DOI: 10.1111/cgf.12894
Eurographics Conference on Visualization (EuroVis) 2016
K.-L. Ma, G. Santucci, and J. van Wijk
(Guest Editors)

Volume 35 (2016), Number 3

Evaluating Viewpoint Entropy for Ribbon Representation of Protein Structure

J. Heinrich[1], J. Vuong[1], C. J. Hammang[2], A. Wu[3], M. Rittenbruch[4], J. Hogan[4], M. Brereton[4], S. I. O'Donoghue[1,2]

[1]CSIRO, Australia
[2]The Garvan Insitute of Medical Research, Australia
[3]The University of Sydney, Australia
[4]Queensland University of Technology, Australia

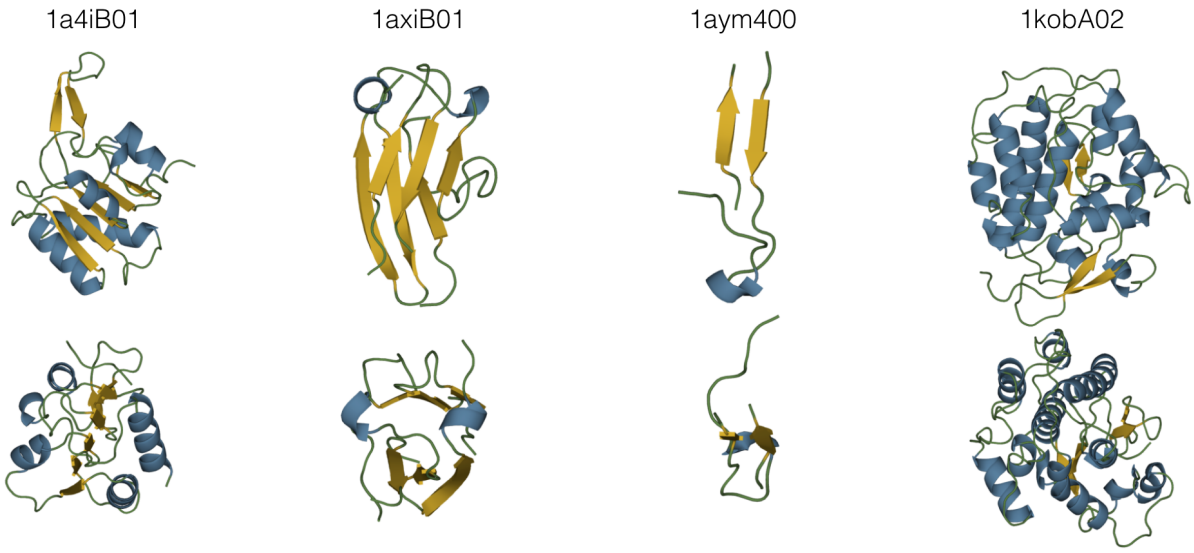

Figure 1: High (top) and low entropy (bottom) viewpoints of ribbon representations of 4 protein structural domains used as test structures in our evaluation. Every domain is a member of 1 of 4 classes in the top level of the CATH hierarchy (from left to right: *alpha and beta*, *mainly beta*, *few secondary structures*, and *mainly alpha*).

Abstract

While many measures of viewpoint goodness have been proposed in computer graphics, none have been evaluated for ribbon representations of protein secondary structure. To fill this gap, we conducted a user study on Amazon's Mechanical Turk platform, collecting human viewpoint preferences from 65 participants for 4 representative superfamilies of protein domains. In particular, we evaluated viewpoint entropy, which was previously shown to be a good predictor for human viewpoint preference of other, mostly non-abstract objects. In a second study, we asked 7 experts in molecular biology to find the best viewpoint of the same protein domains and compared their choices with viewpoint entropy.

Our results indicate that viewpoint entropy overall is a significant predictor of human viewpoint preference for ribbon representations of protein secondary structure. However, the accuracy depends on the type and composition of the structure: while most participants agree on good viewpoints for structures with mainly beta sheets, viewpoint preference varies considerably for complex arrangements of alpha helices. Finally, experts tend to choose viewpoints of both low and high viewpoint entropy to emphasize different aspects of the respective structure.

1. Introduction

Structural biology has produced over 100,000 experimentally-determined macromolecular structures at near atomic resolution. These data give invaluable insight into the molecular machinery of life; however, using these data effectively is challenging, even for experts. A very common task involves choosing optimal 3D viewpoints for these complex structures: This task is done by student and expert users alike, and for a wide range of tasks, for example, in creating figures for scientific publication, or animations for public communication [JH14]. Currently, this task is almost always done manually; here, we evaluate viewpoint entropy as a strategy for selecting viewpoints automatically.

Protein structures are typically defined by their amino acid sequence (primary structure), the presence of local features known as *alpha helices* and *beta sheets* (secondary structure), and the overall spatial layout (tertiary structure). Ranges of residues without secondary structure are often referred to as *coils*. There are a multitude of visualization techniques available for macromolecules (see [KKL*15] for a recent overview), each conveying different aspects of the 3D structure of a molecule. Common examples are ball-and-stick representations of individual atoms and bonds, surface representations, or secondary structure representations such as ribbons. In this work, we focus on the latter, as secondary structure visualizations are frequently used to obtain the overall shape of a macromolecule, and many web-resources such as the Protein Data Bank (PDB) [BWF*00] use ribbon representations as the default view. In ribbon representations, the type of secondary structure of a residue is typically mapped to both geometry and color. In this article, we use a blue color for helices, yellow for sheets, and green for coils (see Figure 1 for an example).

While structural biologists typically use sophisticated software packages for visualizations of macromolecules [OGF*10], the advent of web-based molecular graphics [MKK*14] now allows anyone interested in molecular structures (including the general public, teachers and students, molecular biologist, or bioinformaticians) to access and visually explore any protein deposited in public databases such as the PDB or Aquaria [OSK*15]. Considering the complexity of interacting with 3D objects in general [KHC*08] and with complex macromolecules in particular, we anticipate that pre-determined viewpoints would be of great benefit to experts and naïve users alike. Providing sensible default views as a starting point for interactive exploration or computing camera paths for periodic orbiting are example applications of automatic viewpoint detection.

