respect to the experimental structure exceeds 5 Å, implying that these parts of the structures differ too substantially to allow a reliable structural comparison.

Finally, we note that, consistent with the broader conformational sampling and larger RMSDs observed for UNRES, this force field also explores higher radii of gyration, ranging from 11.29 to 11.96 Å in the analysed structures, whereas AMBER samples a narrower interval of 11.14–11.61 Å, but both of them were sampled close to the crystallographic value of 11.29 Å.

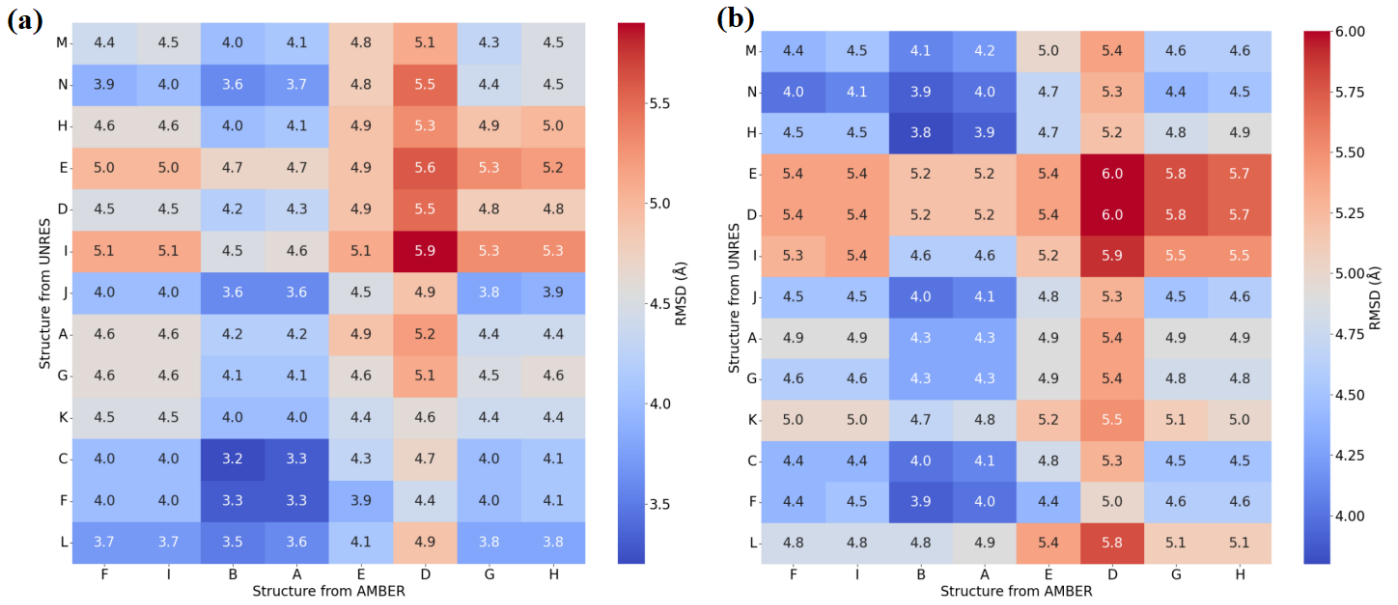

Figure 6: RMSD between representative AMBER and UNRES minima. (a) RMSD using reconstructed and AMBER-optimised UNRES structures. (b) RMSD using reconstructed, unoptimised UNRES structures.

