Geometry-based Computation of Symmetric Homo-oligomeric Protein Complexes

Christopher Miles* Brian Olson* Amarda Shehu *,†

Abstract

The need to engineer novel therapeutics and functional materials is driving the in-silico design of molecular complexes. This paper proposes a method to compute symmetric homo-oligomeric protein complexes when the structure of the replicated protein monomer is known and rigid. The relationship between the structure of a protein and its biological function brings the in-silico determination of protein structures associated with functional states to the forefront of computational biology. While protein complexes, arising from associations of protein monomers, are pervasive in every genome, determination of their structures remains challenging. Given the difficulty in computing structures of a protein monomer, computing arrangements of monomers in a complex is mainly limited to dimers. A growth in the number of protein complexes studied in wet labs is allowing classification of their structures. A recent database shows that most naturally-occurring protein complexes are symmetric homo-oligomers. The method presented here exploits this database to propose structures of symmetric homo-oligomers that can accommodate spatial replications of a given protein monomer. The method searches the database for documented structures of symmetric homo-oligomers where the replicated monomer has a geometrically-similar structure to that of the input protein monomer. The proposed method is a first step towards the in-silico design of novel protein complexes that upon further refinement and characterization can serve as molecular machines or fundamental units in therapeutics or functional materials.

1 Introduction

Protein chains assemble as building blocks into structures of greater complexity in cells. Protein complexes play central roles in ion transport and regulation in membranes, transduction of signal down chemical pathways, degradation of proteins, and even transcriptional regulation [1]. Fig. 1 shows one such complex, the GroEL chaperonin, a heptamer that corrects structural defects in newly-synthesized proteins [2]. Interactions between the seven monomers give GroEL both its three-dimensional (3D) structure and biological function.

Evidence of protein structure determining protein function has made structure determination a major focus of molecular biology [3]. Driven by the need for novel therapeutics and functional materials, decades of research have been devoted to structure determination both in wet labs and in silico [4–9]. Such research has targeted mainly the characterization of protein monomers, single polypeptide chains that assume a unique structure under native conditions [10].

Computing native or native-like structures of a protein monomer is a challenging problem. The space of possible arrangements, conformations, of a protein chain is vast and high-dimensional. The energy surface associated with the space, rising from interactions among atoms in the chain, is rugged and can be probed only with empirical energy functions [11]. Though native conformations are associated with the global minimum in the funnel-like protein energy surface [12,13], searching the conformational space for such conformations remains computationally challenging [14,15].

Given the difficulty in computing native conformations of a single protein chain, research in determining the structure of a protein complex has only recently begun to gain attention [16, 17]. Computing structures of a complex involves exploring different ways of arranging and positioning monomers in a complex. The computational complexity of deriving the native structure of a monomer from knowledge of its amino-acid sequence makes it infeasible to approach structure determination of a protein complex ab initio.

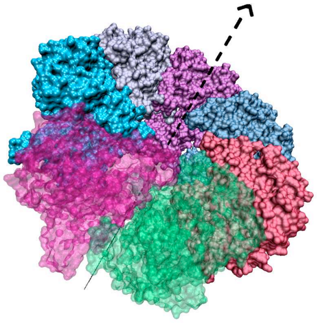


Figure 1: The seven monomers in GroEL are in different colors, with the front two drawn in transparent. The black line denotes the symmetry axis.

The vast conformational space of a protein chain presents a challenge to algorithms that explore this space in search of relevant monomeric structures. This problem is only exacerbated when considering direct combinatorial extensions of these algorithms for the simultaneous exploration of possible structures for all monomers in a complex [18,19]. Employing protein monomers with known native structures allows approaching the in-silico structure determination of a protein complex as a search for possible arrangements of fixed monomeric structures in the complex.

About 60%-70% of the proteins in every genome are homo-oligomers, complexes assembled from identical monomers [20]. On one hand, such pervasiveness of protein complexes warrants more urgency into structural studies to offer mechanistic insight into their biological functions. On the other hand, these estimates allow computational methods

[*] Department of Computer Science, George Mason University

[†] Corresponding Author, amarda@cs.gmu.edu

to focus on structure determination of homo-oligomers. That is, given a protein chain of known structure, compute different ways of spatially replicating it in a complex so that the resulting complex structure is physically-realistic and valid.

The large size of a protein complex, typically tens of thousands of atoms, poses challenges to structure determination in wet labs [2, 21]. Size also challenges the scalability of computational methods that simulate physical motions of individual atoms [22]. Assuming that the structure of the protein monomer is not only known, but also rigid, allows dedicating resources on searching for rigid-body transformations that properly align monomers in a complex.

Research in computing such transformations for dimers, protein complexes consisting of two monomers, is active [18, 23–27]. The problem is referred to as the docking problem. The typical approach in addressing the docking problem is to keep one monomer fixed and compute possible dimer structures by sampling the space of rotations and translations of the docked monomer with respect to the fixed one. This approach is not feasible on arbitrary complexes, as the dimensionality of the search space increases exponentially with the number of monomers in a complex.

Two key features of protein complexes provide auspicious opportunities for efficient in-silico characterization of their structures. Recent studies have shown that functional, genetic, and physicochemical needs have driven an evolutionary selection of protein complexes towards symmetric homo-oligomers [1]. Symmetry, nature's way of employing hierarchy to achieve complexity, and redundancy, nature's parsimony in reusing identical protein monomers, are two key features that have begun to be exploited for an efficient computation of structures of symmetric homo-oligomers.

The number of protein complexes whose structures have been determined in wet labs is growing. This growth is spawning new databases devoted to classification of complex structures, which are revealing that most naturally-occurring homo-oligomers exhibit either cyclic or dihedral symmetry [20]. The GroEL chaperonin shown in Fig. 1 is an example of a heptamer with cyclic symmetry. Such findings have prompted computational methods to focus on computing structures of protein homo-oligomers that observe cyclic or dihedral symmetry [16, 17].

While focusing on symmetric homo-oligomers narrows the number of ways monomers can be positioned in a complex, the search space remains high-dimensional. In addition, the number of monomers in the complex (also referred to as the oligomeric number) is not known a priori. Current methods iterate over possible oligomeric numbers, exploring a high-dimensional space to compute complex structures of a given oligomeric number. Structures where monomers are not in serious collision with one another are ranked according to agreement with experimental data, often obtained from Nuclear Magnetic Resonance (NMR) experiments [16, 17]. The NMR data allow determining whether computed structures of a homo-oligomer are feasible.

Work in [16] focuses on cyclic homo-oligomers. For each oligomeric number, the space of possible symmetry axes is systematically explored in search of valid complex structures. The search space in [16] is subdivided into regions worthy of further exploration and regions corresponding to structures in direct violation of NMR data. The NMR data employed are distance constraints due to Nuclear Overhauser Effects (NOE), which allow determining whether the packing of monomers in a complex structure is valid or not. Other work computes cyclic complex structures through a grid search algorithm, restricting the number of proposed structures by their agreement with NMR residual dipolar couplings (RDCs) [17].

While docking research cannot readily be extended to homo-oligomers, most recent methods on computing homo-oligomeric structures are limited by the availability of experimental data on the complex at hand. These methods cannot be employed to design new symmetric homo-oligomeric structures. However, growing databases of structures of naturally-occurring protein complexes can be leveraged to design novel complexes similar to how the Protein Data Bank (PDB) [28] of native protein structures is employed to design native structures of single polypeptide chains [6, 9, 29–31].

