

RESEARCH ARTICLE

Study of correlations between protein peptide plane dynamics and side chain dynamics

Yanzhen Hou¹, Jiaojiao Liu¹, Jianfeng He¹, Xubiao Peng^{1,2}, Antti J. Niemi^{1,3,4,5*}

1 School of Physics, Beijing Institute of Technology, Beijing 100081, P.R. China, **2** Center for Quantum Technology Research and School of Physics, Beijing Institute of Technology, Beijing 100081, P.R. China, **3** Laboratoire de Mathématiques et Physique Théorique CNRS UMR 6083, Fédération Denis Poisson, Université de Tours, Parc de Grandmont, F37200, Tours, France, **4** Nordita, Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden, **5** Laboratory of Physics of Living Matter, Far Eastern Federal University, 690950, Sukhanova 8, Vladivostok, Russia

* Antti.Niemi@su.se

Abstract

Protein dynamics is pivotal to biological processes. However, experiments are very demanding and difficult to perform, and all-atom molecular dynamics simulations can still not provide all the answers. This motivates us to analyze protein dynamics in terms of different reduced coordinate representations. We then need to resolve how to reconstruct the full all-atom dynamics from its coarse grained approximation. Accordingly we scrutinize all-atom molecular dynamics trajectories in terms of crystallographic Protein Data Bank (PDB) structures, and inquire to what extent is it possible to predict the *dynamics* of side chain $C\beta$ atoms in terms of the *static* properties of backbone $C\alpha$ and O atoms. Here we find that simulated $C\beta$ dynamics at near physiological conditions can be reconstructed with very high precision, using the knowledge of the crystallographic backbone $C\alpha$ and O positions. The precision we can reach with our PDB-based *Statistical Method* reconstruction exceeds that of popular all-atom reconstruction methods such as *Remo* and *Pulchra*, and is fully comparable with the precision of the highly elaborate *Scwrl4* all-atom reconstruction method that we have enhanced with the knowledge of the backbone $C\alpha$ and O atom positions. We then conclude that in a dynamical protein that moves around at physiological conditions, the relative positions of its $C\beta$ atoms with respect to the backbone $C\alpha$ and O atoms, deviate very little from their relative positions in static crystallographic PDB structures. This proposes that the dynamics of a biologically active protein could remain subject to very similar, stringent stereochemical constraints that dictate the structure of a folded crystallographic protein. Thus, our results provide a strong impetus to the development of coarse grained techniques that are based on reduced coordinate representations.

OPEN ACCESS

Citation: Hou Y, Liu J, He J, Peng X, Niemi AJ (2019) Study of correlations between protein peptide plane dynamics and side chain dynamics. PLoS ONE 14(4): e0215141. <https://doi.org/10.1371/journal.pone.0215141>

Editor: L. Michel Espinoza-Fonseca, University of Michigan, UNITED STATES

Received: January 24, 2019

Accepted: March 27, 2019

Published: April 12, 2019

Copyright: © 2019 Hou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data can be found from Protein Data Bank <https://www.wwpdb.org/> by selecting structures with better than 1.0 Å resolution.

Funding: The research of A.J. Niemi has been funded by Swedish Research Council under Contract No. 2018-04411.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The $C\alpha$ atoms are located at the branch points of a protein, they connect the backbone and the side chains. As a consequence their positions are subject to relatively tight stereochemical

constraints. Indeed, in the case of static crystallographic proteins, the all-atom structure can be often determined with a good precision from the knowledge of the $C\alpha$ positions [1–9]. This motivates the so-called $C\alpha$ trace problem where the goal is to construct an accurate all-atom model of a crystallographic folded protein structure solely from the knowledge of the positions of its $C\alpha$ atoms [10–15].

Protein dynamics is pivotal to biological function and many proteins are presumed to be flexible under physiological conditions. Unlike in the case of static crystallographic protein structures, the relative positions of the various atoms can then become variable. There is no *a priori* reason why the dynamical all-atom structure could be determined solely from the knowledge of the $C\alpha$ backbone. Instead the peptide planes and the side chains are expected to twist and bend with respect to the $C\alpha$ backbone, presumably in a complicated fashion. However, high precision experiments in the case of a dynamical protein are very difficult to perform, and our knowledge of atomic level protein dynamics remains limited [16–19]. Presently, the best sources of information are all-atom molecular dynamics simulations. These are probably best exemplified by the very long *Anton* folding trajectories [20, 21].

In [22] a dynamical variant of the $C\alpha$ trace problem is addressed. For example, two all-atom *Anton* trajectories were analyzed in detail, with focus on the peptide plane O atoms and side chain $C\beta$ atoms: The O atom is the only heavy peptide plane atom that is not covalently connected to the $C\alpha$ and the $C\beta$ defines the direction of the side chain. Thus, these two atoms have a particularly important structural information content. Unexpectedly, it was found that the motions of the peptide plane O atoms and the side chain $C\beta$ atoms could be reconstructed with high precision solely from the knowledge of the $C\alpha$ atom dynamics, using a statistical analysis of *static* crystallographic protein structures in Protein Data Bank (PDB) [23]. The results suggest that in an isolated dynamical protein, at near physiological conditions, the motions of its O and $C\beta$ atoms remain strongly slaved to the $C\alpha$ dynamics, to the extent that they are more or less only subject to modest thermal fluctuations. If this is indeed the case, the problem of protein dynamics at physiological conditions could be much simpler than expected: If the motions of different atoms are strongly correlated and in a systematic manner, many aspects of protein dynamics could be modelled by an effective coarse grained energy function, formulated in terms of a reduced sets of coordinates that describe only the $C\alpha$ atoms [24–26] or a group of atoms including the $C\alpha$ as effective interaction centers [27–29].

Here we present an analysis of simulated all-atom protein dynamics in terms of reduced coordinate sets. We focus on the interrelations between the $C\alpha$, peptide plane O and side chain $C\beta$ atoms. This choice is partly motivated by a series of articles [30–32] that propose to investigate the interrelations between the $C\alpha$, the peptide plane O and the side chain $C\beta$ atoms in a Hamiltonian context. Specifically, we inquire how accurately can the dynamics of the $C\beta$ atoms be determined from the knowledge of the backbone $C\alpha$ and O atom dynamics, using solely a statistical analysis of *static* crystallographic PDB structures. We select these three groups of atoms for their pivotal structural role: Alongside the $C\alpha$ atoms the peptide plane O atoms and side chain $C\beta$ atoms are both structurally highly important. The O atoms are the only backbone heavy atoms that have no covalent bond with $C\alpha$ atoms, but they have many very important formative functions. For example, in combination with the peptide plane H atoms they forge and stabilize regular secondary structures including α -helices and β -strands. Similarly, the $C\beta$ atoms determine the orientations of the side chains, strongly affecting the positions of all the higher level side chain atoms. Indeed, in the case of crystallographic protein structures, the knowledge of the peptide planes and the $C\beta$ atom positions is quite sufficient to determine the other side chain atom positions in terms of rotamer libraries [33–37]. This is amply demonstrated by the success of reconstruction programs such as *Pulchra* [13], *Remo* [14] and *Scwrl4* [15].

Our aim is to understand how protein dynamics could be modelled in an effective manner. For this we propose to investigate and develop different coordinate descriptions that can adapt to the dynamics. Previously, several different coordinate systems have been proposed in the literature. Particularly notable are the Ramachandran angles [38, 39], widely used to validate crystallographic protein structures. But these angles provide only local information, they are not sufficient to describe the three dimensional shape, and thus, the dynamics, of a protein [40]. For this alternative coordinate systems have been developed, often starting by envisioning the $C\alpha$ backbone as a piecewise linear polygonal chain for which the $C\alpha$ coordinates \mathbf{r}_i^{α} ($i = 1, \dots, N$) define vertices; here N is the number of $C\alpha$ atoms and we start the indexing from the N terminus. As a linear chain the $C\alpha$ backbone can be framed, and one way is to employ a discrete version of the Frenet frames [41, 42]. The Frenet frames are purely geometric, they do not make any reference to the other heavy atoms of the protein. Nevertheless, discrete Frenet frames provide a very convenient description of a crystallographic protein structure, as it turns out that the $C\beta$ and peptide plane heavy atoms are all highly localized in terms of Frenet frames, quite independently of the proteins amino acid content.

Other purely geometric ways to frame the $C\alpha$ backbone can also be introduced. An example is the parallel transport (Bishop) framing [43]. But in terms of parallel transport frames the $C\beta$ and peptide plane atoms fail to localize [41]. Thus, these frames do not seem to be very convenient, in the study of protein structure. We also note that reconstruction programs such as *Pulchra*, *Remo* and others, commonly use their own, specially designed framings that are built on the $C\alpha$ backbone geometry.

Besides geometry based frames, proteins can also be framed using their structure; the Ramachandran angles are an example. In particular, the combination of $C\alpha$ and $C\beta$ atoms defines such a structure based framing [41, 42]. It employs the covalent bond that connects these two atoms, in addition of the virtual bonds that connect neighboring $C\alpha$ atoms, to define the local frame. Such a $C\beta$ framing adapts well to all amino acids except for glycine that has no $C\beta$ atom. Since the $C\beta$ atoms are highly inert in the discrete Frenet frames, such a $C\alpha C\beta$ based framing yields a description of the protein structure that is very similar to the discrete Frenet framing description.

Here we develop an alternative structure based framing, using the peptide planes: The normal vector of a peptide plane, in combination with the virtual bonds between the $C\alpha$ atoms, can be employed to define such a framing. Our aim is to study how such a peptide plane based framing describes protein dynamics. For this we analyze data from two extended *Anton* simulations [20] that are performed with *Charmm22** force field [21]. The trajectories that we consider are the villin and the ww-domain, reported in [20]. Villin is α -helical and ww-domain is β -stranded in a folded state. The *Anton* simulations observe several transitions between structures that are (apparently) unfolded and that are (apparently) folded. Thus, the combination of these two trajectories cover many local structures, with all the biologically relevant amino acids appearing, except CYS with its unique potential to form sulphur bridges. Moreover, the villin in [20] involves a NLE mutant and the HIS in [20] is protonated. Thus, our analysis includes the effects of mutations and pH variations. All this ensures that there is a good diversity of dynamical details for us to analyze, in terms of these two trajectories.

The length of the villin trajectory is 120 μ s, and we have selected every 20 th simulated structure for our prediction analysis, for a total of 31395 structures. The length of the ww-domain trajectory is 651 μ s and we have chosen every 40 th simulated structure for our prediction analysis, for a total of 60814 structures. We use the coordinates of the backbone $C\alpha$ and O atoms along the *Anton* trajectories, to try and reconstruct the coordinates of the $C\beta$ atoms using a variety of methods. We then compare the results with the $C\beta$ coordinates along the *Anton* trajectories. Besides a $C\alpha$ -Frenet frame based reconstruction [22], and the peptide plane based

Statistical Method reconstruction that we develop here, we also analyze the results that we obtain using commonly available all-atom structural reconstruction programs *Remo* and *Scwrl4*; here we use an enhanced variant of *Scwrl4* that accounts for the positions of the backbone $C\alpha$ and O atoms. We have also performed the analyses using *Pulchra*, with results that comparable to those we obtain with *Remo*. For clarity of representation, we do not present the *Pulchra* results explicitly.