We focus on the evaluation of one particular measure to quantify the goodness of a viewpoint: Viewpoint entropy [VFSH01] is an established, appearance-based measurement of viewpoint goodness based on information theory. It can be applied to any three-dimensional scene with known geometry and has been proposed to be useful with molecular structures as well [VFSL02, DCMP10]. However, no formal evaluation has been conducted yet to show the extent to which viewpoint entropy matches human preference of viewpoints for ribbon representations of protein structures.

Hence, we ran 2 studies to compare human preference with viewpoint entropy for 4 representative molecular structures: first, we recruited individuals without specific knowledge about proteins via Amazon's Mechanical Turk (MTurk) platform to build a single-attribute model of viewpoint preference. To train this model, we carefully selected a set of macromolecular structures that represents a large fraction of structurally similar domains, which are functional sub-units of proteins. We anticipate this set of structures will serve future investigations as benchmark models. In an additional study, we collected preferred viewpoints of the same set of structures from experts in molecular biology, and compared those to the data of the non-experts. The results of both groups suggest that viewpoint entropy is significantly better than random at predicting human preference overall, with varying performance.

In summary, the contributions of this paper are (i) a set of benchmark domains that represent 7,819 entries in the PDB, (ii) a quantitative evaluation of viewpoint entropy for non-experts, and (iii) a qualitative evaluation of viewpoint preference for experts to guide future research.

2. Related Work

Viewpoint goodness has been investigated extensively in computer graphics [SLF*11] and volume visualization [MNTP07, TFTN05]. While much work has been done to find models of viewpoint goodness for real-world objects, only little is known as to which extent these models match the human preference of complex shapes such as molecules. Here, we focus on a single type of object, namely ribbon representations of protein secondary structure. We therefore report related work relevant for the evaluation of viewpoint goodness for macromolecules in this section and refer the reader to the work of Secord et al. [SLF*11] for an overview of the methods that have been developed for other types of objects.

Viewpoint entropy was introduced by Vázquez et al. [VFSH01] for general scenes and has been applied to image-based rendering [VFSH03] and macromolecules [VFSL02, VFSL06] in subsequent work. While Vázquez et al. applied viewpoint entropy to ball-and-stick representations of molecules, we use our notion of features for the underlying probability distribution and provide a quantitative analysis of the performance of viewpoint entropy as a predictor of viewpoint goodness for ribbon representations.

Doulamis et al. [DCMP10] proposed a personalized viewpoint selection for atom-based representations of macromolecules and trained a non-linear classifier with input from domain experts. While their approach is similar to ours in that a model is fitted to human preference data, our model is aimed at being specific for ribbon representations, while performing well for non-experts in structures. Furthermore, Doulamis et al. did not provide an evaluation of their approach.

Takahashi et al. [TFTN05] proposed viewpoint entropy to locate feature-driven viewpoints in a volume-rendering context. While we use a similar approach to compute viewpoint entropy for features of an object, our definition of features is different and can not be applied to volumes. Also, Takahashi et al. did not conduct a quantitative evaluation of their approach.

Ji and Shen [JS06] based their selection of single and multiple viewpoints on analyzing the opacity, color and curvature of the viewed object. While they get promising results on using this technique on different types of objects (real-world and abstract), the

parameters they used to select a good viewpoint would not be suitable to apply to protein structures due to their unique features and complexity.

Some molecular visualization packages [Sch10] implement the "best" view by providing a simple principal-components view, which is typically computed from the 3D coordinates of C_α atoms. However, such views do not take into account the visibility of residues and thus may hide important features of a structure. In contrast to such object-space methods, image-based measures of viewpoint goodness evaluate the image of a molecular structure after projecting it to 2D. While such an approach loses the 3D context of the structure, such appearance-based methods have the advantage of measuring exactly what the viewer perceives.

Our methodology is based on the meta-study conducted by Secord et al. [SLF*11]. In their work, a set of 14 attributes measuring viewpoint goodness are investigated towards their contribution in predicting human viewpoint preference. While Secord et al. give recommendations for which attributes perform well when used as a single predictor and in combination with others, their model was not trained with nor tested on protein structures. In contrast, the training data set of Secord et al. consisted of 14 real-world objects and only two abstract objects. Considering the vast amount of possible 3D shapes, the number of known structures [PaHS*15] is a rather small subset of these shapes, which eventually allows us to design specialized, but well-performing models of viewpoint goodness. Here, we investigate the feasibility of measuring viewpoint goodness for these complex shapes and hence focus on evaluating a single, image-based attribute only.

We collected human preference data using the MTurk platform. Over the last years, MTurk has become increasingly popular for perception studies in a variety of disciplines; examples include computer graphics [SLF*11], human-computer interaction [HB10], and information visualization [CSR*14, DBH14]. MTurk facilitates online user studies by providing 'requesters' with an interface to design human-intelligence-tasks (HITs), which can then be completed by individuals signed up as 'workers'. For every HIT completed and approved, Amazon transfers a pre-defined and prepaid micro-payment to the worker. While this service supports researchers in recruiting participants and collecting results in a timely manner, it does not allow for the same amount of monitoring as experiments conducted in a controlled environment.

3. Viewpoint Entropy

In this work, we evaluate viewpoints of protein secondary structure representations. We hypothesize that good viewpoints of such representations maximize both the number and the area of *visible* distinct secondary structure elements, i.e. we seek to maximize the information conveyed by alpha helices, beta sheets and coils. Below, we summarize how this can be achieved using viewpoint entropy.

The basic idea behind viewpoint entropy is to maximize the information that can be perceived from a given scene across all candidate viewpoints. Viewpoint entropy is based on the Shannon Entropy of a discrete random variable X:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i,$$

where $p_i = P(X = a_i)$ is the probability of the event a_i. By using the relative area of projected faces of a scene as probability distribution, Vázquez et al. [VFSL02] define the viewpoint entropy of a scene S from a viewpoint v as

$$I_P(v) = -\sum_{i=1}^{N_b} \frac{A_i}{4\pi} \log_2 \frac{A_i}{4\pi},$$

where A_i is the projected area of face i, N_b is the number of faces and 4π is the solid angle of the sphere (of possible camera positions). Using orthogonal projection and assuming that we are operating on the projected image of the scene with N pixels, this can further be simplified to

$$I_F(v) = -\sum_{i=1}^{N_f} \frac{Np_i}{N} \log_2 \frac{Np_i}{N},$$

where N_f is the number of objects or *features* and Np_i is the number of projected pixels of feature i, and $i = 0$ denotes the background of the image. For convenience, we use the normalized viewpoint entropy as proposed by Takahashi et al. [TFTN05]:

$$normaI(v) = -\frac{1}{\log_2(N_f + 1)} I_F(v). \quad (1)$$

In contrast to previous work where viewpoint entropy was applied to faces [VFSH01] or atoms and bonds [VFSL02], we use *residue features* instead for the underlying probability distribution. We define a residue feature as a set of one or more consecutive amino acids carrying a common semantic meaning, and count all pixels of a feature in the projected image to determine Np_i. In this work, we focus on secondary structure as a feature, such that N_f is the number of distinct secondary structure elements and the respective Np_i are the number of pixels per distinct alpha helix, beta sheet or coil.