The method proposed here focuses on computing structures of symmetric homo-oligomers that can accommodate spatial replications of a given monomeric structure. The method exploits the 3D complex database, a recent database that gathers and classifies experimentally-obtained structures of naturally-occurring protein complexes [20]. The database allows extracting protein homo-oligomers of cyclic or dihedral symmetry. The proposed method searches over this database to obtain a set of symmetric homo-oligomeric structures preferred by protein monomers of geometrically similar structure to the structure of the given monomer. The method is referred to as `Espreso`, for gEometric structure prediction in symmetric homo-oligomers.

The presented `Espreso` method draws inspiration from fragment assembly methods that address structure prediction of a protein monomer. In these methods, the chain of a monomer of unknown structure is segmented in consecutive overlapping fragments. Non-redundant subsets of known protein structures, extracted from the Protein Data Bank (PDB), are then exploited to compute possible structures preferred by short protein fragments. The structure of the protein chain is then assembled from structures of its fragments [6, 9, 29–31].

`Espreso` exploits the 3D complex database to propose structures of symmetric homo-oligomers that can accommodate spatial replications of a given protein monomer of known structure. The structure of the given monomer is compared to that of monomers in the complexes stored in the database. Complexes whose monomeric structures are geometrically similar to the structure of the input protein monomer are proposed as possible complexes that can accommodate replications of the given monomer.

Espreso works under the assumption that monomeric structure is the first determinant of both the number and the way monomers are arranged in a complex. Such an assumption is warranted, given that the packing of monomers in a complex is geometrically-constrained by the structure of each monomer. Designing efficient yet effective measures of geometric similarity for structures of protein monomers, however, is an active area of research [32].

Part of the inability to design effective similarity measures for protein structures originates from the potentially large number of atoms in a protein monomer. Measures such as least root-mean-squared-deviation (lRMSD) average over spatial deviations of atoms in two monomeric structures under comparison, effectively masking away differences. Moreover, since lRMSD requires that compared structures are aligned to one another, the measure is too expensive to execute on a database of potentially thousands of structures. In addition, lRMSD cannot be readily extended to compare structures of monomers of different number of atoms. Correspondence of the atoms becomes a problem.

Simple and fast measures such as comparing radii of gyration (Rg) between two monomeric structures (Rg refers to the average atomic distance from the center of mass) are too coarse for any practical purpose. Structure comparison methods, while abundant, are beyond the scope of a comprehensive summary in this work. The comparison of monomeric structures in this work focuses on capturing overall shape similarity rather than fine structural details.

The Espreso method presented here captures the overall shape similarity between two monomers through the recently proposed ultrafast shape recognition (USR) features [33]. These features, detailed in section 3, are proposed in [33] to efficiently compare the structure of a small ligand against millions of ligand structures stored in pharmaceutical databases. The features have been recently employed to keep track of computed conformations of a protein chain in a low-dimensional projection space for an efficient exploration of the protein conformational space [34].

The USR features in this work are employed to define coordinates of a monomeric structure in a low-dimensional projection space. Projections of two monomeric structures are compared through a novel normalized Manhattan-based similarity score, which builds on the similarity score proposed in [33]. Section 3 shows that USR-based projections capture well the similarity between two monomeric structures without getting lost in fine details such as content of secondary structure segments in a protein chain.

The proposed Espreso method is applied on seven protein monomers that have diverse lengths and native structures. The chosen monomers are known to be the building block of homo-oligomers that are important in molecular biology research in the context of therapeutic applications. The results in section 4 show that the method correctly captures the complexes populated in nature by the chosen monomers. Additional complexes are obtained, showing that the input monomers can spatially replicate in novel ways.

Further tests shown in section 4 quantify the richness of the 3D complex database. The method is applied to more than $36,000$ non-redundant protein chains extracted from the PDB. Complex structures are proposed for each chain at different thresholds of similarity between the given monomeric structure and that of the monomers of the homo-oligomers in the 3D complex database. Predictions are shown as a function of similarity.

The proposed method is intended as a first rapid step when searching for different ways of arranging copies of a monomer of known structure in a complex. The 3D complex database allows Espreso to focus only on structures of naturally-occurring complexes. By employing the database as the set of solutions preferred by nature (in so far as the richness of the database allows, an issue discussed further in sections 3 and 5), the method circumvents computationally-expensive combinatorial searches over viable arrangements of an oligomeric number of monomeric structures.

Complex structures proposed by Espreso for a given monomer are not by any means complete. Alternative arrangements of the monomeric structure may exist that are not yet represented in the database. The issue of the current status of the database is further discussed in sections 3 and 5. Moreover, the complex structures obtained by the method need to be further evaluated with energetic considerations. Refinement of obtained complex structures can be carried out in computationally-demanding simulations that employ physically-realistic energy functions. The lowest-energy resulting complexes can then be employed to guide wet-lab research in engineering novel complexes or studying properties such as stability and function in greater detail.

The complex design aspect of the proposed method has far-reaching implications for proposing physically-realistic models of molecular complexes. Novel complexes that act as molecular machines can be designed to have specific structural morphologies and functionalities. Proposing such complexes in the dry lab and then synthesizing and characterizing them in greater detail in the wet lab has the potential to push forward molecular biology research that encompasses drug design and material science.

The related work places the proposed Espreso method in context in section 2. Espreso is then described in detail in section 3. Applications on $36,512$ non-redundant protein chains and seven chosen monomers are presented in section 4. The work concludes with a discussion in section 5.

2 Related Work

Resolving the structure of a protein complex poses significant challenges in the wet lab due to complex size and resolution quality [2,21]. Traditional protocols based on solution NMR employ local optimization techniques that often get trapped in local minima, consequently missing the true structure of a complex [35]. For these reasons, in-silico approaches to structure prediction are gaining ground.

Computational methods on complex structure prediction have focused primarily on dimers. Docking two monomers follows a three-stage procedure [36]. The first stage searches for a set of physically-realistic structures of the monomers under consideration. In the second stage, one monomer is kept fixed, while structures of the second monomer are transformed through rigid-body transformations to dock them onto the first monomer. In the final stage, the resulting dimer structures are ranked according to energetic criteria, similarity to experimentally-determined native structures of the dimer, or agreement with other available experimental data.

Literature on searches for physically-realistic structures of a protein monomer is rich, as the problem of structure prediction is central in molecular biology [3,37]. While a detailed summary is beyond the scope of this work, methods include Fast Fourier Transforms (FFT) on a voxel grid that discretizes a monomeric structure [38], Monte Carlo or Molecular Dynamics [18,22], genetic algorithms [39], fragment assembly [9,29,40], and many more (cf. to [6]).

Docking methods that forego the first stage assume rigid or semi-rigid monomeric structures. These methods address rigid-body docking or docking without flexibility. They employ either one or a manageable few structures of each monomer under consideration. These methods are very effective when employing co-crystalized monomeric structures. When working on separately crystalized structures, these methods tend to yield many false positives for the native structure of the dimer [25].

Docking methods that include the first stage consider the conformational flexibility of the protein monomers. Work in [23,24] removes the effects of van der Waals (vdw) interactions from the edges of docked monomers. These interactions are then reintroduced systematically, allowing the monomers to shift and sample low-energy conformations. Because of the additional search for alternative conformations, flexible docking methods are inherently more computationally demanding than rigid-body docking methods. Flexible docking methods can also yield too many conformations that, while relevant on their own, do not allow docking the monomers onto each-other.