Finally, we remind that the *Anton* simulations provide us with *dynamical* data at physiologically relevant temperature values while all the reconstruction algorithms that we use, are based on *static* crystallographic structures that are commonly measured at very low (liquid nitrogen) temperatures. Thus, our results should also reveal how the structure deforms due to temperature effects in addition of dynamics.

Methods

Framing a protein backbone

Purely geometric discrete Frenet frames. The discrete Frenet frames are purely geometric frames, they are built solely in terms of the $C\alpha$ atoms of a protein backbone. Thus, these frames can be useful *e.g.* for reconstruction of the atomic structure when only the $C\alpha$ atom positions are known. We define the discrete Frenet frames as follows: We assign to each $C\alpha$ position \mathbf{r}_i an orthonormal triplet $(\mathbf{n}_i, \mathbf{b}_i, \mathbf{t}_i)$ [41, 42]. We construct these vectors by first identifying \mathbf{t}_i with the vector that points from the center of the i^{th} $C\alpha$ towards the center of the $(i + 1)^{\text{st}}$ $C\alpha$,

$$\mathbf{t}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|} \tag{1}$$

The binormal vector is normal to the plane formed by three consecutive $C\alpha$ carbons,

$$\mathbf{b}_i = \frac{\mathbf{t}_{i-1} \times \mathbf{t}_i}{|\mathbf{t}_{i-1} \times \mathbf{t}_i|} \tag{2}$$

and the normal vector is

$$\mathbf{n}_i = \mathbf{b}_i \times \mathbf{t}_i \tag{3}$$

For visualization we use a color coding with $\mathbf{n} \sim x \sim \text{red}$ (r), $\mathbf{b} \sim y \sim \text{green}$ (g) and $\mathbf{t} \sim z \sim \text{blue}$ (b) in terms of a right-handed Cartesian $(xyz) \sim (rgb)$ coordinate system; see Fig 1.

We use the discrete Frenet frame vectors to define the (virtual) $C\alpha$ backbone bond (κ) and torsion (τ) angles as follows,

$$\begin{aligned} \kappa_i &= \arccos(\mathbf{t}_{i+1} \cdot \mathbf{t}_i) \\ \tau_i &= \text{sign}(\mathbf{b}_{i+1} \cdot \mathbf{n}_i) \arccos(\mathbf{b}_{i+1} \cdot \mathbf{b}_i) \end{aligned} \tag{4}$$

These angles are shown in Fig 2 We identify them as the canonical latitude (κ) and longitude (τ) angles on the surface of a unit radius (Frenet) sphere \mathbb{S}_x^2 that is centered at the $C\alpha$ atom; the sphere is oriented so that the north pole is in the direction $\kappa = 0$. Thus, the discrete Frenet frame vector \mathbf{t} points to the direction of the north pole; it coincides with the z -direction in a $C\alpha$ centered Cartesian coordinate system. The torsion angle τ coincides with the longitude angle and takes values $\tau \in [-\pi, \pi)$, increasing in the counterclockwise direction around the positive z -axis *i.e.* around vector \mathbf{t} . The great circle $\tau = 0$ passes through the north pole and the tip of the normal vector \mathbf{n} that lies at the equator.

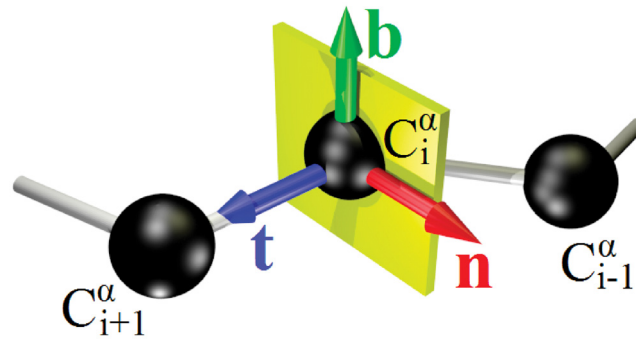


Fig 1. Discrete Frenet frame vectors (1), (2) and (3).

<https://doi.org/10.1371/journal.pone.0215141.g001>

Structure based peptide plane $C\alpha O$ frames. When in addition of the $C\alpha$ atom positions also *e.g.* the peptide plane O atom positions are all known, we can proceed and frame the protein backbone in a structure dependent fashion: The $C\alpha$ atoms $C\alpha(i)$ and $C\alpha(i + 1)$ that are located at the i^{th} and $(i + 1)^{st}$ sites define two corners of the (virtual) peptide plane. The third corner coincides with the i^{th} O atom. Accordingly, when we continue to denote the $C\alpha(i)$ coordinates by \mathbf{r}_i and denote the $O(i)$ coordinates by \mathbf{r}_{O_i} , the O_i centered $C\alpha O$ frames are defined by the following right-handed orthonormal triplet $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}}) \sim (xyz) \sim (rgb)$

$$\hat{\mathbf{u}}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|} \tag{5}$$

$$\hat{\mathbf{w}}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|} \times \frac{\mathbf{r}_{O_i} - \mathbf{r}_i}{|\mathbf{r}_{O_i} - \mathbf{r}_i|} \tag{6}$$

$$\hat{\mathbf{v}}_i = \hat{\mathbf{w}}_i \times \hat{\mathbf{u}}_i \tag{7}$$

In Fig 3 we show these frames. In our reconstruction we shall assume that the peptide planes are ideal. The vector $\hat{\mathbf{w}}_i$ is then a normal vector of the (ideal) peptide plane and we can estimate the coordinates of the remaining peptide plane atoms C_β , N_{i+1} and H_{i+1} , starting from the ideal peptide plane (IPP) geometry that we display in Fig 4.

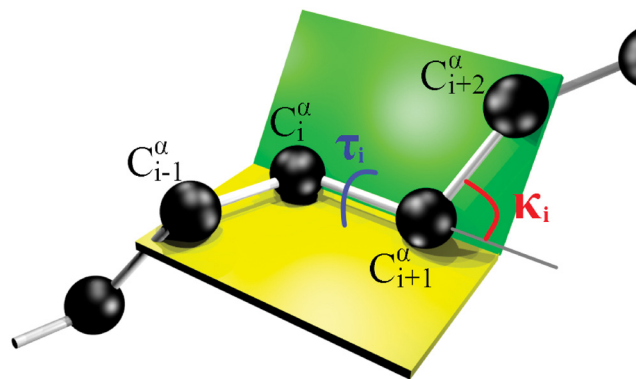


Fig 2. Definition of (virtual) $C\alpha$ backbone bond (κ) and torsion (τ) angles.

<https://doi.org/10.1371/journal.pone.0215141.g002>

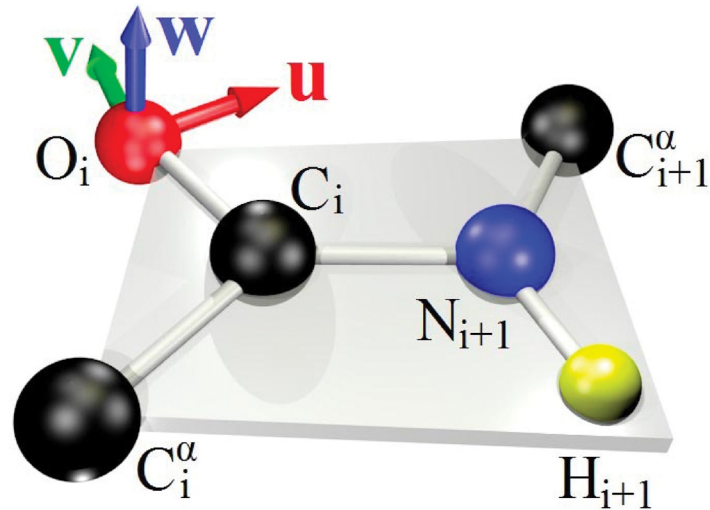


Fig 3. The $C\alpha O$ frames are centered at the O atom of the peptide plane.

<https://doi.org/10.1371/journal.pone.0215141.g003>

Visualizing a protein structure

$C\alpha$ atom visualization. The PDB data pool that we use is the same as used in [22]; see also [40]. It is updated continuously, to consist of all contemporary crystallographic protein structures that have been measured with better than 1.0 Å resolution. We trust that such ultra high resolution structures have been subjected to (at most) very minimal refinement.

For visualization purposes, we use the statistical distribution of the $C\alpha$ backbone bond and torsion angles on a stereographically projected two-sphere S^2 , shown in Fig 5 and constructed

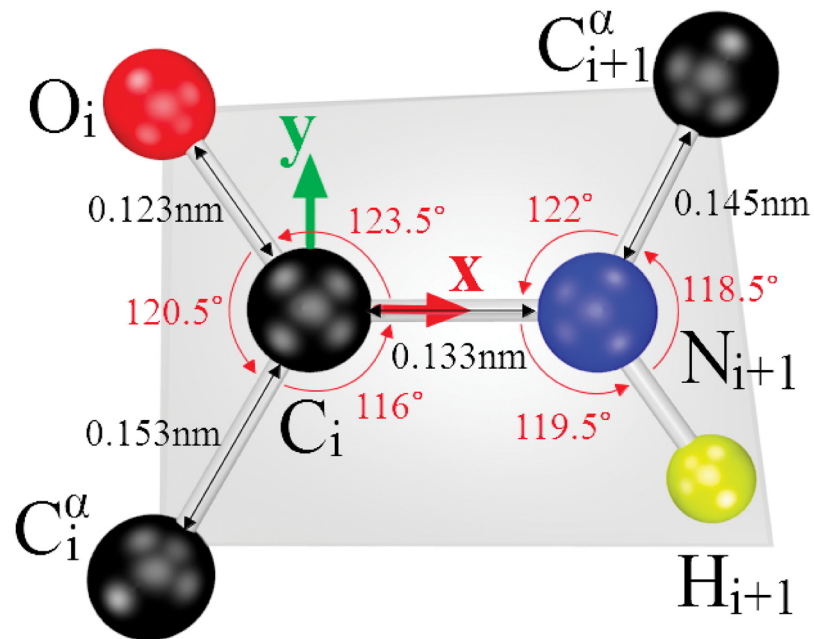


Fig 4. The detailed structure of the ideal peptide plane (IPP) that we use.