To compute $I(v)$, we assign a unique color to every distinct secondary structure element of a ribbon representation and then use an image-based pipeline to compute the number of pixels Np_i for every distinct color in the resulting image, rendered from viewpoint v.

4. Evaluation

In order to evaluate viewpoint entropy for secondary structure features, we conducted a study on Amazon's Mechanical Turk platform. We deliberately decided to recruit non-experts in structural biology to ensure that our results are valid for the general public. This is necessary as users of molecular visualization today cannot only be found in all areas of the life sciences (e.g. medical practitioners, biomedical animators, molecular biologists, and bioinformaticians), but also in the general public (e.g. in outreach and education). We therefore did not require participants to have any knowledge about protein structures and thus did not ask to perform a qualification test to participate in our study, except for collecting user agreement on the participation information.

We were interested in investigating the correlation between viewpoint entropy and the observed goodness of a viewpoint for a representative set of structures and viewpoints. This can be achieved by collecting human viewpoint preference for every structure (as the observed goodness) and by comparing it to the goodness as measured by our model. In line with the work of Secord et al. [SLF*11], we employed a two-alternative forced choice (2AFC) methodology to collect human preference data. This protocol has been shown to be effective with respect to the time and effort required by the subjects, while delivering good results in terms of statistical power [MTM12]. In this type of experiment, participants are presented with two stimuli and are asked to make a choice between them. As this task is straight-forward, 2AFC experiments were found to produce more accurate results than alternative rating models [MTM12].

In our study, a trial consisted of selecting one image out of two in a side-by-side layout, each of which showed a ribbon representation of the same structure but from a different viewpoint. Inspired by previous studies on viewpoint goodness [BTB99, SLF*11], we instructed participants to choose the view that "best reveals the shape of the object" without imposing a time-limit. Structures were colored according to the Aquaria [OSK*15] coloring scheme.

To avoid participant duplication, trials were submitted to MTurk in batches of HITs (one batch per structure). Hence, participants were allowed to accept each trial exactly once and were able to choose voluntarily how many trials they want to complete.

4.1. Pilot Study

In order to test our setup and estimate the time a participant requires to complete a trial, we conducted a full 2AFC experiment as a pilot study on MTurk using a within-subjects design: for 10 viewpoints of the domain with CATH ID `2bg9B01` of type *mainly beta* (see Figure 2), we showed every combination of dissimilar pairs of viewpoints twice to 13 naïve participants: After presenting 90 pairs in random order, we shuffled the order of pairs, swapped sides (such that every image appears exactly once on the left and once on the right side), and showed all pairs again to the same participant. With an initial assumption that every choice takes an approximately of 10 seconds to complete, we set the reward per trial to USD 0.02, based on the US minimum wage of USD 8.00/hour.

As every pair of images was shown to a participant twice, we can compute the *consistency* at which one viewpoint was preferred over the other for every structure, pair of viewpoints and participant. If the same image was picked twice by a participant, the consistency for that participant and pair is $c = 1$, else it is $c = 0$. We can then use the number of consistent votes to compute the probability of obtaining such a result purely by chance using a binomial test and filter for participants that appear to be choosing images randomly at the desired level of confidence (e.g. $\alpha = 0.05$). This strategy allowed us to filter for careless subjects while keeping those cases where participants could not make up their mind as the two views appeared to reveal the shape of the structure equally well.

From 13 participants, we identified 3 as careless subjects ($\alpha = 0.05$) and obtained votes from 10 participants in 830 trials (note that not all participants completed all trials). With every view being compared to every other view, we can then use the number of votes per view v_i divided by the total number of votes for all pairs containing v_i as an approximation of the probability $P(v_i)$ that view v_i is preferred over any other view, and rank all viewpoints accordingly from *most preferred* to *least preferred*. Figure 2 (left) compares these ranks with a ranking from highest to lowest viewpoint entropy. Overall, there is a clear positive correlation between ranks, and 4 out of 10 views are on the same rank (including ranks number 1 and 10). Hence, for this structure, naïve observers prefer views of high viewpoint entropy over low viewpoint entropy views. The plot on the right of Figure 2 further illustrates the relation of viewpoint entropy $I(v_i)$ with $P(v_i)$. The best-fit logistic regression curve emphasizes the expected sigmoid shape of the correlation [SLF*11].

For this structure and the set of 10 viewpoints, we obtained an average consistency of $\approx 81\%$, indicating that for 81% of pairwise comparisons, a winner could be determined. The median completion time was 10 seconds as anticipated.

4.2. Main Study

After finishing the pilot study, we conducted a larger study on MTurk using 4 *benchmark structures*, 100 viewpoints and 104 participants. In this section, we describe our approach to sample these structures and viewpoints, respectively.

4.2.1. Sampling Structures

As our aim is to evaluate the performance of viewpoint entropy to predict good viewpoints for protein secondary structure, we need to find a set of benchmark structures that are as diverse as possible with respect to their composition of secondary structure elements and at the same time represent as many proteins as possible.

Compared to the huge number of real-world shapes, the pool of available protein structures is rather small: The protein databank (PDB) [BWF*00] currently holds only slightly more than 100,000 entries, many of which consist of arrangements of *domains*, which form the building blocks of molecular evolution to create proteins with different functions. Due to the complexity of molecular structures, our overall strategy is to test the feasibility of viewpoint selection algorithms on a small set of the most common domain architectures in this work, and leave the investigation of more complex, composite models for future work.