To handle the computational demands of flexible docking, many methods consider limited local flexibility [38]. These methods implement the docking-with-some-flexibility approach, where the monomeric structures are considered semi-rigid. The method in [38], for instance, maintains the overall monomeric structure rigid, but allows details at the interaction interface to change slightly. More recent docking methods allow the backbones of the docked monomers to move as well [41]. While more comprehensive in their search of docked structures, these methods can generate many more structures than are practical to score.

Scoring functions can be computationally-demanding to execute on a large number of dimer structures. These functions all attempt to correctly identify the dimer structure with the lowest energy. Scoring functions typically use electrostatic, vdw, and hydrostatic energetic interactions [23, 25,26]. Even when employing physically-realistic energy functions to rank computed dimer structures, many docking methods are not complete. Dimer structures that rank low in energy have been shown to disagree with experimental data such as NOE distance constraints (cf. to [16]).

Extending docking methods to compute structures of complexes of more than two monomers is not practical, as the dimensionality of the search space increases exponentially with the number of monomers in the complex. Recently proposed methods for arranging an arbitrary number of monomers in a complex move beyond the docking framework. To efficiently search for viable arrangements of monomers of known rigid structure, current methods focus on symmetric homo-oligomeric complexes, where the monomers can be arranged together only in a limited number of ways. Specifically, current methods focus on cyclic symmetry, which makes it easier to arrange monomeric structures around a rotational axis. These methods further limit the number of ways monomers are arranged together by employing experimental data. The method in [16] employs NMR NOE distance constraints, whereas that in [17] employs NMR RDC data.

The method proposed in [16] uses a branch and bound algorithm to subdivide the conformational space. Regions of the space that correspond to cyclic homo-oligomeric structures that are either in serious vdw clashes or violate the NOE distance constraints are discarded and not considered for further subdivision. The method simultaneously evaluates interactions among all monomers in the complex rather than iterate over pairwise interactions. The method in [17] proposes using NMR RDC data to evaluate computed oligomeric structures due to possible ambiguity in the experimental assignment of intra and inter monomeric NOE distance constraints through NMR.

The `Espreso` method proposed in this work considers both cyclic and dihedral homo-oligomers, as long as the monomeric structure in the symmetric homo-oligomer under consideration is geometrically similar to the structure of the input monomer. By sampling solutions from the 3D complex database, `Espreso` circumvents the problem of determining oligomeric number and searching for physically-realistic arrangements of that number of monomers in a complex.

3 Methods

Espreso searches the 3D complex database for complex structures that can spatially accommodate a given monomeric structure. While the database contains a non-redundant set of naturally-occurring complex structures determined in the wet lab, the subset considered here contains only symmetric homo-oligomers. This subset is extracted from the database through the functionality provided in [20].

The extracted subset is considered as the possible search space of symmetric homo-oligomers preferred by proteins in cells. The basic process in Espreso is to iterate over this subset and identify those symmetric homo-oligomers whose monomers are geometrically similar in structure to the structure of the given monomer.

Geometric similarity between two monomeric structures is estimated not over the cartesian coordinates of atoms in the monomers, but rather in a low-dimensional projection space. A monomeric structure is projected on an eight-dimensional (8d) space and represented through a vector of eight coordinates. Geometric similarity between two monomeric structures is then measured through a novel similarity function that operates on two 8d vectors.

Since the solution set considered here consists of symmetric homo-oligomers only, the monomeric structures in such a complex are identical to one another within rigid-body transformations. The projection of the monomeric structures on the 8d projection space naturally removes differences due to rigid-body transformations. Therefore, each symmetric homo-oligomer in the considered solution set is represented through the 8d vector of coordinates of one its monomers.

Small structural deviations among the monomers due to flexible side-chains in a protein structure are removed by considering only the backbone of each monomer in the projection. Additional structural deviations arising from slight fluctuations of monomer backbones in a symmetric homo-oligomer are recorded in an average deviation value associated with each complex. It is worth mentioning that the eight coordinates employed in the projection are a subset of the twelve coordinates proposed in [33]. The four coordinates removed from consideration allow discarding noise due to local backbone fluctuations of the monomers in a complex.

The quality of the complex structures proposed for a given monomeric structure depends both on the richness of the 3D complex database and the assumption that geometric similarity is the primary determinant whether copies of a monomeric structure can be accommodated in a symmetric homo-oligomer. Various statistics are compiled over the 3D complex database and its subset of symmetric homo-oligomers to quantify its richness. Representation of a complex, the projection procedure, and the similarity score proposed to estimate placement of a monomer in a given symmetric homo-oligomer are described next. Implementation details conclude the description of the method.

3.1 Estimating Richness of the 3D Complex Database

As of January 2009, the 3D complex database contained $30,475$ structures of non-redundant protein complexes. Redundant PDB submissions of the same protein complex were removed in the compilation of the database [20]. Out of $30,475$ complex structures, $26,831$ structures belong to homo-oligomers. Symmetric homo-oligomers consist of 8939 cyclic and 2613 dihedral complexes.

Fig. 2(a) shows the distribution of complexes in the 3D complex database as a function of oligomeric number. As also observed in [20], the database is heavily biased towards small complexes. The peak of the distribution in Fig. 2(a) is reached on dimers. This is not surprising, since wet-lab experiments have an easier time resolving structures of small complexes. Evolution of oligomeric complexes in cells also seems to favor the formation of dimers through pre-positioned interaction interfaces [1].

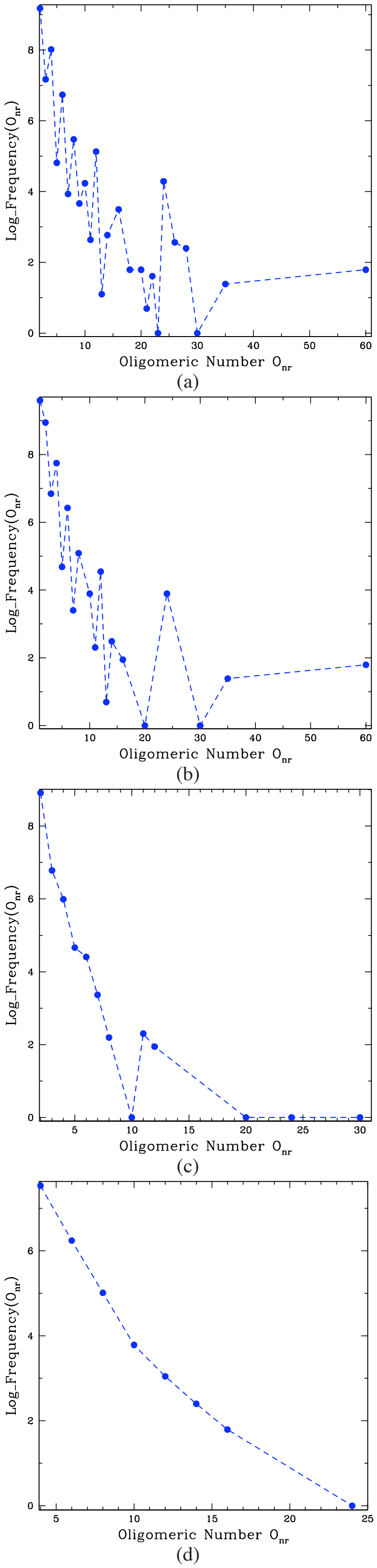
Fig. 2(b) plots the number of homo-oligomers as a function of oligomeric number, whereas (c) and (d) focus on cyclic and dihedral homo-oligomers, respectively. Figs. 2(b)-(d) also highlight that dimers dominate homo-oligomers. Interestingly, as also noted in [20], symmetric homo-oligomers of an even number of monomers seem to be more prevalent than those of odd oligomeric number.