4. Conclusion

The aim of this contribution was to compare the performance of the all-atom AMBER potential and the coarse-grained UNRES potential for BPTI by analysing their energy landscapes and low-lying minima. Despite the very different levels of resolution, both models produce broadly similar folded ensembles, strongly constrained by the three disulphide bonds. This similarity is reflected in native-like RMSD values, comparable average per-atom deviations, and closely matching polar and hydrophobic SASA fractions. For AMBER, the representative minima are highly compact and remain very close to the crystallographic structure in terms of RMSD, radius of gyration, and SASA. The UNRES minima, as expected for a coarse-grained potential that neglects some microscopic interactions, show slightly larger RMSDs, radius of gyration values, and total SASA, indicating more conformational flexibility while still sampling folded, native-like states. We also examined the reconstruction of UNRES minima to all-atom structures. Significant structural changes occur during the necessary optimisation stage with AMBER, reflecting the presence of unphysical geometries (e.g. unusual backbone bond distances) in the raw reconstructed configurations with PULCHRA and confirming that relaxation is essential before quantitative comparison. The correspondence between minima of the two potential energy surfaces obtained is not one-to-one: for instance, the global minima in the two force fields are not mutually closest to the crystallographic structure in RMSD. This mismatch highlights both the intrinsic differences in the underlying potentials and the non-trivial nature of mapping between all-atom and coarse-grained landscapes which posses different conformational spaces by design.

Authors contribution

Songsong Zhou: Performed energy landscape searches, prepared figures, performed data and structural analysis, wrote the original draft and edited the manuscript. **Patryk A. Wesołowski**: Conceptualised the study, supervised the project, performed data analysis, wrote the original draft and edited the manuscript. **Yifei Wang**: Co-supervised the project, contributed to energy landscape searches. **Vilmos Neuman**: Wrote the original draft and edited the manuscript. **Krzysztof K. Bojarski**: Reconstructed UNRES minima, performed data and structural analysis. **David J. Wales**: Conceptualised the study, supervised the project, developed and maintains the Cambridge software for energy landscape exploration. All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

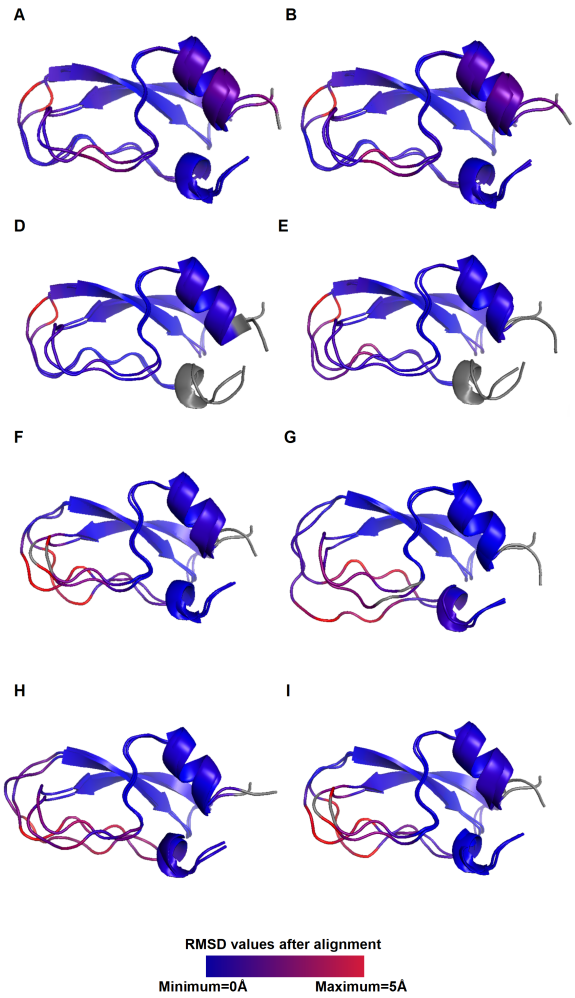

Figure 7: The spectrum of RMSD values for local minima obtained with the AMBER potential, compared to the experimental BPTI structure. Blue regions indicate minimal deviations from the experimental structure, red regions highlight pronounced local differences, and grey regions correspond to parts of the structures that are not included in the comparison due to pronounced structural dissimilarity.