<https://doi.org/10.1371/journal.pone.0215141.g004>

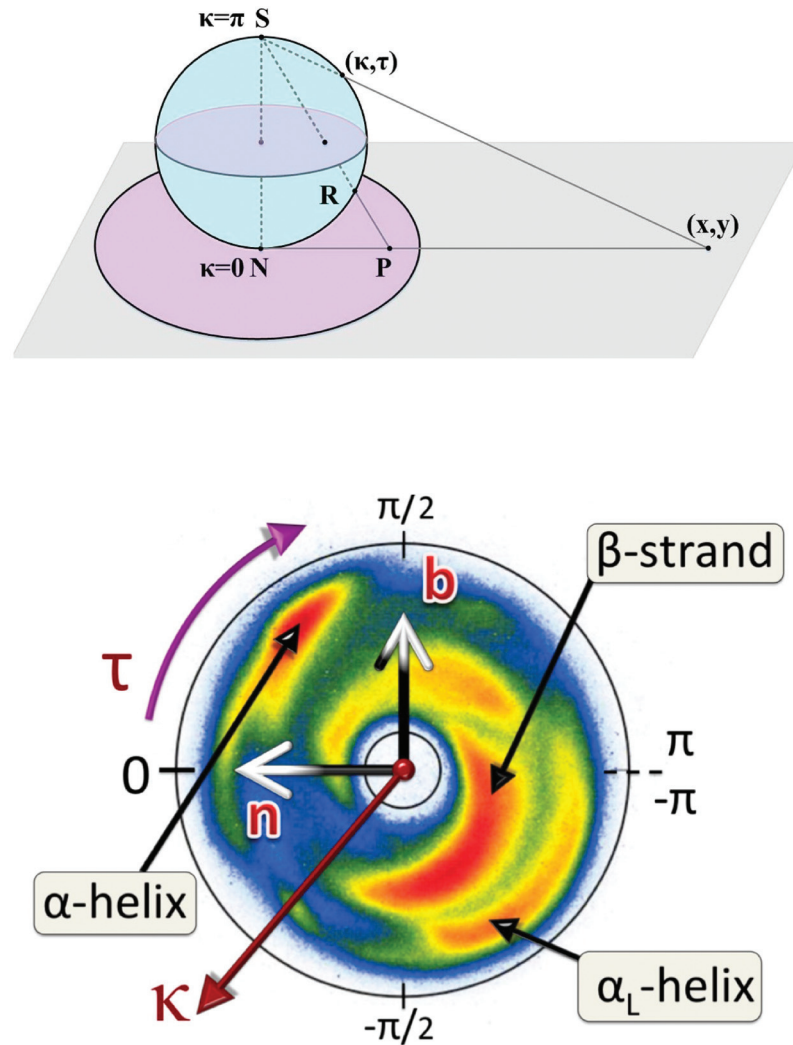


Fig 5. Top: Stereographic projection of two sphere onto plane, with the projection taken from the south pole. Bottom: Statistical distribution of all $C\alpha$ atoms in our pool of PDB structures, on stereographically projected sphere.

<https://doi.org/10.1371/journal.pone.0215141.g005>

as follows: Let $S_x^2(i)$ be a (unit radius) sphere that is centered at the i^{th} $C\alpha$ atom of a given PDB structure. The vector t_i then has its tail at the location of $C\alpha(i)$ and its head lies at the north pole of $S_x^2(i)$, pointing towards the $C\alpha(i + 1)$ atom. The following atom $C\alpha(i + 2)$ is located similarly, in the direction of the discrete Frenet frame vector t_{i+1} that points from the center of $S_x^2(i + 1)$ towards its north pole. We parallel transport t_{i+1} , with no rotation, until its tail coincides with the origin of $S_x^2(i)$. Let (κ_i, τ_i) be the discrete Frenet frame coordinates of the parallel transported head of t_{i+1} on the surface of $S_x^2(i)$. These coordinates describe how a miniature observer at the position of $C\alpha(i)$ atom, with head oriented towards the north pole of $S_x^2(i)$, finds the backbone twisting and bending when she proceeds along the chain to the position of $C\alpha(i + 1)$.

After we repeat this construction for all the $C\alpha$ atoms of all PDB structures in our PDB data pool, we obtain a statistical distribution of coordinates (κ, τ) . We visualize this distribution by projecting the sphere S_x^2 stereographically onto the complex plane as

shown in Fig 5. The projection is defined by

$$x + iy = \tan\left(\frac{\kappa}{2}\right)e^{i\tau}$$

In Fig 5 we also show the (κ, τ) distribution of all the $C\alpha$ atom coordinates in our PDB data set, on the stereographically projected Frenet sphere. The distribution is largely concentrated inside an annulus, with inner circle $\kappa_{in} \approx 1$ and outer circle $\kappa_{out} \approx \pi/2$. In the Figure we have identified the regular secondary structure regions corresponding to α -helices, β -strands and left-handed α -helices.

Visualization of peptide plane O atoms and side chain $C\beta$ atoms. To visualize the peptide plane O atom and side chain $C\beta$ atom distributions in our PDB data set, we use the $C\alpha$ centered Frenet spheres S_x^2 but with no stereographic projection. We depict the direction of these atoms on the surface of S_x^2 , exactly as they are seen by a discrete Frenet frame observer who stands at the position of the $C\alpha$ atom. In Fig 6 we show the ensuing statistical distributions that we obtain for the O and $C\beta$ atoms in our statistical pool of PDB structures. Again, the major regular secondary structures have been identified.

The Statistical Method reconstruction algorithm

Our *Statistical Method* reconstruction algorithm is an extension of the algorithm developed in [22]. It builds on the $C\alpha O$ framing, it aims to use the known positions of $C\alpha$ and O atoms to predict the positions of the remaining atoms along a protein chain. As an example, we here predict the spherical coordinates (θ, ϕ) i.e. directions of $C\beta$ atoms on the surface of a proper two-sphere S^2 , using the knowledge of the $C\alpha$ coordinates (κ, τ) and O coordinates that we denote (ϑ, ψ) in the sequel.

We do not address fluctuations in the covalent bond lengths and other distance scales. By construction, any reconstruction that builds on crystallographic data can only reproduce (at most) very long time scale average values of these quantities. Besides, our analyses confirm that the radial distance variations in *Anton* trajectories are minor in comparison to the

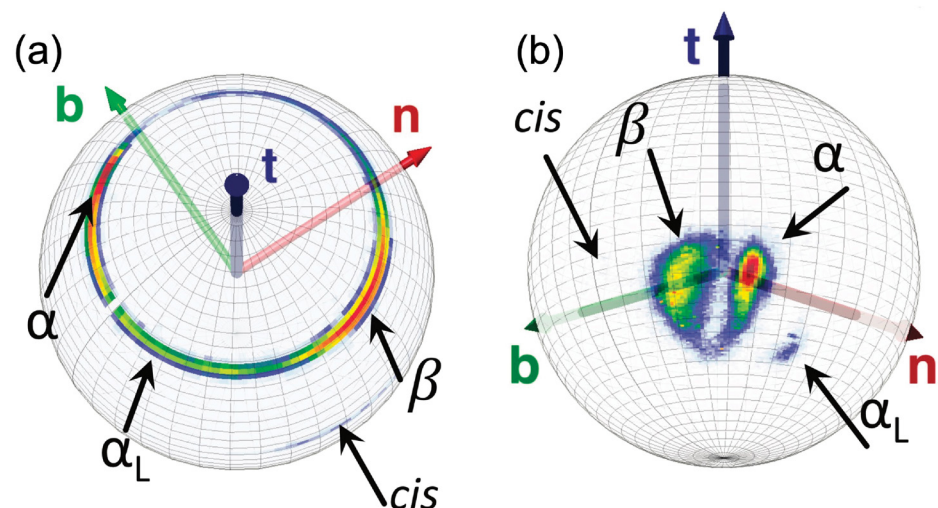


Fig 6. Fig a): The distribution of peptide plane O atom coordinates (ϑ, ψ) in our PDB pool of structures on the surface of the Frenet sphere S_x^2 . Fig b): The distribution of $C\beta$ atom coordinates (θ, ϕ) in our PDB pool of structures on the surface of the Frenet sphere S_x^2 . In both Figures we have identified the α -helices (α), β -strands (β), left-handed α -helices (α_L) and *cis*-peptide planes (*cis*).

<https://doi.org/10.1371/journal.pone.0215141.g006>

directional (angular) variations. Thus, for the $C\alpha$ - $C\beta$ covalent bond length we use the *Anton* average value 1.55 Å and for the other distances we refer to Fig 4.

We start as follows [22]: We divide the statistical distribution of our PDB data shown in Fig 5 (bottom) into 60 equal sectors with angular size $\Delta\tau = \pi/30$ radians. We then divide each of these sectors into two radial sets, one corresponding to bond angle values $\kappa > 1.2$ (rad) and the other to $\kappa \leq 1.2$ (rad): The value $\kappa = 1.2$ divides the annulus in Fig 5 (roughly) into two annuli corresponding to α -helix-like ($\kappa > 1.2$) and β -strand-like ($\kappa < 1.2$) structures.

Next, we observe that the statistical distribution of peptide plane O atoms shown in Fig 6 (top) forms a very narrow circular strip with latitude $\vartheta \approx \pi/4$. Accordingly, we further divide the PDB data in terms of the O atom coordinate values, by longitudinally dividing the statistical O distribution of Fig 6 into 90 equal length $\Delta\psi$ segments with length $\Delta\psi = \pi/45$.

We use our division of PDB structures according to the (κ, τ, ψ) values, to predict the $C\beta$ coordinates (θ, ϕ) of a given structure as follows:

Step 1. We start from the known τ_i value of the given $C\alpha(i)$ atom, to select the pertinent sector $\tau_i \in \Delta\tau$. We then use the κ_i value of $C\alpha(i)$ together with κ_{i+1} of the $C\alpha(i + 1)$ atom, to further assign $C\alpha(i)$ to one of the four sets

$$\begin{aligned} \text{Set } \Delta\kappa_1 : & \quad \kappa_i < 1.2 \quad \& \quad \kappa_{i+1} < 1.2 \\ \text{Set } \Delta\kappa_2 : & \quad \kappa_i < 1.2 \quad \& \quad \kappa_{i+1} \geq 1.2 \\ \text{Set } \Delta\kappa_3 : & \quad \kappa_i \geq 1.2 \quad \& \quad \kappa_{i+1} < 1.2 \\ \text{Set } \Delta\kappa_4 : & \quad \kappa_i \geq 1.2 \quad \& \quad \kappa_{i+1} \geq 1.2 \end{aligned}$$

We then use the ψ_i value of $O(i)$ to select the $\Delta\psi$ segment. These steps identify a subset of $[\Delta\kappa; \Delta\tau; \Delta\psi]$ to which we putatively assign the $C\beta(i)$ atom.

Step 2. We proceed to consider those PDB structures for which the (κ, τ, ψ) values are in the same subset $[\Delta\kappa; \Delta\tau; \Delta\psi]$. Among these we search for a PDB structure that has equal amino acids at two consecutive sites, as the $C\beta$ atom of interest.

- If we find only one such matching PDB structure we use the coordinates of its $C\beta$ atom for our prediction.
- If we have more than one pair of matching amino acids in the PDB subset, we use the average values of the ensuing $C\beta$ coordinates for our prediction.
- If there are no PDB entries with matching amino acid pairs, we use the average value of $C\beta$ coordinates (θ, ϕ) of all PDB structures in the given subset $[\Delta\kappa; \Delta\tau; \Delta\psi]$ for our prediction.
- Finally, if the PDB subset $[\Delta\kappa; \Delta\tau; \Delta\psi]$ has no entries, we use a neighboring subset for our prediction. In selecting the neighboring subset, we move clockwise in along the strip of O atoms, then clockwise in terms of the torsion angle.

