

# SCIENTIFIC REPORTS

OPEN

## Assemble-And-Match: A Novel Hybrid Tool for Enhancing Education and Research in Rational Structure Based Drug Design

Pouya Tavousi<sup>1</sup>, Reza Amin<sup>2</sup> & Sina Shahbazmohamadi<sup>3</sup>

Rational drug design is the process of finding new medication that can activate or inhibit the biofunction of a target molecule by binding to it and forming a molecular complex. Here, shape and charge complementarities between drug and target are key. To help find effective drug molecules out of a huge pool of possibilities, physical and computer aided tools have been developed. Former offers a tangible experience of the molecular interactions yet lacks measurement and evaluation capabilities. Latter enables accurate and fast evaluations, but does not deliver the interactive tangible experience of physical models. We introduce a novel hybrid model called "Assemble-And-Match" where, we enhance and combine the unique features of the two categories. Assemble-And-Match works based on fabrication of customized molecular fragments using our developed software and a 3D printer. Fragments are hinged to each other in different combinations and form flexible peptide chains, conformable to tertiary structures, to fit in the binding pocket of a (3D printed) target molecule. Through embedded measurement marks, the molecular model is reconstructed *in silico* and its properties are evaluated. We expect Assemble-And-Match tool can enable combination of visuospatial perception with *in silico* computational power to aid research and education in drug design.

Traditional drug discovery methods have a basis in trial-and-error<sup>1</sup>. In contrary, rational drug design approaches attempt to modulate a specific biological target in hope for a therapeutic value<sup>2</sup>. Typically, the ligand drug molecule must be designed such that it can attach to the binding site of the target strongly (i.e. with high affinity) and sometimes uniquely (i.e. with high specificity). This happens if there is a complimentary relationship between the drug and the target in shape (for affinity) and in charge (for specificity)<sup>3</sup>. Finding the best drug molecule candidate among a large library of possible ligands requires a way to evaluate the individual items<sup>4,5</sup>. Due to cost and time limitations, experimental screening of all possibilities is ruled out<sup>6</sup>. Therefore, several tools are employed to help narrow down the primary set of candidates. These tools that, are used both in research labs for conducting more efficient research studies and in the educational environments to raise the students' comprehension of molecular structures and interactions, can be classified into two categories of physical<sup>7,8</sup> and computer aided molecular models<sup>9</sup>.

Computational tools attempt to rank a set of either collected or enumerated candidates based on a scoring function<sup>10</sup>. The scoring function which is supposed to reflect the binding affinity between the drug and target molecules, is commonly derived semi-empirically<sup>11</sup> but can also be obtained from density functional theory<sup>12</sup> or ab-initio quantum chemistry techniques<sup>13</sup>. Scoring function can also be the result of knowledge based methods, where statistical and data mining tools are used for pattern extraction and data fitting from experimental binding affinity examples<sup>14</sup>. In a perfect world, the drug design screening tool will result in only one candidate solution. So, in turn, chemical synthesis of only one molecule will be sufficient. However in reality, because of the imperfection of the models, even after several iterations of computational evaluations, still several candidates must be chemically synthesized to compare their properties through experimental assays<sup>15</sup>.

<sup>1</sup>Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT, 06269, USA. <sup>2</sup>Department of Mechanical Engineering, University of Connecticut, Storrs, CT, 06269, USA. <sup>3</sup>Department of Biomedical Engineering, University of Connecticut, Storrs, CT, 06269, USA. Correspondence and requests for materials should be addressed to P.T. (email: [pouya.tavousi@uconn.edu](mailto:pouya.tavousi@uconn.edu))

The powerful features offered by computer aided drug design packages promise great advancements in the field. However, from the perspective of an expert pharmacist or immunologist who seeks a tangible and interactive experience of intermolecular interactions, computer tools may seem less handy in terms of the interfaces they provide (limited to computer-human communication only via keyboard, mouse and monitor<sup>16</sup>). Instead, the scientist would benefit much more from a model that, in addition to providing accurate computations and computer visualizations, also engages their visual and tactile sensory to enable them to more effectively apply their insight and experience in solving the drug design problem. In fact, although the human brain is usually slower than the computer to accomplish a certain number of mathematically well-defined tasks, but in many cases it is superior in solving complex tasks such as matching and assembling<sup>17–20</sup>. Virtual reality tools<sup>21</sup> and haptic technology<sup>22</sup> have been employed in an attempt to create environments that can provide a better sense of molecular complexes for drug designers. However, such technologies are still at their infancy and are often very costly. In a study conducted on the usage of tangible interfaces for structural molecular biology, reported in ref.<sup>23</sup>, the authors concluded that such interfaces are advantageous over solely computer visualization methods as they can produce a multisensory engagement, provide the capability of analog computation, offer a natural and intuitive mechanism for manipulation and exploration, and finally provide overview and detail simultaneously.