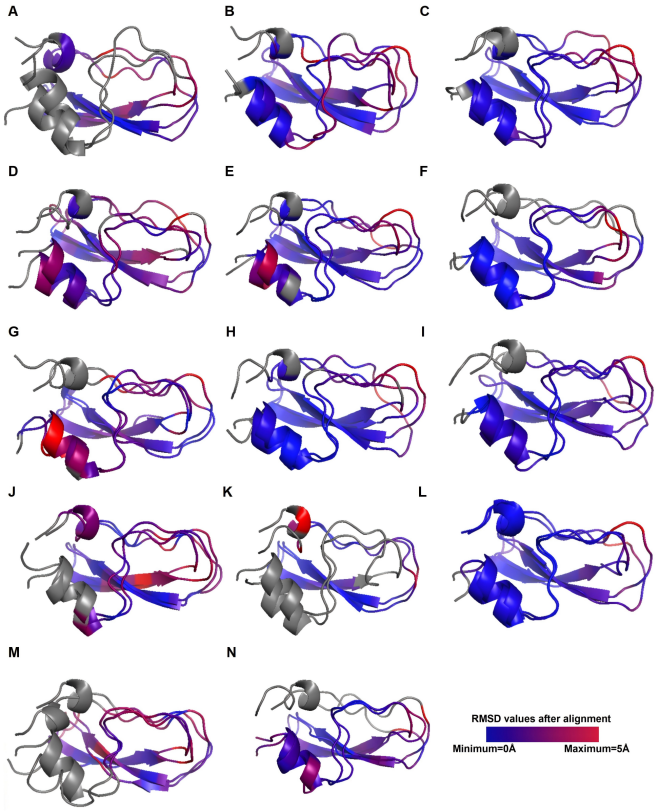

Figure 8: The spectrum of RMSD values for selected minima obtained with the UNRES potential, compared to the experimental BPTI PDB structure. Blue regions indicate minimal deviations from the PDB structure, red regions highlight pronounced local differences, and grey regions correspond to parts of the structures that are not included in the comparison due to pronounced structural dissimilarity.

References

[1] P. Ascenzi, A. Bocedi, M. Bolognesi, A. Spallarossa, M. Coletta, R. D. Cristofaro, and E. Menegatti, "The bovine basic pancreatic trypsin inhibitor (kunitz inhibitor): a milestone protein.," *Curr. Protein Pept. Sci.*, vol. 4, no. 3, pp. 231–251, 2003.

[2] M. Dittrich and C. Kanchanawarin, *Case Study: BPTI.* http://www. ks. uiuc. edu/Training/CaseStudies.: Theoretical and Computational Biophysics Group., 2008.

[3] A. Wlodawer, J. Walter, R. Huber, and L. Sjölin, "Structure of bovine pancreatic trypsin inhibitor: Results of joint neutron and x-ray refinement of crystal form ii," *J. Mol. Biol.*, vol. 180, no. 2, pp. 301–329, 1984.

[4] R. Helland, J. Otlewski, O. Sundheim, M. Dadlez, and A. O. Smalås, "The crystal structures of the complexes between bovine -trypsin and ten p1 variants of bpti," *J. Mol. Biol.*, vol. 287, no. 5, pp. 923–942, 1999.

[5] J. S. Weissman and P. S. Kim, "Reexamination of the folding of bpti: predominance of native intermediates," *Science*, vol. 253, no. 5026, pp. 1386–1393, 1991.

[6] D. P. Goldenberg, "Native and non-native intermediates in the bpti folding pathway," *Trends Biochem. Sci.*, vol. 17, no. 7, pp. 257–261, 1992.

[7] G. I. Makhatadze, K. S. Kim, C. Woodward, and P. L. Privalov, "Thermodynamics of bpti folding," *Protein Sci.*, vol. 2, no. 12, pp. 2028–2036, 1993.

[8] U. Kahler, A. S. Kamenik, F. Waibl, J. Kraml, and K. R. Liedl, "Protein-protein binding as a two-step mechanism: Preselection of encounter poses during the binding of bpti and trypsin," *Biophys. J.*, vol. 119, no. 3, pp. 652–666, 2020.

[9] J. S. Deissman and P. S. Kim, "A kinetic explanation for the rearrangement pathway of bpti folding.," *Nat. Struct. Biol.*, vol. 2, no. 12, pp. 1123–1130, 1995.