Figs. 2(a)-(d) show that the current state of the 3D complex database favors proposing homo-oligomers of cyclic symmetry over those of dihedral symmetry. Proposed symmetric homo-oligomers for a given monomeric structure are also likely to have a small and even number of monomers.

The functionality associated with the 3D complex database allows extracting complexes according to oligomeric number and symmetry. This functionality is important, as it facilitates extracting from the database the subset of symmetric homo-oligomers. While queries with a given monomeric structure by Espreso are not limited to homo-oligomers of a specific symmetry class, narrowing the scope of the search may be useful. Inexpensive wet-lab experiments can detect symmetry or approximate oligomeric number without the need to determine complex structure in detail. If such information is available, the search can focus on symmetric homo-oligomers with the properties observed in the wet lab.

Since queries by Espreso focus on symmetric homo-oligomers, their distribution in the 3D complex database is analyzed in more detail. Fig. 3(a) plots the number of homo-oligomers as a function of monomeric size. Monomeric size is defined as the number of amino acids in a monomer. Figs. 3(b)-(c) focus on homo-oligomers of cyclic and dihedral symmetry, respectively. As expected, Fig. 3 shows that symmetric homo-oligomers in the database prevalently contain small monomers. The current state of the database seems to better support queries with small monomers.

The distribution of homo-oligomers as a function of



(a)

(b)

(c)

(d)

Figure 2: (a) plots the number of complexes in logarithmic scale as a function of oligomeric number. (b) extracts homo-oligomers. (c) and (d) focus on homo-oligomers of cyclic and dihedral symmetry, respectively.

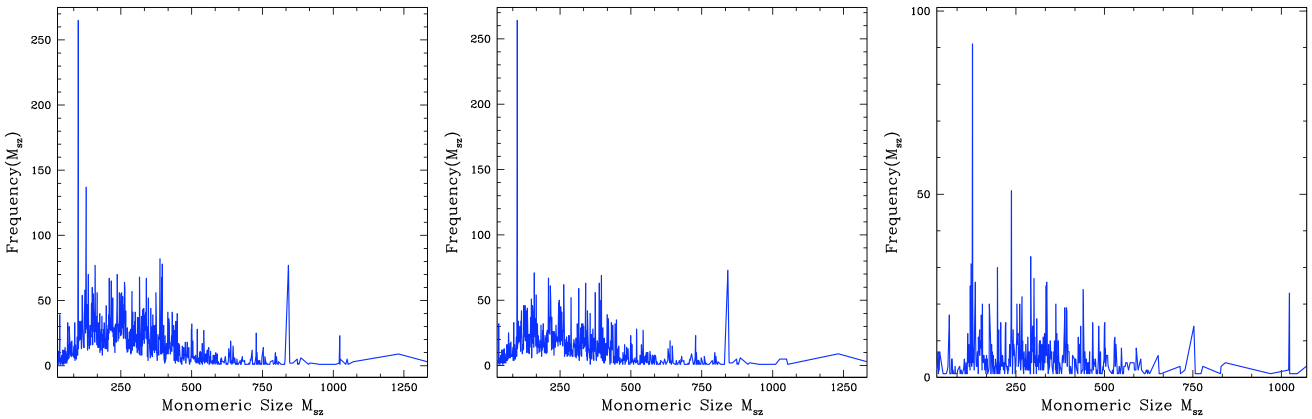
monomeric fold is detailed in Table 3.1. Fold here refers to the content of secondary structure segments. An interesting feature emerges. Monomers that contain exclusively β-sheets are under-represented in the database. This trend persists on symmetric homo-oligomers. Symmetric homo-oligomers, whether cyclic or dihedral, are dominated by monomers that contain both α-helices and β-sheets. The results in Table 3.1 suggest that querying the database with a monomer containing only β-sheets will be limited by the scarcity of complexes with such monomeric folds.

Fold	α	β	α/β
Cyclic Symmetry	614	94	8231
Dihedral Symmetry	93	39	2481
Homo-oligomers	824	138	11117

3.2 Employing a Backbone Representation `Espreso` does not consider all atoms of a monomer when computing projection coordinates of a monomeric structure. Instead, the method considers only the cartesian coordinates of the backbone atoms of a protein chain. The backbone atoms include the N, C_α, C, and O main-chain atoms shared among the twenty classic amino acids. There are several reasons for employing a backbone-resolution representation of a monomeric structure. First, structures resolved in the wet lab can be incomplete. There is often missing information on the cartesian coordinates of side-chain atoms. In addition, side-chain atoms are more flexible in protein structures. While the backbones of monomers in a complex may not be able to move as freely, local fluctuations of side-chains can be accommodated in a complex.

3.3 Representing a Complex Let n be the oligomeric number of a complex. The complex consisting of n monomers can be denoted by $U[n] = \{U_1, \ldots, U_n\}$. U_i refers to the i^{th} monomer, for $1 \leq i \leq n$. Homo-oligomers of n monomers that exhibit the cyclic symmetry are said to belong to the Cn class, whereas those that exhibit the dihedral symmetry are said to belong to the Dn class. For instance, the GroEL chaperonin shown in Fig. 1 belongs to the C7 class. While the 3D complex database includes single monomers, they are removed from consideration here. The `Espreso` method focuses on finding complexes of more than one monomer that can accommodate a given monomeric structure.

3.4 Projecting a Monomeric Structure Employing only the $\{x, y, z\}$ cartesian coordinates of the backbone atoms of a monomer U_i, `Espreso` projects these coordinates onto a low-dimensional vector p_i. The projection builds on the one proposed in [33] for rapid comparison of ligand structures. Using the backbone coordinates of a monomer, the projection procedure designates four atoms to serve as



(a) (b) (c)

Figure 3: (a) plots the number of homo-oligomeric complexes as a function of monomeric size M_{sz}, where M_{sz} refers to the number of amino-acids in a monomer. The complexes are split according to symmetry, cyclic in (b) and dihedral in (c).

reference atoms: the centroid (ctd), the atom closest to the centroid (cst), the atom farthest from the centroid (fct), and the atom furthest from the fct atom (ftf). These four atoms capture the center and corners of a structure.

Employing the above four atoms as references, distances of all other atoms are then computed from the references. Let us denote the vector of distances from the centroid ctd as $\{d_{\rm ctd}\}$, that from the cst atom as $\{d_{\rm cst}\}$, that from the fct atom as $\{d_{\rm fct}\}$, and that from the ftf atom as $\{d_{\rm ftf}\}$. These four vectors contain distributions of distances from the four reference atoms. The work in [33] represents the four distributions through a 12d vector by extracting the first three momenta of each distribution.

The first three momenta computed over a distribution refer to the mean, variance, and skew, respectively. For instance, the three momenta computed over $\{d_{\rm ctd}\}$ are denoted by $\mu^1_{\rm ctd}$, $\mu^2_{\rm ctd}$, and $\mu^3_{\rm ctd}$, where $\mu^1_{\rm ctd} = \langle d_{\rm ctd}\rangle$, $\mu^2_{ctd} = \langle d_{\rm ctd} - \mu^1_{\rm ctd}\rangle$, and $\mu^3_{\rm ctd} = \langle d_{\rm ctd} - \mu^2_{\rm ctd}\rangle$. It is worth noting that various definitions exist for the skew. The one employed here is the simple skew of a distribution.