In Step 2, the average latitude value κ_{ave} is simply

$$\kappa_{ave} = \frac{1}{N} \sum_{i=k}^N \kappa_k$$

where the summation is over all elements in the given subset $[\Delta\kappa; \Delta\tau; \Delta\psi]$. For the average longitude value τ_{ave} we first define

$$X = \frac{1}{N} \sum_{i=k}^N \cos \tau_k \quad \& \quad Y = \frac{1}{N} \sum_{i=k}^N \sin \tau_k$$

and

$$R = \sqrt{X^2 + Y^2}$$

The average value is then obtained from

$$\cos \tau_{ave} = \frac{X}{R} \quad \& \quad \sin \tau_{ave} = \frac{Y}{R}$$

We compute the average longitude value ψ_{ave} in the same way.

The present *Statistical Method* reconstruction algorithm is extremely simple and computationally very fast, in fact much faster than any of the other reconstruction programs that we have considered even though we have not optimized the search algorithm, but use a straightforward MatLab code. We note that the sizes of the subsets $\Delta\kappa$, $\Delta\tau$, $\Delta\psi$ can be changed and optimized, the choice we have made here is only to exemplify the method.

Results

We report on results that we have obtained using our *Statistical Method* reconstruction algorithm, and the publicly available reconstruction algorithms *Remo* and *Scwrl4*. Since *Scwrl4* requires that the peptide planes are known, we first construct the peptide planes using the $C\alpha$ and O atom coordinates from the *Anton* simulation. We then estimate the positions of peptide plane C, N and H atoms using the ideal peptide plane structure shown in Fig 4. Accordingly we denote this enhanced variant *ScwrIPP* in the sequel. The present *Statistical Method* reconstruction algorithm will be denoted *Stat* in the sequel. Note that we do not use the full *Anton* peptide planes in *ScwrIPP*. This is so that the information content in *ScwrIPP* and *Stat* are comparable.

Description of Anton data distributions

We start by displaying the various data distributions in the *Anton* simulations; for the bond and torsion angles of the $C\alpha$ backbone, for the peptide plane structures and for the $C\beta$ atoms on the surface of a two sphere.

Bond and torsion angle distributions in Anton simulations. In Fig 7 we show the distributions of the $C\alpha$ bond and torsion angles on the stereographically projected two sphere of Fig 5, for the villin and ww-domain trajectories in the *Anton* data.

Both distributions in Fig 7 are somewhat different from the PDB distribution in Fig 5. Specifically, the villin distribution is largely concentrated in the vicinity of the α -helical region while the ww-domain distribution has a wider spread. The data distributions in Fig 7 both confirm that the *Anton* data has a substantial dynamical content.

Peptide plane structure in Anton simulations. We proceed to confirm that the ideal, crystallographic peptide plane structure of Fig 4 is indeed preserved, in the course of *Anton* simulations. For this we plot the statistical distributions of the $C(i)$ and $N(i+1)$ directions as seen in the $O(i)$ centered $C\alpha O$ frames, in the *Anton* trajectories of villin and ww-domain. The results are shown in Fig 8: We observe that all four distributions are consistent with the ideal peptide plane structure of Fig 4, to the extent that any (minor) deviation could be attributed to a thermal fluctuation.

$C\beta$ distributions in Anton simulations. In Fig 9 we show the $C\beta$ density distributions of Fig 6, along the *Anton* trajectories. In each case the distributions have a support which is very similar to that in Fig 6; the villin distribution has a higher population in the vicinity of the α -helical region of Fig 6 and the ww-domain has a higher population in the vicinity of the β -stranded region, as expected.

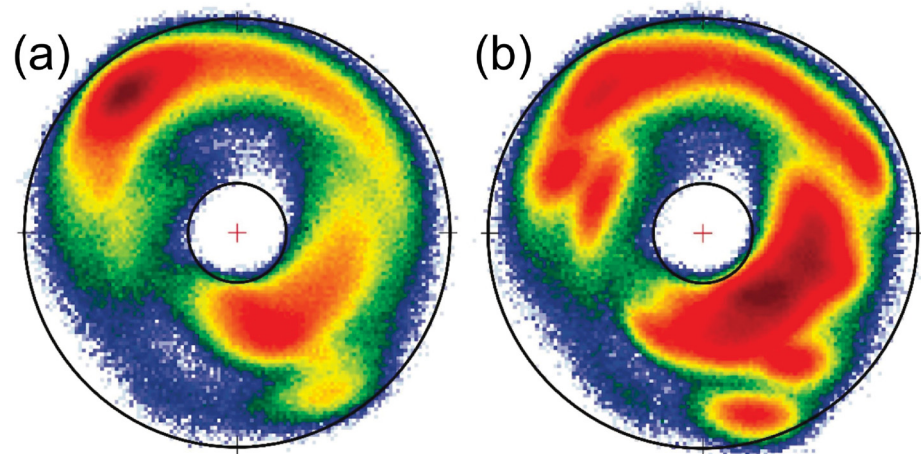


Fig 7. a) Statistical distribution of $C\alpha$ atoms in villin structures of *Anton*. b) Statistical distribution of $C\alpha$ atoms in ww-domain structures of *Anton*.

<https://doi.org/10.1371/journal.pone.0215141.g007>

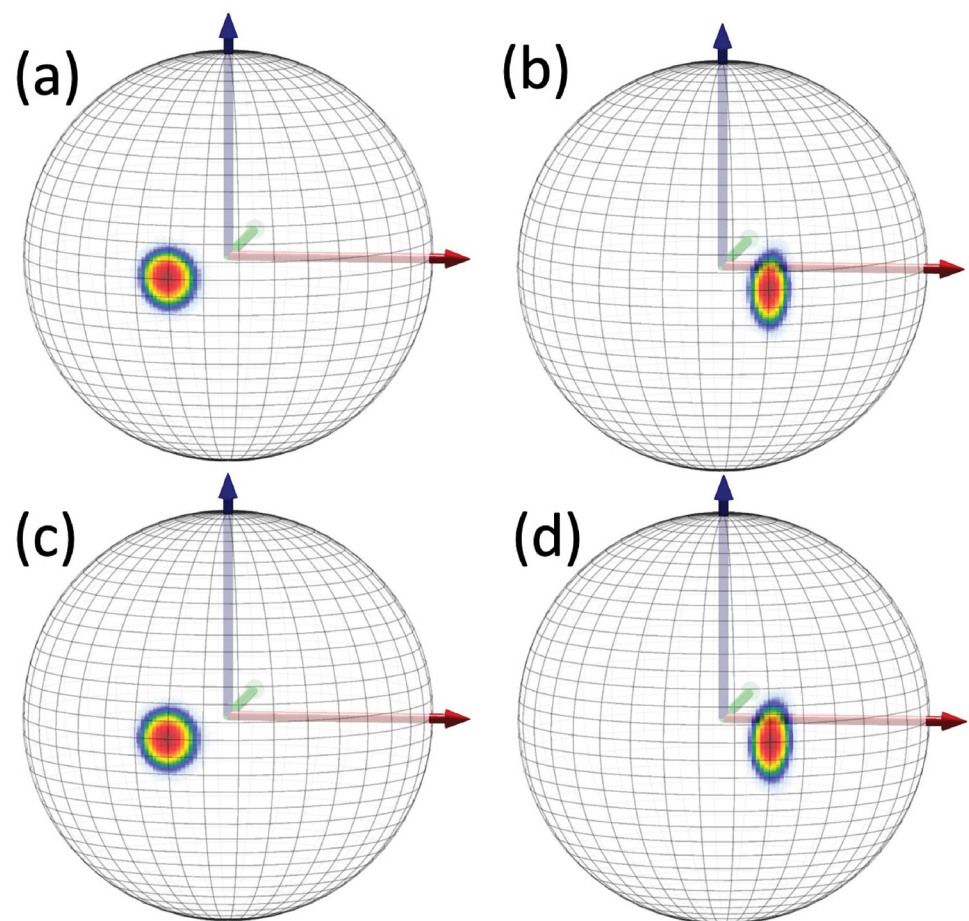


Fig 8. Statistical distribution of peptide plane $C(i)$ and $N(i + 1)$ atoms in *Anton* simulations, as seen in the $O(i)$ centered $C\alpha O$ frames. a) $C(i)$ in villin, b) $N(i + 1)$ in villin, c) $C(i)$ in ww-domain, d) $N(i + 1)$ in ww-domain.

<https://doi.org/10.1371/journal.pone.0215141.g008>

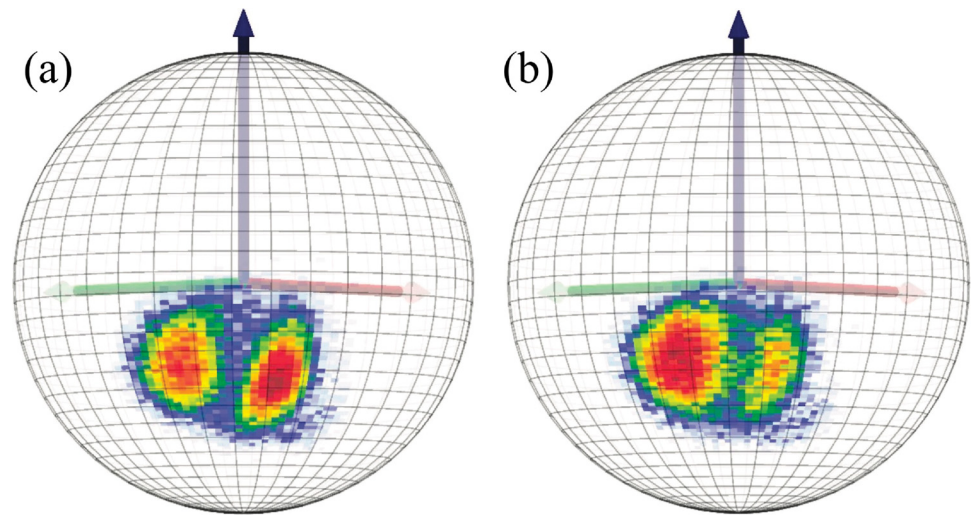


Fig 9. Statistical distribution of side chain $C\beta(i)$ atoms as seen from the position of $C\alpha(i)$ atoms along the *Anton* trajectory; a) villin and b) ww-domain.