The CATH database [SCD*13] provides a hierarchical classification of 26 million domains into 2,738 superfamilies, 1,375 folds, 40 architectures and 4 classes, based on the (i) secondary structure composition, (ii) overall shape, (iii) connectivity and (iv) homology of secondary structure elements. For this study, we were particularly interested in the first level, as it summarizes the protein structural universe into 4 high-level classes with similar secondary structure compositions: (i) *mainly alpha*, (ii) *mainly beta*, (iii) *alpha and beta*, and (iv) *few secondary structures*.

For every class, we used the first domain of the superfamily (of domains) that represents the largest fraction of domains in the respective class. For example, the domain with ID `1a4iB01` is a member of the superfamily with ID `3.40.50.720` which represents a total of 6,092 domains of similar shape in class 3 (alpha and

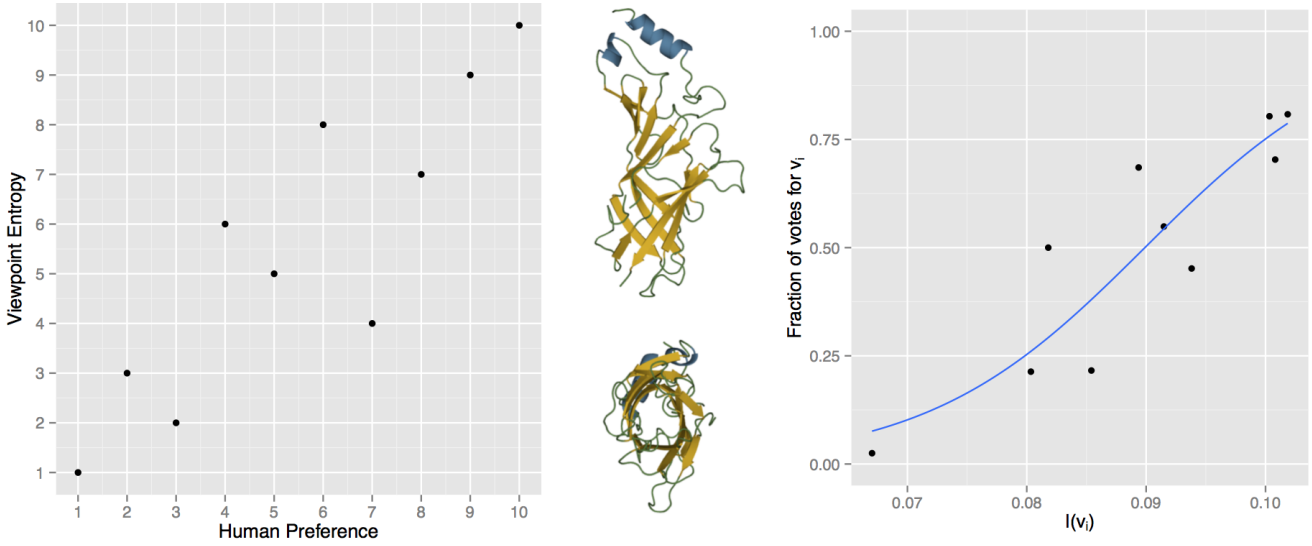

Figure 2: Ranking of viewpoints by human preference (naïve) and by viewpoint entropy (left) for the domain with CATH ID `2bg9B01` of class *mainly beta*. Overall, a clear positive correlation emerges from the plot. In particular, the most and the least preferred views agree with those of highest (center, top) and lowest viewpoint entropy (center, bottom). The correlation of viewpoint entropy $I(v_i)$ with the fraction of times v_i was preferred over any other view (right) reveals a sigmoid shape. The best-fit logistic regression curve is shown in blue.

beta). No other superfamily represents more domains in the same class.

CATH id	Class	#*SS*
`1aym400`	*few secondary structures*	6
`1axiB01`	*mainly beta*	17
`1a4iB01`	*alpha & beta*	25
`1kobA02`	*mainly alpha*	34

Table 1: The 4 domains used for evaluation in this work, sorted by increasing complexity (in terms of #*SS*, the number of distinct secondary structure elements [#*SS* equals $N_f - 1$ to exclude the background feature]).

Using this strategy, we chose the domains with CATH IDs `1a4iB01` (*alpha and beta*), `1axiB01` (*mainly beta*), `1kobA02` (*mainly alpha*) and `1aym400` (*few secondary structures*), see Table 1 and Figure 1. Altogether, the respective superfamilies represent 23,226 domains and 7,819 unique entries from the PDB.

4.2.2. Sampling Viewpoints

2AFC experiments come at the cost of requiring many trials: Ideally, every stimulus is compared to every other, such that on the order of N^2 trials are required for N conditions (here: viewpoints). While this number can be further reduced to $N \log_2 N$ using sorting algorithms [SF01], it quickly becomes impractical for our purpose, as N needs to be large in order to get a good coverage of the viewing sphere. In addition to that, the number of trials per subject should be limited to avoid fatigue and frustration.

In alignment with previous work on real-world objects [SLF*11] and to combine the simplicity of the task (which is particularly important for non-experts in protein structures) with a good coverage of viewpoints, we resorted to sampling random pairs of views instead of conducting an all-versus-all comparison as in Section 4.1. Ideally, we would like to sample the sphere of possible camera positions randomly, but uniformly spaced. To this end, we sampled 100 viewpoints using Hammersley sequences [WLH97], as such low-discrepancy sequences have the property to achieve a more uniformly spaced distribution of points than pseudorandom numbers. To obtain pairs of views, we randomly assigned 2 viewpoints to a pair without replacement, thus obtaining a total of 50 pairs of viewpoints per structure. The caveat of this approach is that every viewpoint is compared to a single other viewpoint, such that a ranking of viewpoint goodness as obtained in the pilot study cannot be directly extracted from the resulting data. Instead, distances in image attributes are used to model the odds of view v_i being preferred over view v_j for any given pair i, j.

After sampling viewpoints, we determined the orientation of the camera as follows: First, we computed the principal components of the spatial coordinates of all backbone C_α atoms in the structure. As a result, we obtained 3 linearly independent vectors, where the first indicates the direction of the largest possible variability, the second indicates the largest variability in an orthogonal direction to the first, and so on. Of these principal components, we then used the first to define the *up*-vector of the structure in object space. After rotating the camera to a viewpoint, we then aligned the camera *up*-vector to the model *up*-vector before creating the final image.