[10] D. J. States, C. M. Dobson, and M. Karplus, "A new two-disulphide intermediate in the refolding of reduced bovine pancreatic trypsin inhibitor.," *J. Mol. Biol.*, vol. 174, no. 2, pp. 411–418, 1984.

[11] C. Eigenbrot, M. Randal, and A. A. Kossiakoff, "Structural effects induced by removal of a disulfide-bridge: the x-ray structure of the c30a/c51a mutant of basic pancreatic trypsin inhibitor at 1.6 Å.," *Protein Eng. Des. Sel.*, vol. 3, no. 7, pp. 591–598, 1990.

[12] J. S. Weissman and P. S. Kim, "The pro region of bpti facilitates folding.," *Cell*, vol. 71, no. 5, pp. 841–851, 1992.

[13] J. S. Weissman and P. S. Kimt, "Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase.," *Nature*, vol. 365, no. 6442, pp. 185–188, 1993.

[14] K. Wüthrich and G. Wagner, "Nmr investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor.," *FEBS Lett.*, vol. 50, no. 2, pp. 265–268, 1975.

[15] G. Otting, E. Liepinsh, and K. Wüthrich, "Protein hydration in aqueous solution.," *Science*, vol. 254, no. 5034, pp. 974–980, 1991.

[16] G. Otting, E. Liepinsh, and K. Wüthrich, "Disulfide bond isomerization in bpti and bpti(g36s): An nmr study of correlated mobility in proteins.," *Biochemistry*, vol. 32, no. 14, pp. 3571–3582, 1993.

[17] A. Wlodawer, J. Nachman, G. L. Gilliland, W. Gallagher, and C. Woodward, "Structure of form iii crystals of bovine pancreatic trypsin inhibitor.," *J. Mol. Biol.*, vol. 198, no. 3, pp. 469–480, 1987.

[18] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins.," *Science*, vol. 330, no. 6002, pp. 341–346, 2010.

[19] E. Persson and B. Halle, "Nanosecond to microsecond protein dynamics probed by magnetic relaxation dispersion of buried water molecules.," *J. Am. Chem. Soc.*, vol. 130, no. 5, pp. 1774–1787, 2008.

[20] P. A. Wesołowski, D. J. Wales, and P. Pracht, "Multilevel framework for analysis of protein folding involving disulfide bond formation.," *J. Phys. Chem. B.*, vol. 128, no. 13, pp. 3145–3156, 2024.

[21] S. Spicher and S. Grimme, "Robust atomistic modeling of materials, organometallic, and biochemical systems.," *Angew. Chem. Int. Ed.*, vol. 59, no. 36, p. 15795–15803, 2020.

[22] C. Bannwarth, S. Ehlert, and S. Grimme, "Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions.," *J. Chem. Theory Comput.*, vol. 15, no. 3, pp. 1652–1671, 2019.

[23] S. Grimme, A. Hansen, S. Ehlert, and J. M. Mewes, "r2scan-3c: A "swiss army knife" composite electronic-structure method.," *J. Chem. Phys.*, vol. 154, no. 064103, pp. 1–6, 2021.

[24] J. N. Murrell and K. J. Laidler, "Symmetries of activated complexes," *Trans. Faraday. Soc.*, vol. 64, pp. 371–377, 1968.

[25] D. J. Wales, *Energy Landscapes.* Cambridge: Cambridge University Press, 2003.

[26] D. J. Wales, "OPTIM: A program for optimizing geometries and calculating reaction pathways." (accessed 2025-09-25).

[27] P. A. Wesołowski, A. K. Sieradzan, M. J. Winnicki, J. W. Morgan, and D. J. Wales, "Energy landscapes for proteins described by the unres coarse-grained potential.," *Biophys Chem.*, vol. 303, no. 107107, pp. 1–13, 2023.

[28] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, "An overview of the amber biomolecular simulation package.," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 3, no. 2, pp. 198–210, 2013.