Although the advent of computer tools somewhat shifted the interest of scientific community from physical to the digital models, the unique insightful and interactive features of physical models keep them still in play. In fact, their popularity for qualitative studies and teaching the molecular concepts has been rising in the past few years<sup>7,24</sup>. The three main categories of physical molecular models are skeletal, ball-and-stick and space-filling. In the skeletal models, bonds and atoms are represented by rods and the intersections of them respectively. In the ball-and-stick model, bonds are rods and atoms are represented by spheres. The fact that atom radii here are significantly scaled down, makes this model incapable of representing surface and space-occupying properties of chemical molecules. The space-filling model represents atoms by spheres where the radius of sphere is proportional to the radius of atom and the distance between the centers of spheres is proportional to the distance between nuclei of atoms. Space-filling models are usually used to represent the interacting surface of the molecule with other molecules<sup>25,26</sup>. Space-filling models are also known as CPK models named after Robert Corey, Linus Pauling, and Walter Koltun who were pioneers in developing molecular models. Depending on its type, a molecular model may only represent the whole molecule as one piece or it may reflect the construction of the molecule from smaller constituents. The composite Nicholson model in which amino acid or nucleic acid residues act as the construction pieces for forming macromolecules<sup>27</sup> is a good example of the latter.

Tracing the development of physical tools in the recent years, one will observe the wide spread usage of 3D printing technology in advancing more sophisticated models<sup>28,29,30</sup>. This has been so, mainly because additive manufacturing has in one way overcome some of the challenges on the way of fabricating complex geometries, and in another way has made custom design more affordable for almost everyone. The current 3D printing technologies and the impact of 3D printed models on students' comprehension and retention of structural information are reviewed in ref.<sup>31</sup>. In one study reported in ref.<sup>32</sup>, the authors attempted to address the difficulty that students face in processing the macromolecular structures by creating four different 3D physical models followed by testing them on students in order to probe the effectiveness of these models on learning. The authors concluded that, with the use of such models, the barrier to understanding complex structures by students can be lowered.

One thing that pertains to all physical models is their role in enhancing the visuospatial ability which seems to be a critical factor in perceiving concepts of chemistry, biochemistry and the related topics. The longitudinal findings reported in ref.<sup>33</sup> support the crucial role of spatial ability in developing expertise in STEM. Also<sup>34</sup>, reports on an examination conducted on the role of visuospatial cognition in chemistry learning, in which the authors suggest that an experience in manipulating concrete models is crucial in helping students solve chemistry problems. Furthermore, the study in ref.<sup>35</sup> compares the learning of students that utilize physical models versus those that use textbooks and 3D computer models for learning imaging anatomy, showing a significant lead of the first group in terms of learning and comprehension. In another case, the controlled study on learning of students conducted in ref.<sup>36</sup> suggested that combining hand-held models with computer imaging programs helps with answering questions on molecular structure and function.

Despite the wide usage of physical molecular models for qualitative representations, they were rarely used for rigorous quantitative evaluation of molecular systems, both in education and research, until very recently. Chakroberty *et al.* demonstrated in refs<sup>16,37,38</sup> a successful attempt to build a flexible scaled physical model so called Peppyptides which provides a coarse grained representation of the peptide chain for protein folding applications. The limited flexibility is manifested through rotations that are present only in the backbone. Further, the embedded magnets in their design simulate the short-range barriers on the backbone torsion angles as well as the long-range hydrogen bonds. A marker-less tracking method has been presented in ref.<sup>39</sup> to recover the 3D structure of these models from a smart phone video. The self-assisting protein folding kit presented in ref.<sup>40</sup> is another good example of flexible models for the polypeptide backbone. With the developed model, the authors produced an idealized model of the Triosephosphate Isomerase Mutase enzyme (TIM). The model is aimed at being used by students for a step-by-step investigation into the nature of protein folding.