<https://doi.org/10.1371/journal.pone.0215141.g009>

Reconstruction of side chain $C\beta(i)$ coordinates

Comparison of $\overrightarrow{C\alpha C\beta}$ direction with Frenet frame based reconstructions. In [22] a Frenet frame based method has been developed to reconstruct the positions of atoms along a protein chain, from the knowledge of the $C\alpha$ atom positions. The present *Statistical Method* builds on that method. Thus, by comparing the results from these two methods we obtain an understanding how the knowledge of O atom positions influences the quality of reconstruction. For this we consider the vectors $\overrightarrow{C\alpha C\beta}$ that point from the $C\alpha$ atom to the ensuing $C\beta$ atom. We evaluate these vectors from *Anton* simulations, and we reconstruct them using the two methods. We then compute the statistical (probability) distributions of the angles between the *Anton* vectors and the reconstructed vectors. The results are shown in Fig 10.

We find that in both methods, the predicted $C\beta$ positions are extremely close to the original *Anton* results: The peaks of all four distributions in Fig 10 are very strongly peaked in the range of 0.08–0.11 radians; the values are slightly larger in the Frenet frames based reconstruction. With the $C\alpha$ - $C\beta$ average covalent bond length ~ 1.55 Å this corresponds to a distance in the range of 0.12–0.15 Å, which is hardly observable with presently available atomic level x-ray spectroscopy. For example, in our pool of ultra high resolution PDB structures, the Debye-Waller B-factor fluctuation distances all have values that are larger than 0.15 Å [22].

We observe from Fig 10 that the present, *Statistical Method* reconstructed $C\beta$ atoms are systematically slightly closer to the $C\beta$ atoms that are reconstructed in the Frenet frame based method. Thus, the inclusion of the peptide plane O atoms does appear to improve the reconstruction precision.

Comparison of $\overrightarrow{C\alpha C\beta}$ direction with ScwrIPP and Remo reconstruction. In Fig 11 we compare the $\overrightarrow{C\alpha C\beta}$ angles between the present *Statistical Method* and *ScwrIPP*. For comparison, we also compare with results from the *Remo* reconstruction program. In that case the only input is the $C\alpha$ positions of the *Anton* trajectory, the $O(i)$ positions are determined by the program. The results shown in Fig 11 display the statistical (probability) distributions of the angles between the *Anton* vectors $\overrightarrow{C\alpha C\beta}$ and the reconstructed vectors $\overrightarrow{C\alpha C\beta}$. The

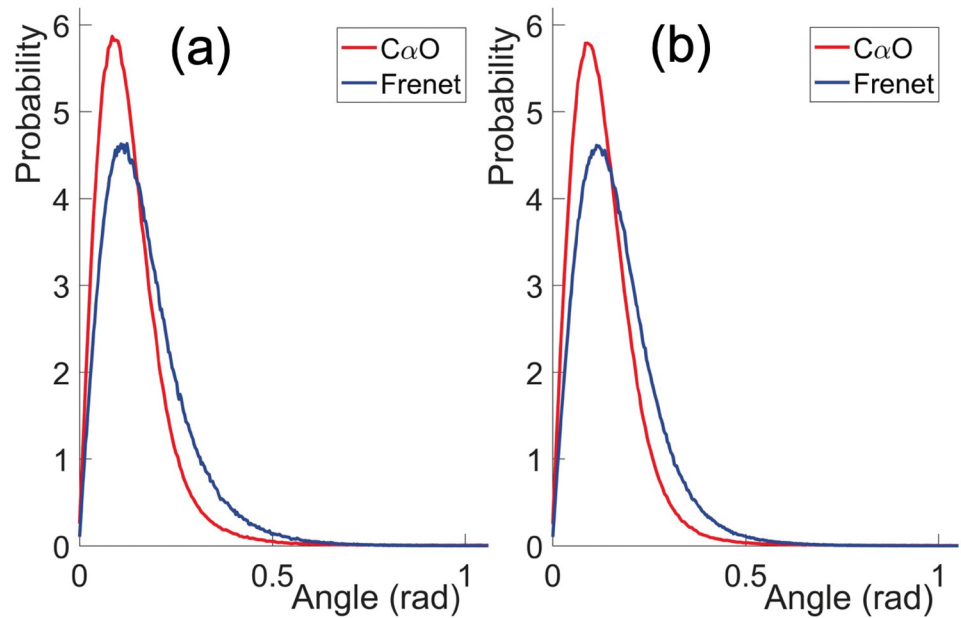


Fig 10. Comparison of the angle between the \overrightarrow{CaO} directions in original *Anton* data and reconstructed data using the present *CaO* based *Statistical Method* and the Frenet frames based method of [22]. a) for the villin data and b) for the ww-domain data.

<https://doi.org/10.1371/journal.pone.0215141.g010>

results are very good in all three methods, those obtained with *ScwrIPP* are slightly better than those obtained with *Statistical Method*. The comparison to *Remo* reconstruction confirms that the inclusion of O atom positions leads to a visible and systematic improvement in the precision.

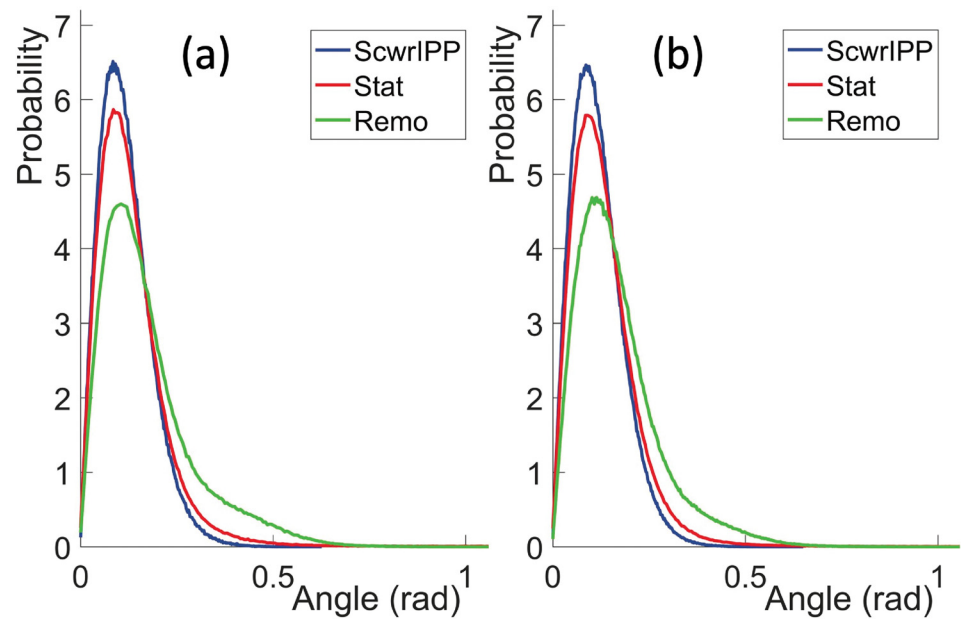


Fig 11. Comparison of the angle between the \overrightarrow{CaO} directions in original *Anton* data and reconstructed data using the present *Statistical Method* (*Stat*), *ScwrIPP* and *Remo*. a) for the villin data and b) for the ww-domain data.

<https://doi.org/10.1371/journal.pone.0215141.g011>

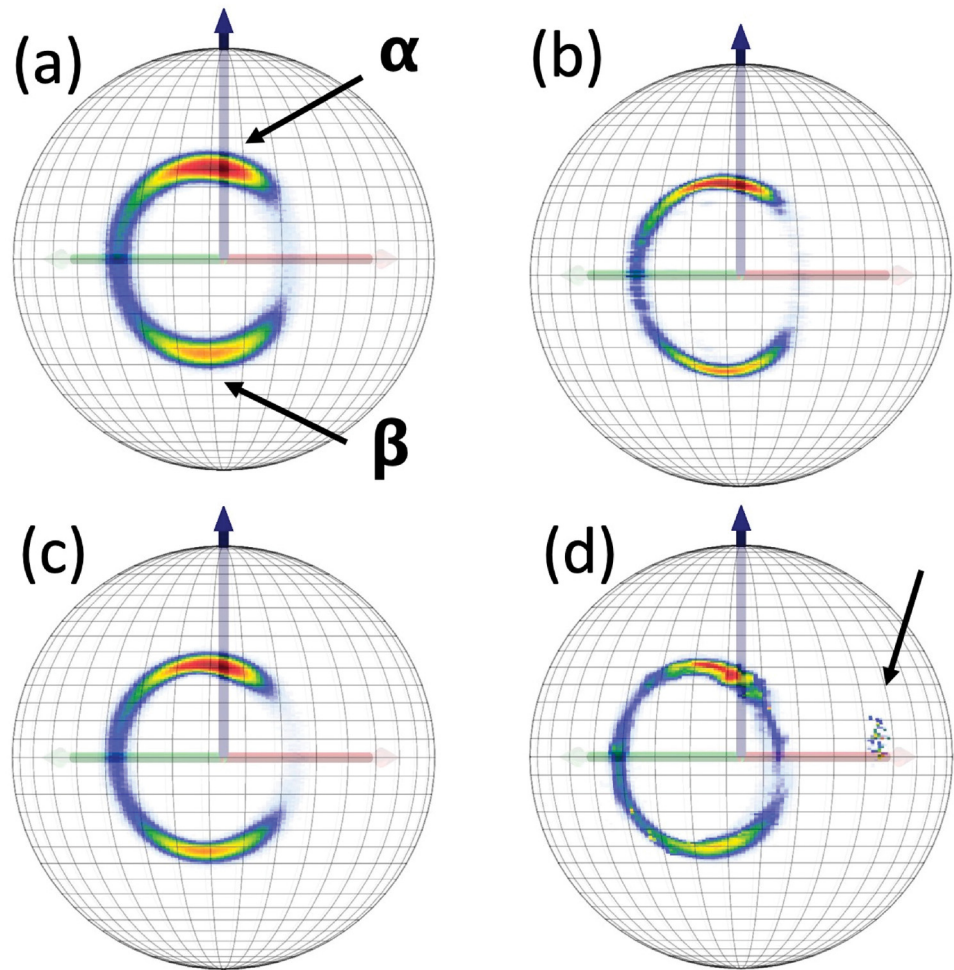


Fig 12. Statistical distribution of villin $C\beta(i)$ atoms, when viewed from the position of $O(i)$ atoms in $C\alpha O$ frames. a) Original *Anton* data, b) *Statistical Method* reconstructed data c) *ScwrIPP* reconstructed data and d) *Remo* reconstructed data. In a) we have identified the α -helical and β -stranded regions. In d) we have pointed out a dislocated region in *Remo* reconstruction.

<https://doi.org/10.1371/journal.pone.0215141.g012>

Statistical distributions of $\overrightarrow{O(i)C\beta(i)}$ on the two-sphere. In Figs 12 and 13 we show various statistical distributions of $C\beta(i)$ atoms, as they are seen in the $C\alpha O$ frames on the surface of a $O(i)$ centered two-sphere. We show the results for the original *Anton* data, and the results from *Statistical Method*, *ScwrIPP* and *Remo* reconstructions. The α -helical and β -stranded regions are clearly identifiable in all distributions. We observe that *Statistical method* and *ScwrIPP* reproduce properly the original villin and ww-domain distributions. But in the case of *Remo* the distributions shows fragmentation and are also slightly, but systematically, shifted to the left, both in the case of villin and ww-domain. There is also an apparent small dislocated region in the villin distribution of *Remo*, pointed out in Fig 12d).