[29] D. A. Case, H. M. Aktulga, K. Belfon, I. Ben-Shalom, S. R. Brozell, D. S. Cerutti, and P. A. Kollman, *Amber 2021.* San Francisco: University of California, 2021.

[30] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," *Proteins*, vol. 65, no. 3, pp. 712–725, 2006.

[31] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3696–3713, 2015.

[32] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[33] H. Nguyen, J. Maier, H. Huang, V. Perrone, and C. Simmerling, "Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent.," *J. Am. Chem. Soc.*, vol. 136, no. 40, pp. 13959–13962, 2014.

[34] P. A. Janowski, C. Liu, J. Deckman, and D. A. Case, "Molecular dynamics simulation of triclinic lysozyme in a crystal lattice.," *Protein Sci.*, vol. 25, no. 1, pp. 87–102, 2016.

[35] A. Liwo, M. Baranowski, C. Czaplewski, E. Gołaś, Y. He, D. Jagieła, P. Krupa, M. Maciejczyk, M. Makowski, M. A. Mozolewska, *et al.*, "A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions," *J. Mol. Model.*, vol. 20, pp. 1–15, 2014.

[36] A. Antoniak, I. Biskupek, K. K. Bojarski, C. Czaplewski, A. Giełdoń, M. Kogut, M. M. Kogut, P. Krupa, A. G. Lipska, A. Liwo, *et al.*, "Modeling protein structures with the coarse-grained unres force field in the casp14 experiment," *J. Mol. Graph. Model.*, vol. 108, p. 108008, 2021.

[37] M. F. Lensink, G. Brysbaert, T. Mauri, N. Nadzirin, S. Velankar, R. A. Chaleil, T. Clarence, P. A. Bates, R. Kong, B. Liu, *et al.*, "Prediction of protein assemblies, the next frontier: The casp14-capri experiment," *Proteins*, vol. 89, no. 12, pp. 1800–1823, 2021.

[38] A. G. Lipska, A. K. Sieradzan, C. Czaplewski, A. D. Lipińska, K. M. Ocetkiewicz, J. Proficz, P. Czarnul, H. Krawczyk, and A. Liwo, "Long-time scale simulations of virus-like particles from three human-norovirus strains," *Journal of Computational Chemistry*, vol. 44, no. 16, pp. 1470–1483, 2023.

[39] A. Liwo, M. Pyrka, C. Czaplewski, X. Peng, and A. J. Niemi, "Long-time dynamics of selected molecular-motor components using a physics-based coarse-grained approach," *Biomolecules*, vol. 13, no. 6, p. 941, 2023.

[40] A. Liwo, C. Czaplewski, A. K. Sieradzan, A. G. Lipska, S. A. Samsonov, and R. K. Murarka, "Theory and practice of coarse-grained molecular dynamics of biologically important systems," *Biomolecules*, vol. 11, no. 9, p. 1347, 2021.

[41] A. K. Sieradzan, C. R. Czaplewski, E. A. Lubecka, A. G. Lipska, A. S. Karczynska, A. P. Gieldon, R. Slusarz, M. Makowski, P. Krupa, M. Kogut, *et al.*, "Extension of the unres package for physics-based coarse-grained simulations of proteins and protein complexes to very large systems," *Biophys. J.*, vol. 120, no. 3, pp. 83a–84a, 2021.

[42] A. G. Lipska, A. M. Antoniak, P. Wesołowski, A. Warszawski, S. A. Samsonov, and A. K. Sieradzan, "Coarse-grained modeling of the calcium, sodium, magnesium and potassium cations interacting with proteins," *J. Mol. Model.*, vol. 28, no. 7, p. 201, 2022.

[43] I. Roterman, A. Sieradzan, K. Stapor, P. Fabian, P. Wesołowski, and L. Konieczny, "On the need to introduce environmental characteristics in ab initio protein structure prediction using a coarse-grained unres force field," *J. Mol. Graph. Model.*, vol. 114, p. 108166, 2022.