Espreso employs eight out of the twelve projection coordinates. The third momenta are removed from consideration. Computing the third momenta over structures of identical monomers in a symmetric homo-oligomer revealed often significant differences. While almost identical ligand structures in [33] had small deviations due to the small number of degrees of freedom in them, in structures of identical monomers even slight backbone fluctuations can significantly affect the third momenta. Since the skew is very sensitive to slight backbone fluctuations, the method extracts only the first and second momenta. A monomeric structure in this work is therefore projected on an 8d vector.

3.5 Computing Geometric Similarity Employing the above momenta, a monomeric structure U_i is represented (in its projected form) through the 8d vector $p_i = \{\mu^1_{\rm ctd}, \mu^2_{\rm ctd}, \mu^1_{\rm cst}, \mu^2_{\rm cst}, \mu^1_{\rm fct}, \mu^2_{\rm fct}, \mu^1_{\rm ftf}, \mu^2_{\rm ftf}\}$. The similarity between two monomeric structures U_i and U_j is computed in the projected space over their respective projection vectors p_i and p_j. The similarity score employed here adapts the one proposed in [33]. Operating on 12d vectors p_i and p_j, the score in [33] was $S_{ij} = \frac{1}{1+\frac{1}{12}\cdot\sum_{k=i}^{12}|p_i[k]-p_j[k]|}$.

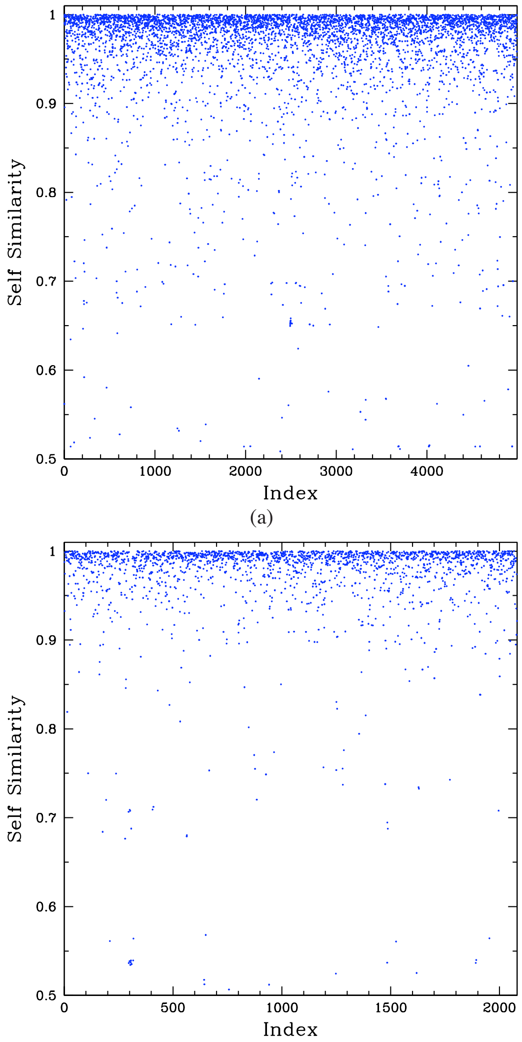
The Manhattan-based similarity score S_{ij} in [33] reaches 1 when the 12d vectors are exactly the same and 0 when these vectors are most dissimilar. Differences are averaged among all twelve coordinates. The mean, variance, and skew of all distributions are considered equally important.

3.6 Proposing a New Similarity Score The similarity score in [33] works well on mostly rigid small ligands, as supported by the results presented in [33]. As mentioned above, backbones of long protein chains are more flexible. However, slight backbone fluctuations should not adversely affect the similarity among two mostly identical monomeric structures. A new similarity score S_{ij} is proposed here that employs only the first and second momenta of the four distance distributions in a monomeric structure.

The proposed score normalizes differences between corresponding momenta of projections p_i and p_j of two monomeric structures U_i and U_j. Normalization is introduced to properly scale differences between corresponding momenta. However, no a priori knowledge is available on the possible range of values of the momenta. Therefore, the normalization in the proposed score scales differences by the maximum value between the compared momenta. Specifically, the similarity score proposed here and employed by Espreso to compare two monomeric structures is:

$$S_{ij} = \frac{1}{1+\frac{1}{8}\cdot\sum_{k=i}^{8}\frac{|p_i[k]-p_j[k]|}{max\{|p_i[k]|,|p_j[k]|\}}}$$

3.7 Estimating Self Similarity in a Symmetric Homo-oligomer Given a symmetric homo-oligomer $U[n] = \{U_1, \ldots, U_n\}$, 8d projections of its monomers result in the set $\{p_1, \ldots, p_n\}$. One of the monomers can be arbitrarily chosen to be the representative. That is, its projection p_1 is associated with the complex so that queries with p_1 will capture the complex $U[n]$. To allow for the fact that there may be slight deviations among monomers even in a symmetric homo-oligomer, an average deviation $\langle\epsilon\rangle$ is maintained. This deviation is a self similarity score that averages over the set $\{S_{12}, \ldots, S_{1n}\}$ of similarity score of all $n-1$ monomers from the one chosen as reference.



(a)

(b)

Figure 4: The x axis in (a) and (b) iterates over homo-oligomers of cyclic and dihedral symmetry, respectively. The y axis plots the self similarity between the monomers in a complex, averaged over the number of monomers as described in the self similarity score defined above.

The self similarity score is defined as $\frac{1}{n-1} \cdot \sum_{j=2}^{n} S_{1j}$. Figs. 4(a)-(b) plot this score computed over symmetric homo-oligomers. The plots show that the lowest self similarity value is higher than 0.5, with most of the symmetric homo-oligomers reaching self similarity values of $0.9-1.0$. The results on the self similarity value of symmetric homo-oligomers confirm that the similarity score proposed in this work captures well the equivalence within rigid-body transformations of identical monomers.