As in the pilot study, we showed every pair of views twice to every participant. This resulted in a maximum of 100 trials available for every participant and structure.

Model	σ	Standard Error	z-value	Fitness	p-value	Average Consistency
1aym400	0.98	0.22	4.55	0.64	< 0.01	74%
1axiB01	0.96	0.2	4.86	0.88	< 0.01	73%
1a4iB01	0.67	0.18	3.63	0.71	< 0.01	66%
1kobA02	0.44	0.16	2.82	0.60	< 0.01	63%
Full	0.77	0.01	59.44	0.71	< 0.01	69%

Table 2: Coefficients (σ), standard error, z-value, relative fitness, and p-values for the fitted Bradley-Terry model for the full data set and individual structures. The p-value was computed on a χ^2 distribution with degrees of freedom equal to the difference in degrees of freedom of the null model and the test model. All models reduce deviance significantly ($p < 0.01$) from the null model, indicating that viewpoint entropy is a significant predictor for the preference of viewpoints for naïve users and across all structures.

5. Results

After filtering for careless subjects using the same approach as for the pilot study, we obtained a total of 10,382 choices expressed by 65 participants. Thus, we had to discard 1,618 trials from 39 subjects, which constitutes approximately 13% of all votes.

On average, each of these participants completed approximately 160 trials, indicating that most subjects completed the full set of image pairs only for a single structure. While the mean completion time per trial was 46.68 seconds, the median was only 7 seconds. Hence some participants may have had a long break during a trial, but most trials were completed in less than 10 seconds. The difference in average consistency is very striking: Here, the fraction of pairs for which participants were indecisive increases with the *complexity* of the structure: For the most complex structure 1kobA02 with the largest number of secondary structure elements, consistency was as low as 63%, as opposed to an average consistency of 74% for 1aym400.

5.1. Modelling Viewpoint Preference

In order to evaluate our data, we need a model that predicts the probability of a viewpoint v_1 to be chosen over any other viewpoint v_2, given the respective viewpoint entropies $I(v_1)$ and $I(v_2)$. The Bradley-Terry model [BT52] derives the probability $P(v_i, v_j)$ that a participant chooses v_i over v_j from the log-odds ratio of their viewpoint entropy:

$$logit(P(v_i, v_j)) = I(v_i) - I(v_j), \tag{2}$$

where $I(v_i) = log(o_i)$ and $\frac{o_i}{o_j}$ are the odds of v_i 'beating' v_j. Then, $P(v_i, v_j)$ can be expressed using the inverse logit function as:

$$P(v_i, v_j) = \frac{1}{1 + e^{-\sigma(I(v_i) - I(v_j))}}. \tag{3}$$

This model can now be fitted to our observed data using maximum-likelihood estimation. In this approach, the likelihood that the observed data can be explained by the model is maximized. We fitted one model each to the individual domains and to the results of all assignments individually. The results are listed in Table 2 along with the standard error and z-value. For our data, all coefficients (σ) were found to be significantly different from 0 ($p < 0.01$). As the response variable in the Bradley-Terry model is expressed as log-odds, coefficients can be interpreted as the logarithm of the increase of odds for every increase in the difference of viewpoint entropy. For example, an increase of 1 unit in the difference in viewpoint entropy for 2 given views v_i and v_j in the full model, the odds of v_i being preferred over v_j increase by $e^{0.77} \approx 2.16$ times.

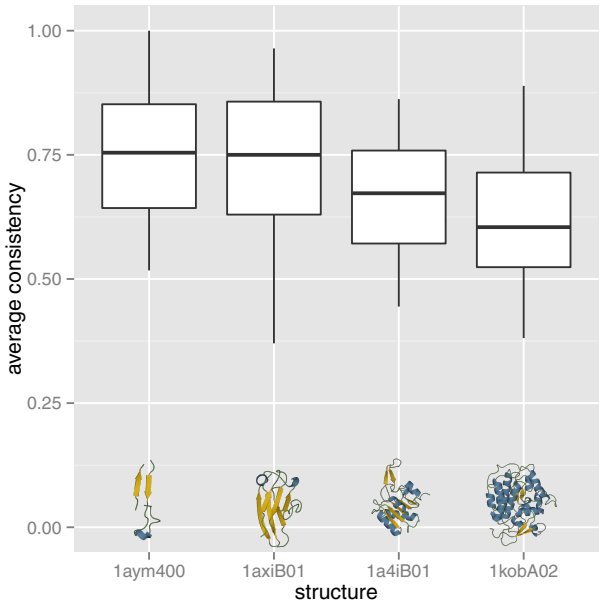

Figure 3: The consistency at which participants picked a viewpoint was lowest for 1kobA02, indicating that participants were most indecisive for that structure.

In logistic regression, the quality of a model is typically expressed as the *deviance* to a fully saturated model, i.e. a model with perfect fit. The deviance is computed from the likelihood of the data to be explained by a model. A good model, i.e. a model that does not deviate substantially from the saturated model, results in a small value for residual deviance. The *null-model*, in contrast, is a model with only the intercept, and the null-deviance is the difference of the null-model to the saturated model. As a saturated model is typically not available, the significance of a model with at least one predictor can instead be tested with a χ^2_{n-m}-test on the difference between the residual deviance and the null deviance. Significant values indicate good model fit and little unexplained variance. However, as the absolute value of deviance is difficult to