Comparison of $\overrightarrow{O(i)C\beta(i)}$ direction with *ScwrIPP* and *Remo* reconstruction. In Fig 14 we compare the deviations in the statistical (probability) distributions for the reconstructed $\overrightarrow{O(i)C\beta(i)}$ angles in the case of *Statistical Method*, *ScwrIPP* and *Remo*. The reference structures are the $\overrightarrow{O(i)C\beta(i)}$ angles along the *Anton* trajectories, for villin and ww-domain. The deviations

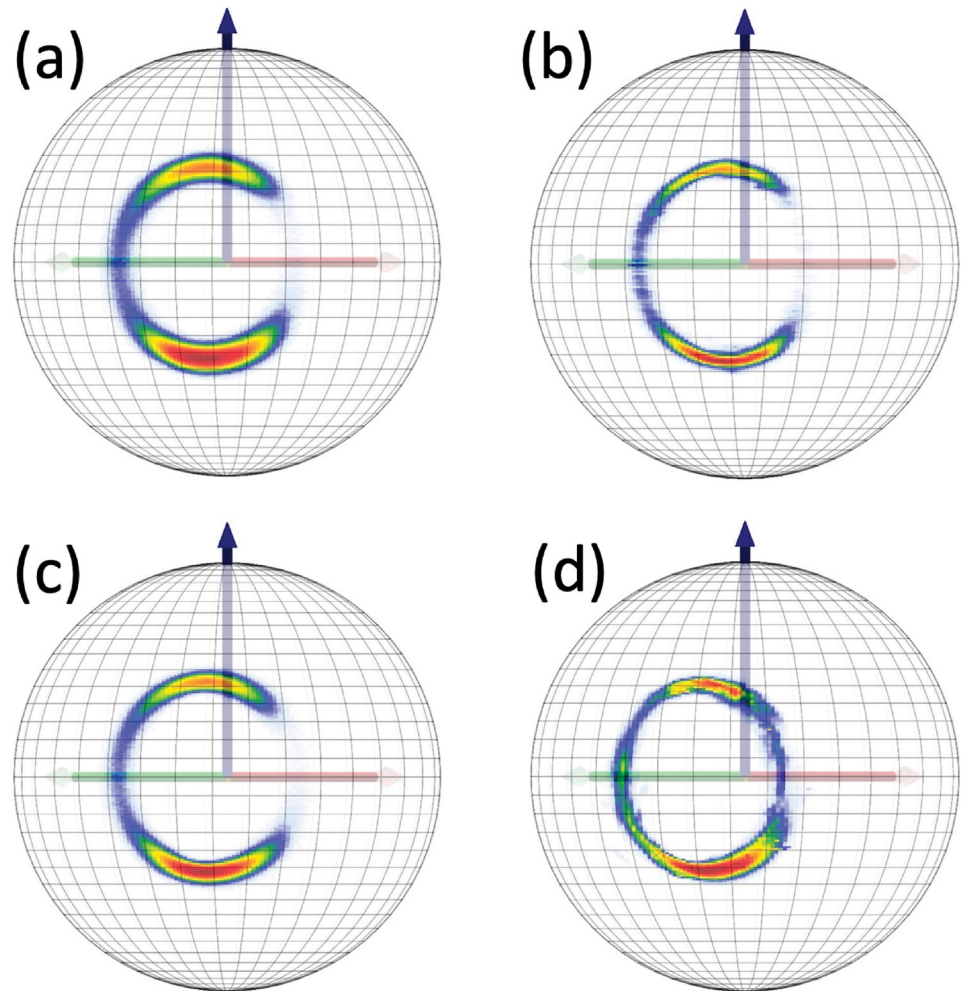


Fig 13. Statistical distribution of ww-domain $C\beta(i)$ atoms, when viewed from the position of $O(i)$ atoms in $C\alpha O$ frames. a) Original *Anton* data, b) *Statistical Method* reconstructed data c) *ScwrIPP* reconstructed data and d) *Remo* reconstructed data.

<https://doi.org/10.1371/journal.pone.0215141.g013>

for *Statistical Method* and *ScwrIPP* are again very small and strongly peaked, at around 0.04 radians; for *Remo* the deviation is also small, but visibly less so. The distance between the $C\alpha(i)$ and the $O(i)$ atom is around 2.40 Å. Thus, a peak at 0.04 radians in angular deviation corresponds to a distance around 0.05 Å. This is hardly observable with present day x-ray protein crystallography techniques.

Statistical distributions of $\overrightarrow{O(i)C\beta(i+1)}$ on a two-sphere. In Figs 15 and 16 we compare the distributions of $C\beta(i+1)$ atoms, as they are seen on the $C\alpha O$ frames at the surface of $O(i)$ centered two-spheres, for villin and ww-domain. We show the results for the original *Anton* data, and from *Statistical Method*, *ScwrIPP* and *Remo* reconstructions. We observe that the *Statistical Method* and *ScwrIPP* reconstructions are remarkably close to the original *Anton* distributions. The *Remo* reconstruction shown in the Fig 16d) displays more dispersion than the other two, shown in Fig 16b) and 16c).

Comparison of the $C\beta(i)-O(i)-C\beta(i+1)$ angle. We combine the results in Figs 12 and 15, respectively in Figs 13 and 16 into analysis of the statistical distributions of the angle

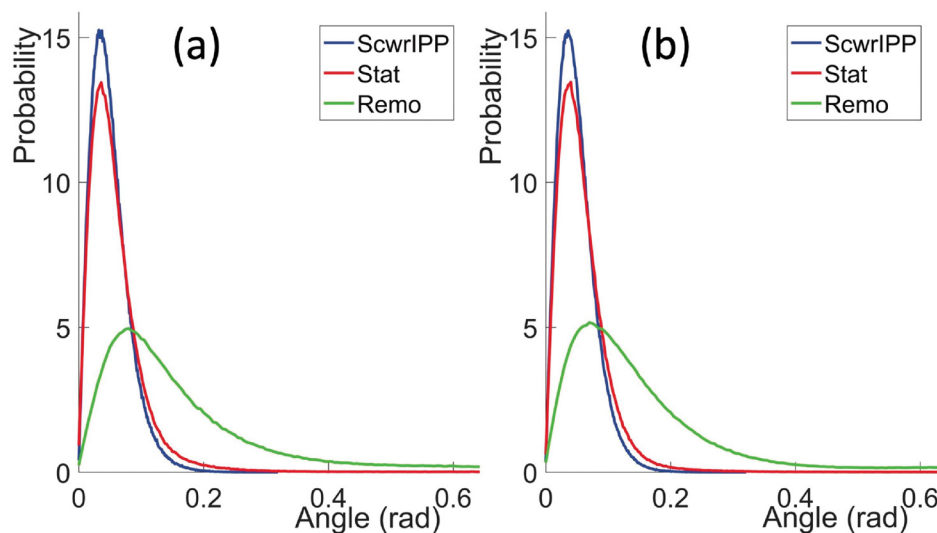


Fig 14. Statistical probability distributions for the angular difference between the *Anton* trajectory vectors $\overrightarrow{OC\beta}$ and the reconstructed vectors in the *Statistical Method*, *ScwrIPP* and *Remo*. a) villin, b) ww-domain.

<https://doi.org/10.1371/journal.pone.0215141.g014>

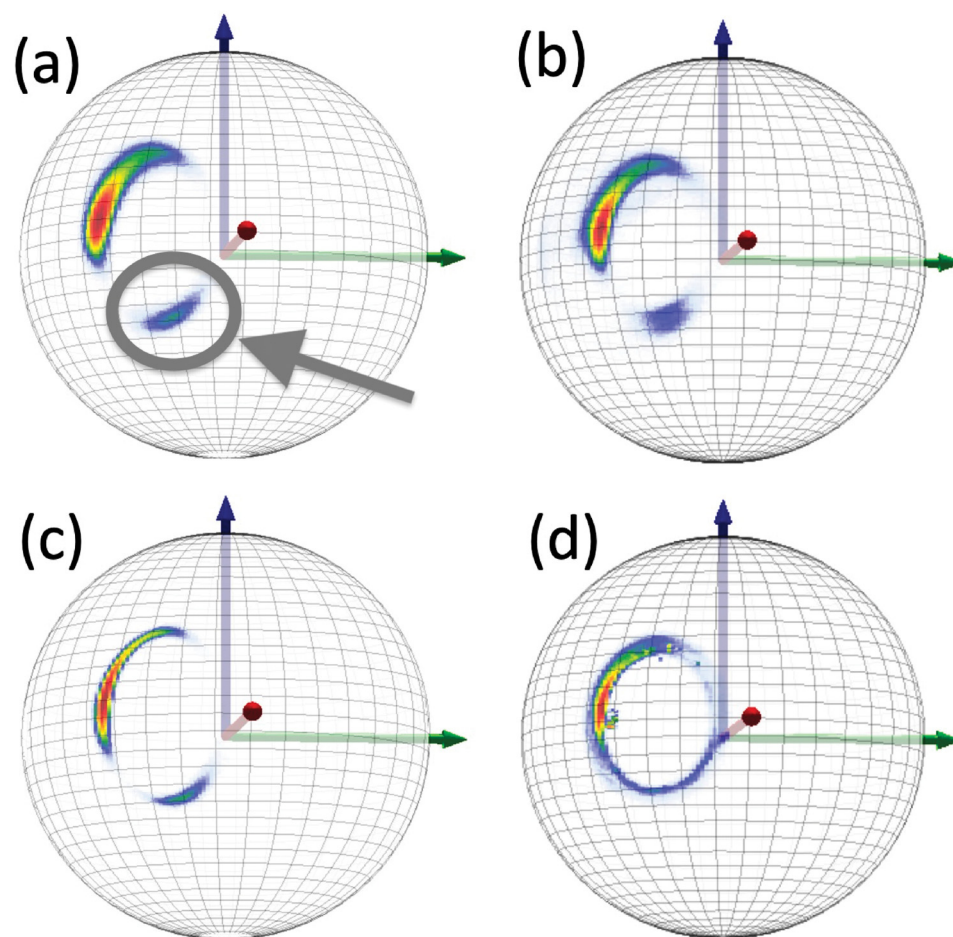


Fig 15. Statistical distribution of villin $C\beta(i + 1)$ atoms viewed from position of $O(i)$ atoms in $C\alpha O$ frames. a) Original *Anton* data b) *Statistical Method* c) *ScwrIPP* d) *Remo*. Note the isolated region that has been identified in Fig a.