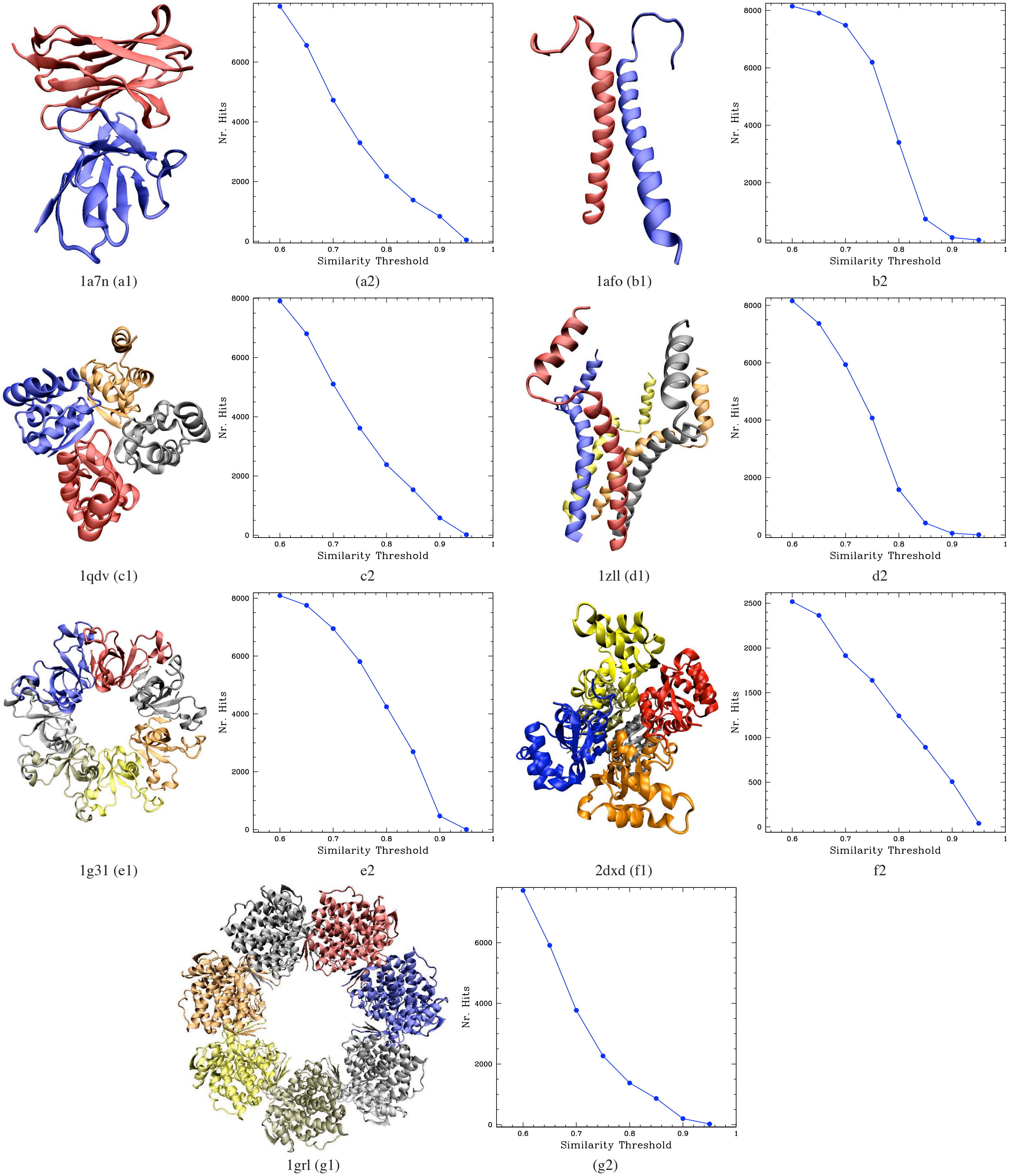
3.8 Putting it All Together: In summary, `Espreso` first extracts all symmetric homo-oligomers from the 3D complex database. The extracted subset is pre-processed to associate a representative 8d projection vector and average deviation value with each symmetric homo-oligomer. An 8d projection vector q is computed on the monomeric structure to be employed in the query. The method scans over the processed subset, computing the similarity score between q and the representative projection of each complex. If this score is within a designated threshold, the complex is proposed as viable for accommodating spatial replications of the input monomeric structure. It is worth mentioning that various thresholds from 0.6 to 1.0 are employed. The lower bound is set to 0.6, because the lowest self similarity value $\langle\epsilon\rangle$ obtained over the symmetric homo-oligomers is around 0.56.

3.9 Implementation Details: `Espreso` is implemented in Python and C++. Experiments are run on an Intel Core2 Duo machine with 4GB RAM and 2.4GHz CPU. Pre-processing of the database takes about 0.2 seconds per complex, which amounts to under 25 minutes for about $7,000$ complexes. The majority of the time, $14/25$ minutes, are spent in unzipping PDB coordinate files corresponding to the complexes. Querying the pre-processed database with a monomeric structure takes about 0.48 CPU seconds. As a result, large-throughput queries of the pre-processed database with a list of $36,512$ non-redundant PDB chains (whose extraction from the PDB is detailed in section 4) amount to 4.87 CPU hours.

4 Results

`Espreso` is first applied to propose symmetric homo-oligomeric structures on seven protein monomers of known structure. The seven monomers are chosen to span different lengths and folds. These monomers have diverse functional roles in cells. They serve as antibodies, potassium channels, kinases, and chaperones.

The monomers include the mouse monoclonal antibody D1.3 (PDB id 1a7n), a C2 dimer of β monomers of 107 amino acids (aas) each [42], the transmembric domain of human glycophorin A (PDB id 1af7o), a C2 dimer of α 40-aa monomers [43], the GP31 co-chaperonin from bacteriophage T4 (PDB id 1g31), a C7 heptamer of mostly β 107-aa monomers [44], the nucleoside diphosphate kinase (PDB id 2dxd), a D3 hexamer of α/β 154-aa monomers [45], the



1a7n (a1)

(a2)

1afo (b1)

b2

1qdv (c1)

c2

1zll (d1)

d2

1g31 (e1)

e2

2dxd (f1)

f2

1grl (g1)

(g2)

Figure 5: (a2)-(g2) shows the number of hits obtained for different query structures over varying thresholds of similarity. The query structures are shown in (a1)-(g1) in their PDB-available complexes. The ribbon representation is employed, and different colors are used to denote the different monomers.

HSP60 bacterial chaperonin GroEL (PDB id 1grl), a C7 heptamer of α/β 107-aa monomers [46], the N-terminal domain of the voltage-gated potassium channel (PDB id 1qdv), a C4 tetramer of α/β 99-aa monomers [47], and the human phospholamban (PDB id 1zll), a C5 pentamer of α 52-aa monomers [48].

The chosen monomers exist in symmetric homo-oligomeric complexes in cells, as Figs. 5(a1-g1) show through the PDB-available structures. The structure of a monomer of each of these proteins is used to query the database, as detailed in section 3. The number of complexes that Espreso proposes as viable over different thresholds of similarity is plotted in Figs. 5(a2-g2).

Figs. 5(a1)-(g1) show that Espreso proposes symmetric homo-oligomers for most monomers, even for similarity thresholds above 0.9. The transmembric domain of human glycophorin A presents a more challenging case. The number of complexes whose monomeric structures are similar to the all α monomer of this protein falls sharply with increasing similarity threshold. It is worth mentioning that most of the monomers used for the queries do not populate complexes in the 3D complex database.

Analyzing symmetries of proposed homo-oligomers reveals that Espreso captures the native symmetry classes. On the mouse monoclonal antibody D1.4, a C2 dimer in its native state, Espreso captures the C2 symmetry on 85% of the symmetric homo-oligomers proposed at the highest similarity threshold of 0.95. On the transmembric domain of human glycophorin A, a native C2 dimer, Espreso captures the C2 symmetry on 55% of the symmetric homo-oligomers proposed at the second-highest similarity threshold of 0.9 (only one complex is proposed at 0.95). On the nucleoside diphosphate kinase, a native D3 hexamer, Espreso captures the D3 symmetry on 30% of the symmetric homo-oligomers proposed at the highest similarity threshold of 0.95. The rest of the proposed complexes for this monomer split between 63% D2, 5% D5, and 2% D6 symmetries.

To assess the ability of Espreso to compute symmetric homo-oligomers at a large scale, the method is employed to query the pre-processed database of symmetric homo-oligomers with a non-redundant set of protein chains. The set is extracted from the PDB through the PISCES server [49], which makes available various pre-compiled lists culled from the January 13, 2009 version of the PDB. The list of protein chains used here is the pdbaanr list.