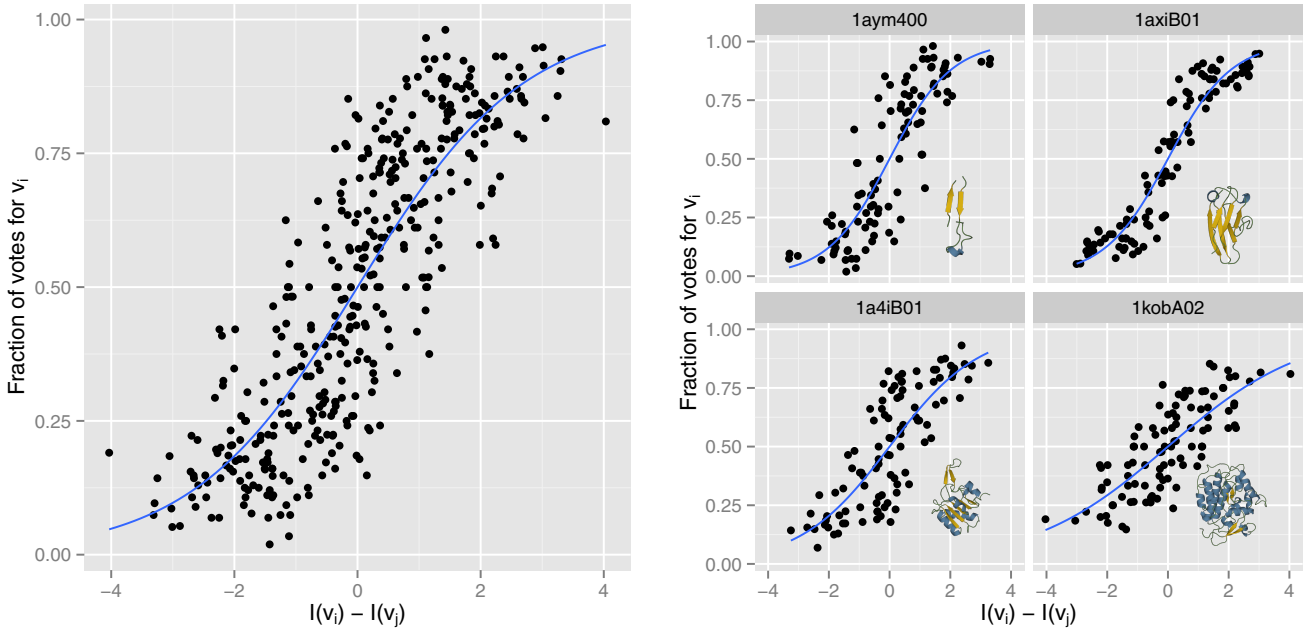

Figure 4: The difference $I(v_i) - I(v_j)$ in viewpoint entropy for viewpoints v_i and v_j (in units of standard deviation), plotted against the fraction of times v_i was chosen over v_j for the full model (left) and all individual test structures (right). In these plots, a point denotes a single viewpoint and the curves show the best-fit logistic regression.

interpret, we only report the *relative fitness* as the log-likelihood of a model, scaled to the range of log-likelihoods between the null-model and the saturated model [SLF*11]. The relative fitness is a value between 0 and 1, where 0 is the fitness of the null-model and 1 denotes the fitness of the saturated model.

As can be seen from Table 2 and Figure 4, viewpoint entropy is a significant predictor of human preference for viewpoints of our test structures for the full model (including all structures), as well as for all structures individually. However, the fitness is highest for `1axiB01` and lowest for `1kobA02`, showing that the quality of the model fit depends on the structure. Overall, these results suggest that viewpoint entropy can be useful in predicting viewpoint preference for ribbon representations of protein structures.

6. Validation

In order to validate our results, we asked experts in molecular biology to provide their preferred viewpoint of the domains used in Section 4, and compared the results to viewpoint entropy. To this end, we recruited 12 participants with a background in molecular biology to view the same 4 structures as were used in the main study. Each participant viewed every structure in PyMOL [Sch10] starting from a random viewpoint, and was asked to rotate the structure with a computer mouse until they decided on a view that best revealed the shape of the object. The wording of the question was chosen such that it matches with the task question posed to the non-experts.

During testing, the screen of the participant as well as their commentary were recorded throughout the whole session. As the selection of viewpoints was part of a larger qualitative experiment on bioinformatics software, the level of expertise with protein structures was not assessed at the beginning of testing. We therefore used the video material to assess the level of expertise of all subjects. As a result of this assessment, the viewpoints chosen by 5 participants were not included in the analysis, as they either specifically mentioned not having prior experience in viewing protein structures and/or other comments implied so. None of the remaining participants expressed familiarity with the test structures throughout the experiment.

Figure 5 (top) shows the distribution of viewpoint entropy for all experts' preferred viewpoints and their locations on the viewing sphere (bottom). While there is a clear preference for high entropy viewpoints in the structure with id `1aym400`, experts seem to pick viewpoints across the whole range of entropies for the other structures. However, across all structures, both low and high entropy viewpoints are among the experts' picks (except for `1aym400`, where mostly high entropy views were chosen). In addition, for `1axiB01` and `1a4iB01`, up to 3 experts chose low to mid-range entropy views. Finally, there is little agreement among experts on the viewpoint entropy for the most complex structure `1kobA02`.

The spatial location of viewpoints on the viewing sphere as illustrated in Figure 5 (bottom) shows a similar pattern: We can see 3 distinct clusters for the simplest structure; for all other structures, there is more variation overall with a single small cluster of 2 to 3 viewpoints. It is important to point out that viewpoints cluster in