[44] P. A. Wesołowski, B. Yang, A. J. Davolio, E. J. Woods, P. Pracht, K. K. Bojarski, K. Wierbiłowicz, M. C. Payne, and D. J. Wales, "Decoding solubility signatures from amyloid monomer energy landscapes," *J. Chem. Theory Comput.*, vol. 21, no. 5, pp. 2736–2756, 2025.

[45] A. Liwo, A. K. Sieradzan, A. G. Lipska, C. Czaplewski, I. Joung, W. Żmudzińska, A. Hałabis, and S. Ołdziej, "A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. iii. determination of scale-consistent backbone-local and correlation potentials in the unres force field and force-field calibration and validation," *J. Chem. Phys.*, vol. 150, no. 15, p. 155104, 2019.

[46] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, and H. A. Scheraga, "Modification and optimization of the united-residue (unres) potential energy function for canonical simulations. i. temperature dependence of the effective energy function and tests of the optimization method with single training proteins," *J. Phys. Chem. B*, vol. 111, no. 1, pp. 260–285, 2007.

[47] R. Kubo, "Generalized cumulant expansion method," *J. Phys. Soc. Jpn.*, vol. 17, no. 7, pp. 1100–1120, 1962.

[48] H. Shen, A. Liwo, and H. A. Scheraga, "An improved functional form for the temperature scaling factors of the components of the mesoscopic unres force field for simulations of protein structure and dynamics," *J. Phys. Chem. B*, vol. 113, no. 25, pp. 8738–8744, 2009.

[49] D. J. Wales, "Some further applications of discrete path sampling to cluster isomerization.," *Mol Phys*, vol. 102, no. 9-10, pp. 891–908, 2004.

[50] T. V. Bogdan, D. J. Wales, and F. Calvo, "Equilibrium thermodynamics from basin-sampling," *J. Chem. Phys.*, vol. 124, p. 044102, Jan. 2006.

[51] D. J. Wales, "Pathsample: A driver for optim to create stationary point databases using discrete path sampling and perform kinetic analysis." (accessed 2025-09-25).

[52] D. C. Liu and J. Nocedal, "Updating quasi-newton matrices with limited storage.," *Math. Comput.*, vol. 35, no. 151, pp. 773–782, 1980.

[53] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Math. Program.*, vol. 45, pp. 503–528, 1989.

[54] S. A. Trygubenko and D. J. Wales, "A doubly nudged elastic band method for finding transition states," *J. Chem. Phys.*, vol. 120, pp. 2082–2094, Feb. 2004.

[55] D. Sheppard, R. Terrell, and G. Henkelman, "Optimization methods for finding minimum energy paths," *J. Chem. Phys.*, vol. 128, p. 134106, Apr. 2008.

[56] G. Mills, H. Jónsson, and G. K. Schenter, "Reversible work transition state theory: application to dissociative adsorption of hydrogen," *Surf. Sci.*, vol. 324, pp. 305–337, Feb. 1995.

[57] G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A climbing image nudged elastic band method for finding saddle points and minimum energy paths," *J. Chem. Phys.*, vol. 113, pp. 9901–9904, Dec. 2000.

[58] G. Henkelman and H. Jónsson, "Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points," *J. Chem. Phys.*, vol. 113, pp. 9978–9985, Dec. 2000.

[59] L. J. Munro and D. J. Wales, "Defect migration in crystalline silicon," *Phys. Rev. B*, vol. 59, pp. 3969–3980, Feb. 1999.

[60] G. Henkelman and H. Jónsson, "A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives," *J. Chem. Phys.*, vol. 111, pp. 7010–7022, Oct. 1999.

[61] Y. Kumeda, D. J. Wales, and L. J. Munro, "Transition states and rearrangement mechanisms from hybrid eigenvector-following and density functional theory.: application to $c_{10}h_{10}$ and defect migration in crystalline silicon.," *Chem. Phys. Lett.*, vol. 341, no. 1-2, pp. 185–194, 2001.

[62] Y. Zeng, P. Xiao, and G. Henkelman, "Unification of algorithms for minimum mode optimization," *J. Chem. Phys.*, vol. 140, p. 044115, Jan. 2014.