The pdbaanr list gives unique entries to non-redundant sequences across all PDB files. Redundant chain ids from all other PDB files are recorded at the end of the title of the representative chain entries. Representative chains are then chosen based on the highest resolution structure available (if the structure is obtained through X-ray crystallography) and the best R-values (if it is obtained through NMR). Priority is given to X-ray structures; non-X-ray structures are considered only after X-ray structures. The resulting list, employed here to represent a non-redundant subset of the PDB, contains in the end 36,512 chains.

The structure of each chain in this non-redundant set is used to query the database at various thresholds of similarity from 0.6 to 0.95. The number of hits, symmetric homo-oligomers obtained at each similarity threshold, is recorded and plotted in Fig. 6 for each of the chains. Hits obtained at different similarity thresholds are plotted in different colors.

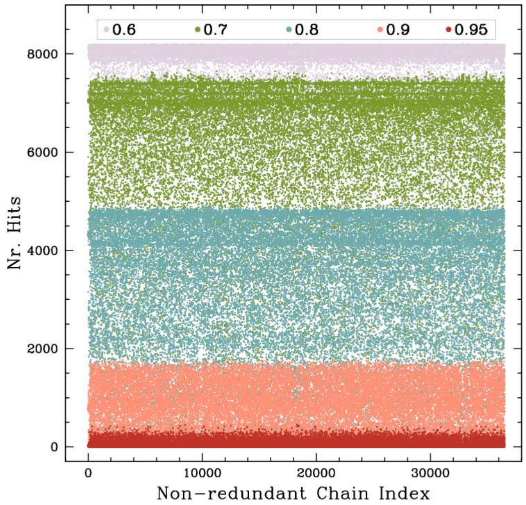


Figure 6: The x axis iterates over all 36,512 chains in the non-redundant set representative of the PDB. The y axis plots the number of symmetric homo-oligomers proposed by Espreso on each chain at different similarity thresholds of 0.6, 0.7, 0.8, 0.9, and 0.95. The number of hits obtained at the different thresholds is plotted in different colors.

Fig. 6 highlights that Espreso is non-discriminating at low similarity thresholds. However, the number of hits drops sharply with increasing threshold. The maximum number of complexes proposed for a chain drops sharply from 8206, 7666, 4952, 1816, to 447 as the similarity threshold increases from 0.6, 0.7, 0.8, 0.9, to 0.95. The number of chains for which Espreso cannot propose more than 1 complex also drops from 5, 32, 176, 1594, to 7996 with increasing similarity threshold. These results suggests that the similarity score employed by Espreso is discriminating at high thresholds and limits the number of complex geometries considered from the 3D complex database.

5 Discussion

The Espreso method proposed in this work is a first step towards designing novel complexes given the structure of a protein monomer. The method computes symmetric homo-oligomeric protein complexes by searching over a database of experimentally-determined naturally-occurring complex structures. Geometric similarity between the structure of a given monomer and the structure of the replicated monomer in a symmetric homo-oligomer is employed as the main determinant of which complex geometries can accommodate a given monomeric structure.

The geometric similarity measure proposed in this work operates on a low-dimensional space where monomeric structures are projected. The projection coordinates employed by `Espreso` build on the ones proposed in [33], but take into consideration possible fluctuations of backbones that can occur even in monomers of a symmetric homo-oligomer. The similarity measure proposed here captures well the overall geometric similarity of monomers in such complexes while filtering away noise in the projection coordinates due to possible backbone fluctuations.

The quality of the proposed complexes depends not only on the geometric similarity criterion employed in this work, but also on the richness of the 3D complex database. Various statistics compiled over this database show that the current set of deposited complex structures are dominated by cyclic homo-oligomers of small and even oligomeric number and small monomers of predominantly α/β folds. As research on protein complexes advances, the richness of databases on protein complexes is expected to increase. Further deposition of complex structures in databases will allow enlarging the search space and improving the quality of proposed symmetric homo-oligomers by `Espreso`.

Further refinement that employs more than geometric considerations may also improve the quality of the complexes proposed by `Espreso`. Since complexes are larger than single protein chains, refinements that employ physically-realistic energy functions are computationally demanding. They can be conducted by detailed studies focused on characterizing specific complexes of particular biological interest. Implementation of efficient coarse-grained scoring functions to improve the quality of proposed complexes provides a natural direction of future work.

Additional considerations beyond energy may improve prediction accuracy. Interaction interfaces in a proposed complex can be analyzed and compared to existing functional interfaces and motifs in protein structures. Interfaces that are commonly captured among protein structures may provide an additional criterion for properly ranking symmetric homo-oligomers proposed by `Espreso`.

In the context of the `Espreso` method proposed in this work, an enhanced complex database and further refinements of proposed complex structures have exciting implications for accurate in-silico design of protein complexes. Complexes designed in the dry lab that go through various refinements and filtering stages can provide good candidates for further characterization in the wet lab, where properties such as stability and biological function can be tested in biological environments. Complexes designed to function as molecular machines are of particular interest, since they can potentially impact the design of both therapeutics and functional materials with novel properties.

References

[1] D. S. Goodsell and A. J. Olson, “Structural symmetry and protein function,” *Annu. Rev. Biophys. and Biomolec. Struct.*, vol. 29, pp. 105–153, 2000.

[2] S. J. Lutdke, D. H. Chen, J. L. Song, D. T. Chuang, and W. Chiu, “Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy,” *Structure*, vol. 12, pp. 1129–1136, 2004.

[3] E. J. Dodson, “Computational biology: Protein predictions,” *Nature*, vol. 450, no. 7167, pp. 176–177, 2007.

[4] L. Stryer, “Implications of X-ray crystallographic studies of protein structure,” *Annu. Rev. Biochem.*, vol. 37, pp. 25–50, 1968.

[5] L. E. Kay, “NMR studies of protein structure and dynamics,” *J. Magn. Reson.*, vol. 173, no. 2, pp. 193–207, 2005.

[6] R. Bonneau and D. Baker, “Ab initio protein structure prediction: progress and prospects,” *Annu. Rev. Biophys. and Biomolec. Struct.*, vol. 30, no. 1, pp. 173–189, 2001.

[7] K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo, “Simultaneous determination of protein structure and dynamics,” *Nature*, vol. 433, no. 7022, pp. 128–132, 2005.

[8] A. Shehu, L. E. Kavraki, and C. Clementi, “Unfolding the fold of cyclic cysteine-rich peptides,” *Protein Sci.*, vol. 17, no. 3, pp. 482–493, 2008.

[9] ——, “Multiscale characterization of protein conformational ensembles,” *Proteins: Struct. Funct. Bioinf.*, 2009, in press.

[10] C. B. Anfinsen, “Principles that govern the folding of protein chains,” *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[11] J. W. Ponder and D. A. Case, “Force fields for protein simulations,” *Adv. Protein Chem.*, vol. 66, pp. 27–85, 2003.

[12] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, “Theory of protein folding: the energy landscape perspective,” *Annual Review of Physical Chemistry*, vol. 48, pp. 545–600, 1997.

[13] K. A. Dill and H. S. Chan, “From levinthal to pathways to funnels,” *Nat. Struct. Biol.*, vol. 4, no. 1, pp. 10–19, 1997.

[14] R. Unger and J. Moult, “Finding lowest free energy conformation of a protein is an NP-hard problem: Proof and implications,” *Bull. Math. Biol.*, vol. 55, no. 6, pp. 1183–1198, 1993.

[15] R. H. Lathrop, “The protein threading problem with sequence amino acid interaction preferences is NP-complete,” *Protein Eng*, vol. 7, no. 9, pp. 1059–1068, 1994.