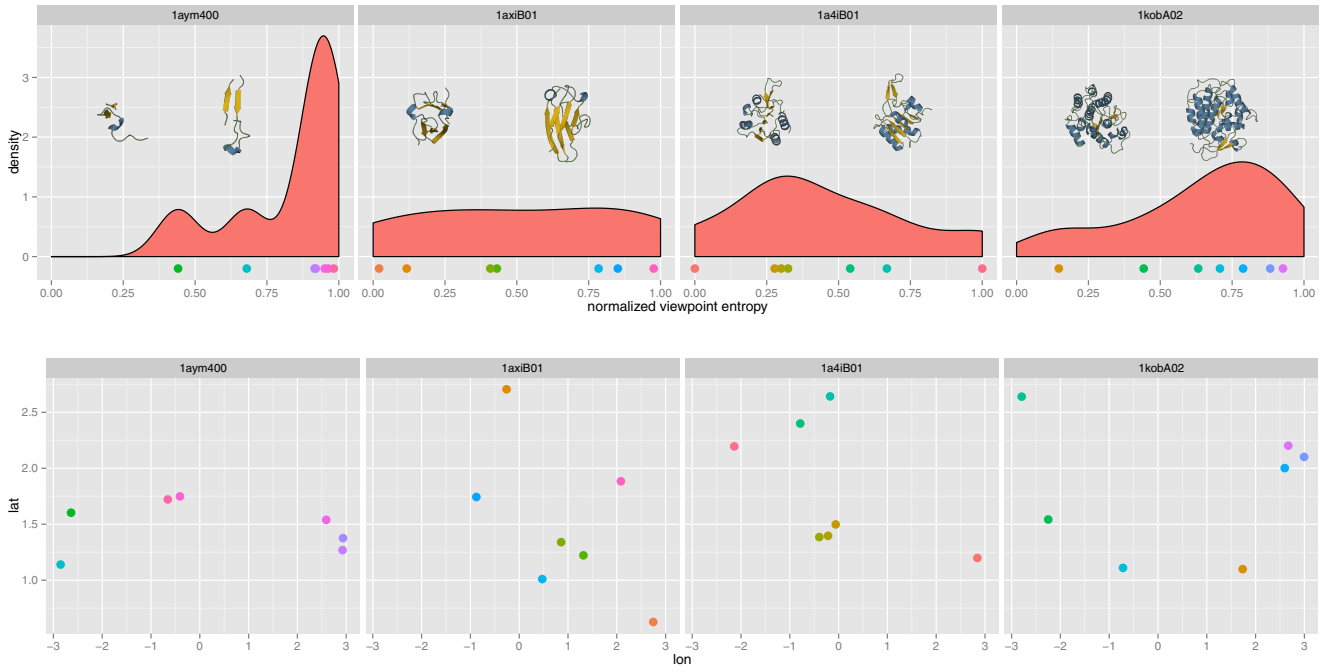

Figure 5: The distribution of viewpoint entropies (top) and spatial locations on the viewing sphere (bottom), for all viewpoints chosen by experts. For every structure, an image of the lowest and highest entropy viewpoint chosen by experts was added to the plot. All of these viewpoints except for the low entropy viewpoint of `1aym400` are either close or antipodal (on the sphere of camera positions) to the highest (lowest) entropy viewpoints from Figure 1.

both the viewpoint entropy space as well as spatially, which suggests that viewpoint entropy might guide the experts' choices in viewpoints.

7. Discussion

In this work, we conducted an evaluation of viewpoint entropy as being a good predictor for the human viewpoint preference of protein structure ribbon representations. Within these constraints, our results suggest that viewpoint entropy performs significantly better than a random guess overall, while revealing that the relative fitness of the model varies considerably with the type of the structure: The viewpoint entropy for structures with mainly beta sheets (`1axiB01`) matches human preference more closely than for structures with a complex arrangement of alpha helices (`1kobA02`). Considering how secondary structure elements are visualized in ribbon representations of proteins, it seems that beta sheets are easier for humans to use as visual landmarks — possibly because they are rendered with an arrow indicating a direction.

Another possible source of this variation is the different spatial composition of secondary structure elements: For non-experts, the average consistency decreases from simple, non-globular structures with a clear principal direction (`1aym400`) to globular, more complex and partly symmetric structures, where all 3 principal components of backbone C_α atoms have a similar length (`1kobA02` and `1a4iB01`). Similarly, the agreement on good viewpoints among experts varies greatly with the type of structure. However, in order to test these observations, a larger set of test structures needs to be investigated.

We investigated viewpoint entropy for 4 diverse structures with respect to their composition of secondary structure elements. Hence, we can only claim validity of our evaluation for these structures and viewpoints, although our pilot shows that the model is able to predict good viewpoints for other structures as well.

One of our aims in this work was to investigate the consistency at which non-experts express a choice for a pair of viewpoints, which allowed us to (i) filter for careless subjects but also (ii) to investigate the effect of the type of structure on consistency. In order to quantify consistency, we required participants to conduct each trial twice. As a trade-off between accuracy and effort for our participants, we therefore decided to use 100 viewpoint samples.

Other parameters that might have an effect on the results of this study are the color map and the shading used to render the sample structures. We have chosen to encode secondary structure both via shape and color in order to facilitate comparison between different viewpoints. The reasoning behind this decision is that protein structures have complex shapes that require redundant encoding to make the study feasible for non-experts.

Both experts and non-experts consistently chose high entropy viewpoints for all 4 domains. In addition, experts were further guided by low entropy viewpoints (for all but the structure with ID

1aym400). The possible use of low entropy viewpoints for selecting good views of small molecules has been proposed previously for ball-and-stick representations [VFSL06]; here we provide evidence that this is also the case for ribbon representations of *single* macromolecules. When asked for the reason of their choice, experts indicated that low entropy viewpoints reveal potentially important global features of the structure, such as the "sandwich" structure in 1a4iB01 or the "tunnel" in 1axiB01.

Some experts also picked views in-between low and high entropy, which can be explained as follows: Even though specifically asked to choose a view that revealed the overall shape of the structure, some experts seemed to pick views that emphasize specific features of the structure instead. As opposed to naïve participants, experts were allowed to explore the structure as long as they wanted to by rotating it manually. By interacting with and acquiring multiple views of the structure, more cues were available to get a better sense of the structure's depth. Even though a specific 3D shape of the protein has been verbally identified by some experts, they still chose a viewpoint that did not show that. One reason could be that the depth information of the structure has already been properly assessed, thus the chosen viewpoint did not need to capture that. Another reason could be that the participants are prioritizing either features or the overall shape; favoring one often requires occluding the other.

Finally, we have chosen to use viewpoint entropy to evaluate viewpoint goodness. As our methodology collecting human preference data is independent of the viewpoint goodness measure, our data can be used to evaluate any other combination of such measures as well.

8. Conclusion and Future Work

In this work, we presented an evaluation of viewpoint entropy to predict good views of ribbon representations of protein structure. Our results indicate that viewpoint entropy is a significant predictor for the human preference of viewpoints overall, while the complexity of the structure seems to have an effect on the accuracy of the predictions. However, to investigate the significance of the structure complexity on viewpoint selection further, more data on additional protein structures need to be collected. We further provided evidence that experts also prefer low entropy viewpoints of macromolecules to highlight important functional elements.

Although we tested only 4 structural models, these samples represent a large number of domains contained in the PDB. In future work, we are seeking to extend our set of benchmark structures to a wider range of models, including samples from the PDB composed of multiple domains and multiple molecules.

Acknowledgements

This work was supported by CSIRO's OCE Science Leader program and its Computational and Simulation Sciences platform, as well as by the Australian Research Council under Linkage Project LP140100574.