<https://doi.org/10.1371/journal.pone.0215141.g015>

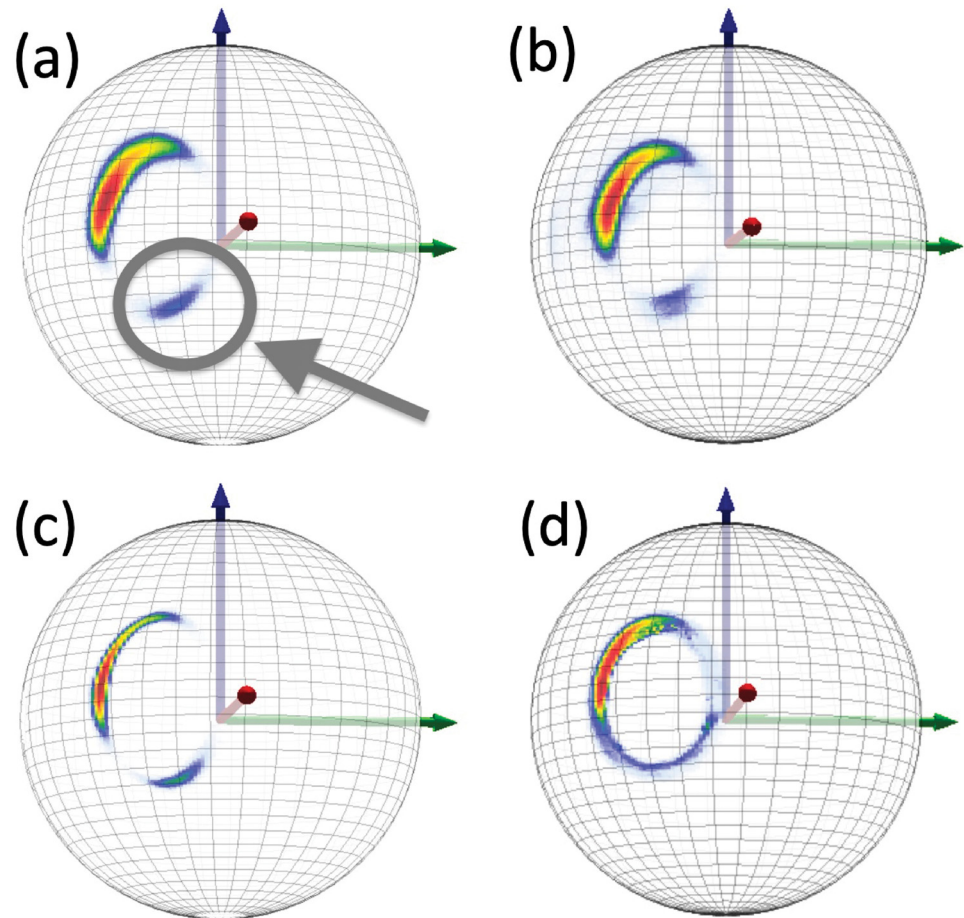


Fig 16. Statistical distribution of ww-domain $C\beta(i+1)$ atoms viewed from position of $O(i)$ atoms in $C\alpha O$ frames. a) Original Anton data b) Statistical Method c) ScwrIPP d) Remo Note the isolated region that has been identified in Fig a.

<https://doi.org/10.1371/journal.pone.0215141.g016>

$\angle C\beta(i)O(i)C\beta(i+1)$ between the directions of the vectors $\overrightarrow{O(i)C\beta(i)}$ and $\overrightarrow{O(i)C\beta(i+1)}$. This measures the quality in the reconstruction of the entire $C\alpha$ - $C\beta$ -peptide plane complex. In Fig 17 we show the results for both villin and ww-domain, in the four cases of original Anton, and Statistical Method, ScwrIPP and Remo reconstruction.

A comparison shows that the Statistical Method provides slight improvement over ScwrIPP. In the case of villin the peak in the Statistical Method is slightly closer to the original Anton peak, and in the case of ww-domain Statistical Method succeeds to reproduce the double peak profile of Anton. The difference between these two methods and Remo reconstruction is apparent, the latter displays visibly larger deviations from the original Anton data. This confirms that the knowledge of the $O(i)$ atom positions brings about clear improvement in the reconstruction.

Isolated region in $C\beta(i+1)$ distributions. In both Figs 15 and 16a), 16b) and 16c) we observe a very similar isolated region; in Fig d) this region becomes connected. In Fig 18 we show the locations of the corresponding regions on the stereographically projected two-sphere of Fig 5. The Fig 18 show that the isolated regions correspond to residues that are positioned either in the left-handed α -helical region or in a region that is the mirror image of the β -stranded region *i.e.* in both cases the torsion angle has become shifted by $\tau \sim \pi$ with respect to

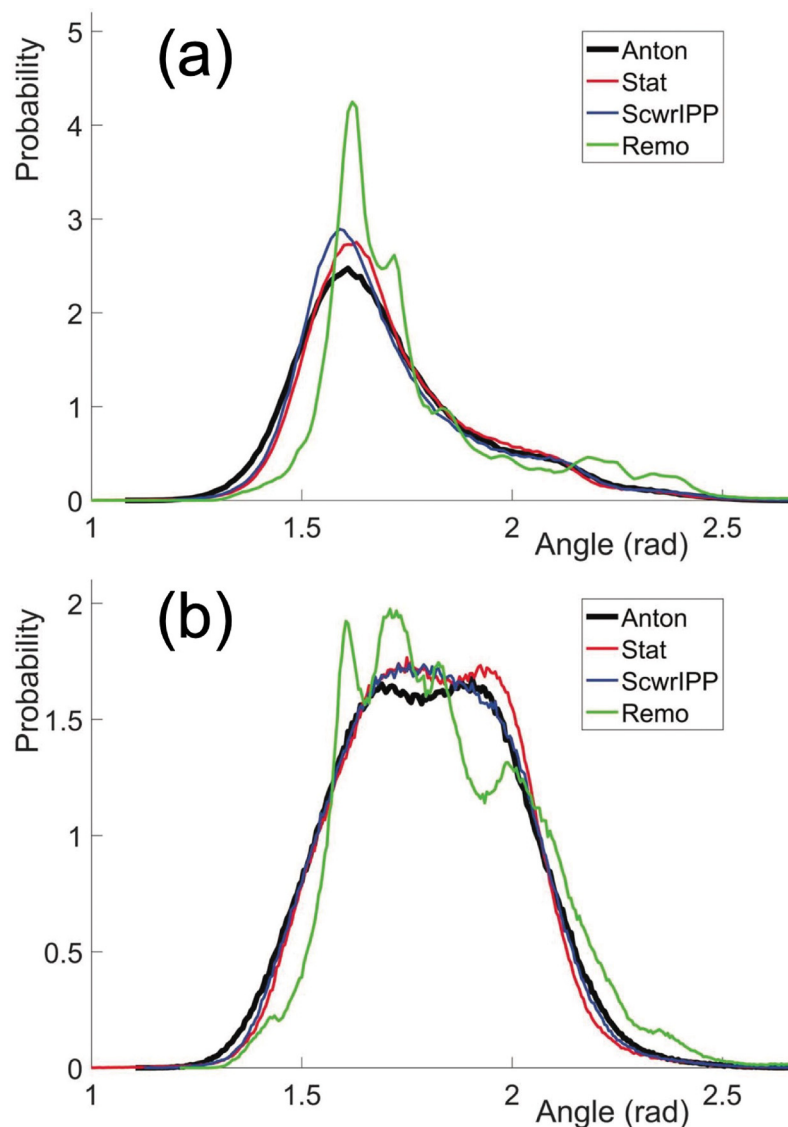


Fig 17. Statistical distribution of the angle $\angle C\beta(i)O(i)C\beta(i + 1)$ for original *Anton* data and its reconstruction using *Statistical Method*, *ScwrIPP* and *Remo*. a) for villin, b) for ww-domain.

<https://doi.org/10.1371/journal.pone.0215141.g017>

the τ -values in the conventional α -helical and β -stranded regions. Note that the π -shifted β -stranded region is very rare among the crystallographic structures shown in Fig 5 which proposes that it might correspond to a higher energy, intermediate dynamical transition state.

Discussion

We have investigated dynamical proteins using results from all-atom molecular dynamics simulation, with *Anton* supercomputer. In particular, we have inspected the very long time period *Anton* simulation trajectories of villin and ww-domain. We have found that the $C\beta$ positions along these trajectories can be accurately reconstructed, solely from the knowledge of the backbone $C\alpha$ and O atom dynamics in combination with a statistical analysis of *static* crystallographic PDB structures.

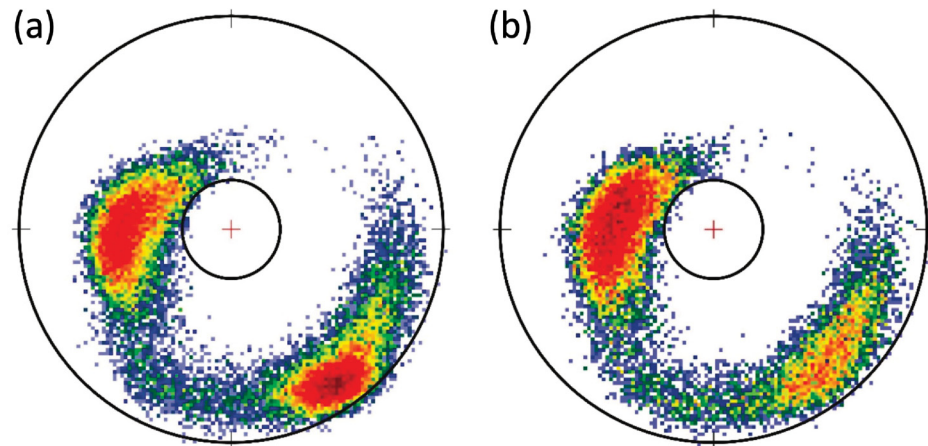


Fig 18. Statistical distribution of the (κ, τ) values that correspond to the isolated regions marked in Figs 15 and 16a. In Fig a) villin and in b) ww-domain.

<https://doi.org/10.1371/journal.pone.0215141.g018>

We have compared the results that we obtain from our *Statistical Method* reconstruction, with publicly available all-atom reconstruction programs *Remo* and *Scwrl4*. *Remo* starts from the $C\alpha$ trace, then reconstructs the peptide planes using elaborate optimization protocols. It then proceeds with *Scwrl* to add the side chain atoms for the full all-atom structure. *Scwrl4* can also be used in a stand-alone mode, and for this we have enhanced it using the knowledge of the $C\alpha$ and O atom coordinates along the *Anton* trajectories, with N, C and H in ideal peptide plane positions. Both *Remo* and *Scwrl4* need to reconstruct the entire all-atom structure, to evaluate the $C\beta$ coordinates. This provides added precision to the reconstruction, as the removal of higher level steric clashes serves as a constraint. On the other hand, the present *Statistical Method* places the $C\beta$ atoms with no regard to higher level side chain structures, or potential steric clashes between side chains and peptide planes. Only the backbone $C\alpha$ and O atoms time evolution is employed in the reconstruction. Nevertheless, the results that we get using the *Statistical Method* are a clear improvement over those obtained with *Remo*, and are fully comparable with those obtained by our enhanced *Scwrl4*. Thus, at the present level of scrutiny, side chain dynamics and in particular that of $C\beta$ atoms, appears to be fully slaved to that of the backbone $C\alpha$ and O atom dynamics. The scrutiny of the higher level side chain atoms does not seem to bring about much improvement in reconstructing the $C\beta$ dynamics.