[16] S. Potluri, A. K. Yan, J. J. Chou, B. R. Donald, and C. Bailey-Kellogg, “Structure determination of symmetric homo-oligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing,” *Proteins: Struct. Funct. Bioinf.*, vol. 65, no. 1, pp. 203–219, 2006.

[17] X. Wang, S. Bansal, M. Jiang, and J. H. Prestegard, “RDC-assisted modeling of symmetric protein homo-oligomers,” *Protein Sci.*, vol. 17, no. 5, pp. 899–907, 2008.

[18] J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker, “Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations,” *J. Mol. Biol.*, vol. 331,

no. 1, pp. 281–299, 2003.

[19] A. M. Ruvinsky and I. A. Vakser, “Chasing funnels on protein-protein energy landscapes at different resolutions,” *Biophys. J.*, vol. 95, no. 5, pp. 281–299, 2008.

[20] E. D. Levy, J. B. Pereira-Leal, C. Chothia, and S. A. Teichmann, “3d complex: A structural classification of protein complexes,” *PLoS. Comput. Biol.*, vol. 2, no. 11, pp. 1395–1406, 2006.

[21] D. A. Snyder, Y. Chen, N. G. Denissova, T. Acton, J. M. Aramini, M. Ciano, R. Karlin, J. Liu, P. Manor, P. A. Rajan, P. Rossi, G. V. Swapna, R. Xiao, B. Rost, J. Hunt, and G. T. Montelione, “Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination,” *J. Am. Chem. Soc.*, vol. 127, no. 47, pp. 16 505–16 511, 2005.

[22] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, H. P. H., M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. van der Vegt, and H. B. Yu, “Biomolecular modeling: Goals, problems, perspectives,” *Angew. Chem. Int. Ed. Engl.*, vol. 45, no. 25, pp. 4064–4092, 2006.

[23] C. J. Camacho, D. W. Gatchell, S. R. Kimura, and S. Vajda, “Scoring docked conformations generated by rigid-body protein-protein docking,” *Proteins*, vol. 40, no. 1, pp. 525–537, 2000.

[24] C. J. Camacho and S. Vajda, “Protein docking along smooth association pathways,” *Proc. Natl. Acad. Sci. USA*, vol. 98, no. 1, pp. 10 636–10 641, 2001.

[25] J. G. Mandell, V. A. Roberts, M. E. Pique, V. Kotlovyi, J. C. Mitchell, E. Nelson, I. Tsigelny, and T.-E. L. F., “Protein docking using continuum electrostatic and geometric fit,” *Protein Eng.*, vol. 14, no. 2, pp. 105–113, 2001.

[26] C. J. Camacho and S. Vajda, “Protein-protein association kinetics and protein docking,” *Curr. Opinion Struct. Biol.*, vol. 12, no. 1, pp. 36–40, 2002.

[27] J. Fernandez-Recio, M. Totrov, and R. Abagyan, “Soft protein-protein docking in internal coordinates,” *Protein Sci.*, vol. 11, no. 2, pp. 280–291, 2002.

[28] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliand, T. N. Bhat, H. Weissig, S. I. N., and P. E. Bourne, “The Protein Data Bank,” *Nucl. Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.

[29] P. Bradley, K. M. S. Misura, and D. Baker, “Toward high-resolution de novo structure prediction for small proteins,” *Science*, vol. 309, no. 5742, pp. 1868–1871, 2005.

[30] A. Colubri, A. K. Jha, M.-Y. Shen, A. Sali, R. S. Berry, T. R. Sosnick, and K. F. Freed, “Minimalist representations and the importance of nearest neighbor effects in protein folding simulations,” *J. Mol. Biol.*, vol. 363, no. 4, pp. 835–857, 2006.

[31] H. Gong, P. J. Fleming, and G. D. Rose, “Building native protein conformations from highly approximate backbone torsion angles,” *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 45, pp. 16 227–16 232, 2005.

[32] P. Das, M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi, “Low-dimensional free energy landscapes of protein folding reactions by nonlinear dimensionality reduction,” *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 26, pp. 9885–9890, 2006.

[33] P. J. Ballester and G. Richards, “Ultrafast shape recognition to search compound databases for similar molecular shapes,” *J. Comput. Chem.*, vol. 28, no. 10, pp. 1711–1723, 2007.

[34] A. Shehu, “Guiding a tree-based search for protein conformations in a projection space,” in *Proceedings of Robotics: Science and Systems*, 2009, submitted.

[35] M. Nilges, “A calculation strategy for the structure determination of symmetric dimers by 1H NMR,” *Proteins: Struct. Funct. Bioinf.*, vol. 17, no. 3, pp. 297–309, 1993.

[36] J. J. Gray, “High-resolution protein-protein docking,” *Curr. Opinion Struct. Biol.*, vol. 16, no. 2, pp. 183–193, 2006.

[37] J. Moult, K. Fidelis, A. Kryshtafovych, B. Rost, T. Hubbard, and A. Tramontano, “Critical assessment of methods of protein structure prediction (CASP) round VII,” *Proteins: Struct. Funct. Bioinf.*, vol. 69, no. S8, pp. 3–9, 2007.

[38] G. R. Smith and M. J. E. Sternberng, “Prediction of protein-protein interactions by docking methods,” *Curr. Opinion Struct. Biol.*, vol. 12, no. 1, pp. 28–35, 2002.

[39] J. T. Pedersen and J. Moult, “Ab initio structure prediction for small polypeptides and protein fragments using genetic algorithms,” *Proteins*, vol. 23, no. 3, pp. 454–460, 1995.

[40] G. Chikenji, Y. Fujitsuka, and S. Takada, “A reversible fragment assembly method for de novo protein structure prediction,” *J. Chem. Phys.*, vol. 119, no. 13, pp. 6895–6903, 2003.

[41] C. Wang, P. Bradley, and D. Baker, “Protein-protein docking with backbone flexibility,” *J. Mol. Biol.*, vol. 373, no. 2, pp. 503–519, 2007.

[42] C. Marks, K. Henrick, and G. Winter, “X-ray structures of d1.3 fv mutants,” to be published.

[43] K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman, “A transmembrane helix dimer: structure and implications,” *Science*, vol. 276, no. 5309, pp. 131–133, 1997.

[44] J. F. Hunt, S. M. van der Vies, L. Henry, and J. Deisenhofer, “Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage,” *Cell*, vol. 90, no. 2, pp. 361–371, 1997.

[45] M. Kato-Murayama, K. Murayama, T. Terada, M. Shirouzu, and S. Yokoyama, “Crystal structure of nucleoside diphosphate kinase in complex with atp analog,” 2007, to be published.

[46] K. Braig, Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich, and P. B. Sigler, “The crystal structure of the bacterial chaperonin GroEL at 2.8 A,” *Nature*, vol. 371, no. 6498, pp. 578–586, 1994.

[47] D. L. Minor, Y. F. Lin, B. C. Mobley, A. Avelar, Y. N. Jan, L. Y. Jan, and J. M. Berger, “The polar T1 interface is linked to conformational changes that open the voltage-gated potassium channel,” *Cell*, vol. 102, no. 5, pp. 657–670, 2000.

[48] K. Oxenoid and J. J. Chou, “The structure of phospholamban pentamer reveals a channel-like architecture in membranes,” *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 31, pp. 10 870–10 875, 2005.

[49] G. Wang and R. L. Dunbrack, “Pisces: a protein sequence culling server,” *Bioinformatics*, vol. 19, no. 12, pp. 1589–1591, 2003.