References

[BT52] BRADLEY R. A., TERRY M. E.: Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika 39*, 3/4 (1952), 324–345. 6

[BTB99] BLANZ V., TARR M. J., BÜLTHOFF H. H.: What object attributes determine canonical views? *Perception 28*, 5 (1999), 575–599. 4

[BWF*00] BERMAN H. M., WESTBROOK J., FENG Z., GILLILAND G., BHAT T. N., WEISSIG H., SHINDYALOV I. N., BOURNE P. E.: The Protein Data Bank. *Nucleic Acids Research 28*, 1 (2000), 235–242. 2, 4

[CSR*14] CHAPMAN P., STAPLETON G., RODGERS P., MICALLEF L., BLAKE A.: Visualizing sets: An empirical comparison of diagram types. In *8th International Conference on Diagrammatic Representation and Inference*. 2014, pp. 146–160. 3

[DBH14] DEMIRALP C., BERNSTEIN M., HEER J.: Learning perceptual kernels for visualization design. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1933–1942. 3

[DCMP10] DOULAMIS N., CHRONIS E., MIAOULIS G., PLEMENOS D.: Personalized View Selection of 3d Molecular Proteins. In *Intelligent Computer Graphics*, no. 321 in Studies in Computational Intelligence. Springer Berlin Heidelberg, 2010, pp. 211–227. 2

[HB10] HEER J., BOSTOCK M.: Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), pp. 203–212. 3

[JH14] JOHNSON G. T., HERTIG S.: A guide to the visual analysis and communication of biomolecular structural data. *Nature Reviews Molecular Cell Biology 15*, 10 (2014), 690–698. 2

[JS06] JI G., SHEN H. W.: Dynamic view selection for time-varying volumes. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 1109–1116. 2

[KHC*08] KEEHNER M., HEGARTY M., COHEN C., KHOOSHABEH P., MONTELLO D. R.: Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science 32*, 7 (2008), 1099–1132. 2

[KKL*15] KOZLIKOVA B., KRONE M., LINDOW N., FALK M., BAADEN M., BAUM D., VIOLA I., PARULEK J., HEGE H.-C.: Visualization of biomolecular structures: State of the art. In *Eurographics Conference on Visualization - STARs* (2015), The Eurographics Association. 2

[MKK*14] MWALONGO F., KRONE M., KARCH G., BECHER M., REINA G., ERTL T.: Visualization of molecular structures using state-of-the-art techniques in WebGL. In *Proceedings of the Nineteenth International ACM Conference on 3D Web Technologies* (2014), pp. 133–141. 2

[MNTP07] MÜHLER K., NEUGEBAUER M., TIETJEN C., PREIM B.: Viewpoint selection for intervention planning. In *Proceedings of the 9th Joint Eurographics / IEEE VGTC Conference on Visualization* (2007), pp. 267–274. 2

[MTM12] MANTIUK R. K., TOMASZEWSKA A., MANTIUK R.: Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum 31*, 8 (2012), 2478–2491. 4

[OGF*10] O'DONOGHUE S. I., GOODSELL D. S., FRANGAKIS A. S., JOSSINET F., LASKOWSKI R. A., NILGES M., SAIBIL H. R., SCHAFFERHANS A., WADE R. C., WESTHOF E., OLSON A. J.: Visualization of macromolecular structures. *Nature Methods 7* (2010), S42–S55. 2

[OSK*15] O'DONOGHUE S. I., SABIR K. S., KALEMANOV M., STOLTE C., WELLMANN B., HO V., ROOS M., PERDIGÃČO N., BUSKE F. A., HEINRICH J., ROST B., SCHAFFERHANS A.: Aquaria: simplifying discovery and insight from protein structures. *Nature Methods 12*, 2 (2015), 98–99. 2, 4

[PaHS*15] PERDIGÃO N., HEINRICH J., STOLTE C., SABIR K. S., BUCKLEY M. J., TABOR B., SIGNAL B., GLOSS B. S., HAMMANG

C. J., ROST B., SCHAFFERHANS A., O'DONOGHUE S. I.: Unexpected features of the dark proteome. *Proceedings of the National Academy of Sciences* (2015), 15898–15903. 3

[SCD*13] SILLITOE I., CUFF A. L., DESSAILLY B. H., DAWSON N. L., FURNHAM N., LEE D., LEES J. G., LEWIS T. E., STUDER R. A., RENTZSCH R., YEATS C., THORNTON J. M., ORENGO C. A.: New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3d structures. *Nucleic Acids Research 41*, Database issue (Jan. 2013), D490–498. 4

[Sch10] SCHRÖDINGER, LLC: The PyMOL Molecular Graphics System, Version 1.3r1. Aug. 2010. 3, 7

[SF01] SILVERSTEIN D. A., FARRELL J. E.: Efficient method for paired comparison. *Journal of Electronic Imaging 10*, 2 (2001), 394–398. 5

[SLF*11] SECORD A., LU J., FINKELSTEIN A., SINGH M., NEALEN A.: Perceptual models of viewpoint preference. *ACM Trans. Graph. 30*, 5 (2011), 109:1–109:12. 2, 3, 4, 5, 7

[TFTN05] TAKAHASHI S., FUJISHIRO I., TAKESHIMA Y., NISHITA T.: A feature-driven approach to locating optimal viewpoints for volume visualization. In *IEEE Visualization* (2005), pp. 495–502. 2, 3

[VFSH01] VÁZQUEZ P., FEIXAS M., SBERT M., HEIDRICH W.: Viewpoint selection using viewpoint entropy. In *Proceedings of the Vision Modeling and Visualization Conference* (2001), pp. 273–280. 2, 3

[VFSH03] VÁZQUEZ P.-P., FEIXAS M., SBERT M., HEIDRICH W.: Automatic view selection using viewpoint entropy and its application to image-based modelling. *Computer Graphics Forum 22*, 4 (2003), 689–700. 2

[VFSL02] VÁZQUEZ P.-P., FEIXAS M., SBERT M., LLOBET A.: Viewpoint entropy: A new tool for obtaining good views of molecules. In *Proceedings of the Symposium on Data Visualisation* (2002), pp. 183–188. 2, 3

[VFSL06] VÁZQUEZ P.-P., FEIXAS M., SBERT M., LLOBET A.: Real-time automatic selection of good molecular views. *Computers & Graphics 30*, 1 (2006), 98–110. 2, 8

[WLH97] WONG T.-T., LUK W.-S., HENG P.-A.: Sampling with Hammersley and Halton Points. *Journal of Graphics Tools 2*, 2 (Jan. 1997), 9–24. 5