All the three methods that we have employed are, by their construction, designed to reconstruct static low temperature crystallographic protein structures. They are not intended to model the all-atom structure of a dynamical and biologically active protein at physiological temperatures. Nevertheless, each of them succeeds in the reconstruction of the $C\beta$ dynamics in *Anton* simulations, with an impressive precision. We find this to be remarkable and very surprising. It motivates the development of coarse grained approaches to model large time scale protein dynamics, in terms of reduced coordinates that describe the dynamics of $C\alpha$, O and $C\beta$ atoms only.

Acknowledgments

We thank D.E. Shaw Research for providing us with the *Anton* data for our analysis. AJN acknowledges support from Vetenskapsrådet and Carl Trygger's Stiftelse för vetenskaplig forskning.

Author Contributions

Conceptualization: Yanzhen Hou, Xubiao Peng, Antti J. Niemi.

Formal analysis: Jiaojiao Liu, Jianfeng He, Xubiao Peng, Antti J. Niemi.

Investigation: Yanzhen Hou, Jiaojiao Liu, Antti J. Niemi.

Project administration: Antti J. Niemi.

Resources: Antti J. Niemi.

Software: Yanzhen Hou, Jiaojiao Liu, Xubiao Peng.

Supervision: Jianfeng He, Xubiao Peng, Antti J. Niemi.

Validation: Antti J. Niemi.

Visualization: Yanzhen Hou, Jiaojiao Liu, Antti J. Niemi.

Writing – original draft: Yanzhen Hou, Jianfeng He, Antti J. Niemi.

Writing – review & editing: Antti J. Niemi.

References

1. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* A47 110–119 (1991) <https://doi.org/10.1107/S0108767390010224>
2. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* 43(D1), D376–D381 (2015) <https://doi.org/10.1093/nar/gku947>
3. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247 536–540 (1995) <https://doi.org/10.1006/jmbi.1995.0159> PMID: 7723011
4. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5 725–738 (2010) <https://doi.org/10.1038/nprot.2010.5> PMID: 20360767
5. Arnold K, Bordoli L, Kopp J, Schwede T. *Bioinformatics* 22 195–201 (2006) <https://doi.org/10.1093/bioinformatics/bti770> PMID: 16301204
6. Zhang Y. Protein structure prediction: when is it useful? *Curr. Opin. Struct. Biol.* 19 145–155 (2009) <https://doi.org/10.1016/j.sbi.2009.02.005> PMID: 19327982
7. Dill K, Ozkan SB, Weikl TR, Chodera JD, Voelz VA. The protein folding problem: when will it be solved? *Curr. Op. Struct. Biol.* 17 342–346 (2007) <https://doi.org/10.1016/j.sbi.2007.06.001>
8. Scheraga HA, Khalili M, Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. *Ann. Rev. Phys. Chem.* 58 57–83 (2007) <https://doi.org/10.1146/annurev.physchem.58.032806.104614>
9. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* 116, 7898–7936 (2016) <https://doi.org/10.1021/acs.chemrev.6b00163> PMID: 27333362
10. Holm L, Sander C. Database algorithm for generating protein backbone and side-chain co-ordinates from a C α trace: application to model building and detection of co-ordinate errors *Journ. Mol. Biol.* 218 183–194 (1991) [https://doi.org/10.1016/0022-2836\(91\)90883-8](https://doi.org/10.1016/0022-2836(91)90883-8)
11. DePristo MA, de Bakker PIW, Shetty RP, Blundell TL. Discrete restraint-based protein modeling and the C α -trace problem *Prot. Sci.* 12 2032–2046 (2003)
12. Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* 50 437–450 (2003) <https://doi.org/10.1002/prot.10286> PMID: 12557186
13. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations *Journ. Comp. Chem.* 29 1460–1465 (2008)

14. Li Y, Zhang Y. REMO: A new protocol to refine full atomic protein models from C α traces by optimizing hydrogen bonding networks *Proteins* 76 665–676 (2009) <https://doi.org/10.1002/prot.22380> PMID: [19274737](https://pubmed.ncbi.nlm.nih.gov/19274737/)
15. Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4 *Proteins* 77 778–795 (2009)
16. Henzler-Wildman K, Kern D. Dynamic personalities of proteins *Nature* 450 964–972 (2007) <https://doi.org/10.1038/nature06522> PMID: [18075575](https://pubmed.ncbi.nlm.nih.gov/18075575/)
17. Frauenfelder H, Chen G, Berendzen J, Fenimore PW, Jansson H, McMahon BH, et al. A unified model of protein dynamics *PNAS* 106 5129–5134 (2009) <https://doi.org/10.1073/pnas.0900336106> PMID: [19251640](https://pubmed.ncbi.nlm.nih.gov/19251640/)
18. Bu Z, Callaway DJE. Chapter 5—Proteins MOVE! Protein dynamics and long-range allostery in cell signaling *Adv. Protein Chem. Struct. Biol.* 83 163–221 (2011)
19. Khodadadi S, Sokolov AP. Protein dynamics: from rattling in a cage to structural relaxation *Soft Matter* 11 4984–4998 (2015) <https://doi.org/10.1039/c5sm00636h> PMID: [26027652](https://pubmed.ncbi.nlm.nih.gov/26027652/)
20. Lindorff-Larsen K, Piana S., Dror RO, Shaw DE. How fast-folding proteins fold. *Science* 334 517–520 (2011) <https://doi.org/10.1126/science.1208351> PMID: [22034434](https://pubmed.ncbi.nlm.nih.gov/22034434/)
21. Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100 L47–L49 (2011) <https://doi.org/10.1016/j.bpj.2011.03.051> PMID: [21539772](https://pubmed.ncbi.nlm.nih.gov/21539772/)
22. Liu J, Dai J, He J, Peng X, Niemi AJ. Can all-atom protein dynamics be reconstructed from the knowledge of C α time evolution? *arXiv:1901.06864*
23. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. (2000) The Protein Data Bank *Nucl. Acids Res.* 28 235–242. (2000) <https://doi.org/10.1093/nar/28.1.235> PMID: [10592235](https://pubmed.ncbi.nlm.nih.gov/10592235/)
24. Danielsson UH, Lundgren M, Niemi AJ. Gauge field theory of chirally folded homopolymers with applications to folded proteins *Phys. Rev.* E82 021910 (2010)
25. Molkenhuth N, Hu S, Niemi AJ. Discrete nonlinear Schrödinger Equation and polygonal solitons with applications to collapsed proteins *Phys. Rev. Lett.* 106 078102–078105 (2011) <https://doi.org/10.1103/PhysRevLett.106.078102> PMID: [21405545](https://pubmed.ncbi.nlm.nih.gov/21405545/)
26. Niemi AJ. in Chamon C., Goerbig M.O., Moessner R., Cugliandolo L.F. (Eds.) *Topological Aspects of Condensed Matter Physics: Lecture Notes of the Les Houches Summer School* Vol. 103 (Oxford University Press, Oxford, 2017)
27. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field *J. Chem. Phys.* 115 2323 (2001) <https://doi.org/10.1063/1.1383989>
28. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Oldziej S, Wachucik K, et al. 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.* B111 260–285 (2007). <https://doi.org/10.1021/jp065380a>
29. Liwo A, Czaplewski C, Oldziej S, Rojas AV, Kazmierkiewicz R, Makowski M, et al. in *Coarse-Graining of Condensed Phase and Biomolecular Systems* Voth G. ed. (CRC Press 2008)
30. Sasai M, Wolynes PG. Molecular theory of associative memory Hamiltonian models of protein folding *Phys. Rev. Lett.* 65 2740 (1990) <https://doi.org/10.1103/PhysRevLett.65.2740> PMID: [10042680](https://pubmed.ncbi.nlm.nih.gov/10042680/)
31. Hardin C, Eastwood MP, Prentis MC, Luthey-Schulten Z, Wolynes PG. Associative memory Hamiltonians for structure prediction without homology: α/β proteins *Proc. Nat. Acad. (USA)* 100 1679–1684 (2003) <https://doi.org/10.1073/pnas.252753899>
32. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing *J. Phys. Chem.* 116 8494–8503 (2012) <https://doi.org/10.1021/jp212541y>
33. Janin J, Wodak S, Levitt M, Maigret B. Conformation of amino acid side-chains in proteins *J. Mol. Biol.* 125 357–386 (1978) [https://doi.org/10.1016/0022-2836\(78\)90408-4](https://doi.org/10.1016/0022-2836(78)90408-4) PMID: [731698](https://pubmed.ncbi.nlm.nih.gov/731698/)
34. Lovell SC, Word J, Richardson JS, Richardson DC The penultimate rotamer library. *Proteins* 40 389–408 (2000)
35. Schrauber H, Eisenhaber F, Argos P Rotamers: to be or not to be?: an analysis of amino acid side chain conformations in globular. *J. Mol. Biol.* 230 592–612 (1993) <https://doi.org/10.1006/jmbi.1993.1172> PMID: [8464066](https://pubmed.ncbi.nlm.nih.gov/8464066/)
36. Dunbrack RL Jr, Karplus M. Backbone-dependent Rotamer library for proteins application to side chain prediction. *J. Mol. Biol.* 230: 543–574. (1993) <https://doi.org/10.1006/jmbi.1993.1170> PMID: [8464064](https://pubmed.ncbi.nlm.nih.gov/8464064/)

37. Shapovalov MS, Dunbrack RL Jr.. A smoothed backbone-dependent Rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* 19 844–858 (2011) <https://doi.org/10.1016/j.str.2011.03.019> PMID: 21645855
38. Ramachandran G, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations, *Journ. Mol. Biol.* 7 95–99 (1963) [https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)
39. Ramakrishnan C, Ramachandran G. Stereochemical Criteria for Polypeptide and Protein Chain Conformations: II. Allowed Conformations for a Pair of Peptide Units, *Biophys. J.* 5 909–933 (1965) [https://doi.org/10.1016/S0006-3495\(65\)86759-5](https://doi.org/10.1016/S0006-3495(65)86759-5) PMID: 5884016
40. Hinsen K, Hu S, Kneller GR, Niemi AJ. A comparison of reduced coordinate sets for describing protein structure *J. Chem. Phys.* 139 124115 (2013) <https://doi.org/10.1063/1.4821598> PMID: 24089758
41. Hu S, Lundgren M, Niemi AJ Discrete Frenet frame, inflection point solitons, and curve visualization with applications to folded proteins *Phys. Rev.* E83 061908 (2011)
42. Peng X, Chenani A, Hu S, Zhou Y, Niemi AJ, A three dimensional visualisation approach to protein heavy-atom structure reconstruction *BMC Struct. Biol.* 14:27 (2014) <https://doi.org/10.1186/s12900-014-0027-8> PMID: 25551190
43. Bishop RL, *Amer. Math. Monthly* 82 246–251 (1974)