A similar way of thinking in structure-based drug design applications and teaching of molecular interactions seems to be missing. Development of physical models that grant flexibility of the ligand, evoke tactile and visual sensory and allow accurate measurements can make a breakthrough in the advancement of drug design tools and further, provide the students with highly interactive tools that can call upon their cognitive and visuospatial abilities toward faster and more efficient learning of the concepts.

In this work, we will report the design and verification of a novel hybrid tool, for enhancing education and research in structure-based drug design, including physical and computational components. The tool which is called Assemble-And-Match enables custom fabrication of ligand molecular fragments and the receptor molecule

using a computer framework and a desktop 3D printer. The user can then exploit their visual and tactile sensory to assemble the fragments in different combinations and alter the conformation of the peptide chain to reach at the best fit between the ligand and the receptor. They will then report the torsion angles through the embedded measurement marks to the software to reconstruct the model and evaluate various properties such as quality of the binding fit.

The Assemble-And-Match model provides a sophisticated and detailed representation that enables side chain rotations in addition to the main chain flexibilities. In addition, per selection of the user, side chains of the receptor can also be mobile. Further, the molecular fragments are not limited to those extracted from standard amino acids. Rather, the user can apply their insight to choose an arbitrary lumped group of atoms as a molecular fragment.

The complexity of assembling task is minimized by 3D printing the molecular fragments and the receptor as one piece components in such a way that no additional components are required for making connections between them (in contrast to the Peppyptide design that uses a set of screw, nut and nylon spacer for constructing a joint and to the model presented in ref.<sup>40</sup> where, chain's alpha carbons must be strung successively using an elastic monofilament). The pieces are printed in such a way that guide the user in making the right connections, eliminating the possibility of any wrong assembly. This promotes fast plug and play which in turn enables rapid testing of several different combinations of construction fragments, leading to a handy and efficient tool for drug design applications. Also, the fact that collision between atoms is automatically avoided in the physical model (as a result of interaction between solid surfaces) results in: trimming a large portion of the configuration space, screening of which takes huge computational time in computer-based drug design; and modeling molecules as impenetrable concrete objects, helping students distinguish between favorable and unfavorable interactions.

The developed software is comprised of two modules, namely design and analysis. Using the design module, CAD models of the molecular fragments and the receptor can be generated in a highly customized fashion, enabling the drug designer to tailor the tool based on their needs. Easy connection between the physical and computer models is facilitated through embedded 3D printed measurement marks, size and accuracy of which are customizable. The analysis module then reconstructs the physical model *in silico* and uses it to evaluate various properties of the molecular system including different forms of energy.