[63] W. Forst, *Theory of Unimolecular Reactions.* New York: Academic Press, 1973.

[64] F. Rao and A. Caflisch, "The Protein Folding Network," *J. Mol. Biol.*, vol. 342, pp. 299–306, Sept. 2004.

[65] F. Noé and S. Fischer, "Transition networks for modeling the kinetics of conformational change in macromolecules," *Curr. Opin. Struct. Biol.*, vol. 18, pp. 154–162, Apr. 2008.

[66] D. Prada-Gracia, J. Gómez-Gardeñes, P. Echenique, and F. Falo, "Exploring the Free Energy Landscape: From Dynamics to Networks and Back," *PLoS Comput. Biol.*, vol. 5, p. e1000415, June 2009.

[67] D. J. Wales, "Energy landscapes: some new horizons," *Curr. Opin. Struct. Biol.*, vol. 20, pp. 3–10, Feb. 2010.

[68] D. J. Wales and J. P. K. Doye, "Stationary points and dynamics in high-dimensional systems," *J. Chem. Phys.*, vol. 119, pp. 12409–12416, Dec. 2003.

[69] J. P. K. Doye and D. J. Wales, "Saddle points and dynamics of Lennard-Jones clusters, solids, and supercooled liquids," *J. Chem. Phys.*, vol. 116, pp. 3777–3788, Mar. 2002.

[70] S. V. Krivov and M. Karplus, "Hidden complexity of free energy surfaces for peptide (protein) folding," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, pp. 14766–14770, Oct. 2004.

[71] S. V. Krivov and M. Karplus, "One-Dimensional Free-Energy Profiles of Complex Systems: Progress Variables that Preserve the Barriers," *J. Phys. Chem. B*, vol. 110, pp. 12689–12698, June 2006.

[72] S. Muff and A. Caflisch, "Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein," *Proteins*, vol. 70, pp. 1185–1195, Mar. 2008.

[73] S. V. Krivov and M. Karplus, "Diffusive reaction dynamics on invariant free energy profiles," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, pp. 13841–13846, Sept. 2008.

[74] D. J. Wales, "Perspective: Insight into reaction coordinates and dynamics from the potential energy landscape," *J. Chem. Phys.*, vol. 142, p. 130901, Apr. 2015.

[75] O. M. Becker and M. Karplus, "The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics," , vol. 106, p. 1495, 1997.

[76] O. M. Becker and M. Karplus, "Archetypal energy landscapes," *Nature*, vol. 394, no. 6695, pp. 758–760, 1998.

[77] S. V. Krivov and M. Karplus, "Free energy disconnectivity graphs: Application to peptide models. the journal of chemical physics," *J. Chem. Phys.*, vol. 117, no. 23, pp. 10894–10903, 2002.

[78] D. A. Evans and D. J. Wales, "Free energy landscapes of model peptides and proteins," *J. Chem. Phys.*, vol. 118, pp. 3891–3897, Feb. 2003.

[79] V. K. De Souza and D. J. Wales, "Energy landscapes for diffusion: Analysis of cage-breaking processes," *J. Chem. Phys.*, vol. 129, p. 164507, Oct. 2008.

[80] V. K. De Souza and D. J. Wales, "Connectivity in the potential energy landscape for binary Lennard-Jones systems," *J. Chem. Phys.*, vol. 130, p. 194508, May 2009.

[81] V. Neuman, P. Wesołowski, K. Bojarski, D. Dewan, M. Schäffler, P. Smardz, and D. Wales, "Visualising the Energy Landscape for a Molecular Dynamics Trajectory," *ChemRxiv*, Oct. 2025.

[82] P. Rotkiewicz and J. Skolnick, "Fast procedure for reconstruction of full-atom protein models from reduced representations," *J. Comput. Chem.*, vol. 29, no. 9, pp. 1460–1465, 2008.

[83] G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack, "Improved prediction of protein side-chain conformations with SCWRL4," *Proteins*, vol. 77, pp. 778–795, Dec. 2009.