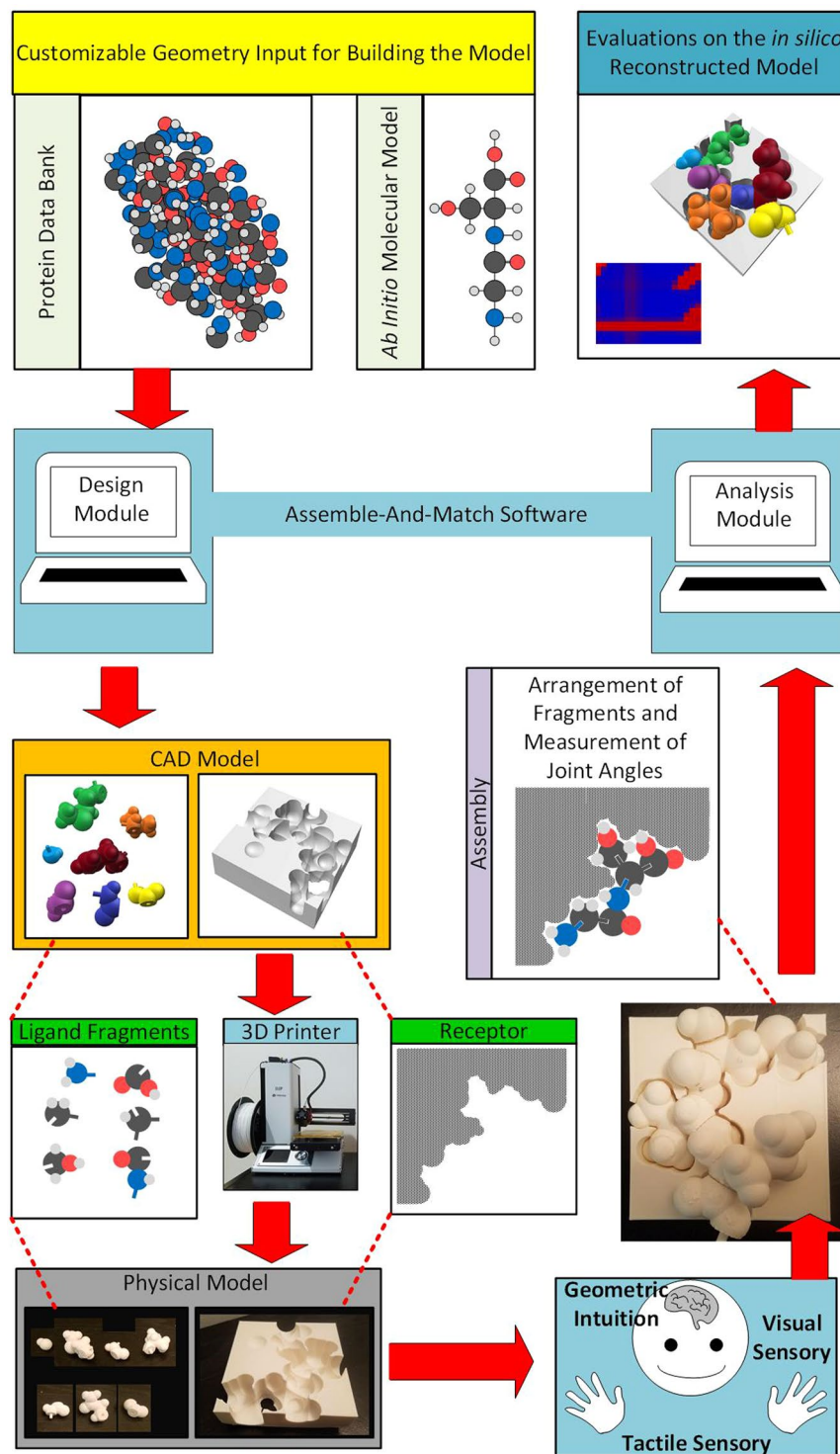
The simple design of the tool has made it easy to afford and use for almost any laboratory that works in the field, as well as any educational institute. Moreover, by exploiting the molding technology, low-cost molecular kits can be manufactured to be affordable by individual students. Figure 1 summarizes the functionality of Assemble-And-Match tool and the connection between its components.

## Material and Methods

**Molecular Model.** The Assemble-And-Match strategy for drug design is based upon the assumption that every drug molecule is made up of a finite number of building block components. One can think of these building blocks as simple LEGO pieces employed for constructing a more complex structure. Herein, we focus only on peptide-based drugs (This however, is potentially extendable to other types of drugs.). Peptides are polymeric chains formed by sequential linking of a number of building blocks so called amino acids. There are only 20 different types of amino acids naturally available. The diversity among different peptides rises from the different sequential arrangements of these amino acids, just like how only 26 letters make all the different English phrases (Fig. 2a).

Each peptide molecule consists of a set of atoms that are put together by means a number of covalent bonds along a chain. These covalent bonds have one thing in common: they restrain the motion of atoms in the molecule relative to each other. However, they do so to different extents. Some completely rigidify a part of the molecule. Some other, allow some flexibility: it is observed that rotation is allowed around certain single covalent bonds within the molecule. In fact, one can think of the peptide chain as number of rigid pieces hinged together by these single bonds. This chain can adopt different conformations as a result of the permitted rotations. These rotations are parameterized by introducing a set of torsion angles, one for each flex region in the main chain or any of the side chains<sup>41</sup>. Each amino acid contributes 3 main chain angles ( $\phi$ ,  $\psi$  and  $\omega$ ) and 0 to 4 side chain angles ( $\chi_1$  to  $\chi_4$ ) to this set. It is known that  $\omega$  always gets locked at around  $0^\circ$  or  $180^\circ$ <sup>41</sup>. Therefore the conformation of a peptide chain can be completely described only with a set of  $\phi$ ,  $\psi$  and  $\chi_i$  ( $1 \leq i \leq 4$ ) values. As a result, two things can completely determine the final shape (i.e., 3D structure) of a peptide: the sequence of amino acids and how the torsion angles are set (Fig. 2b).

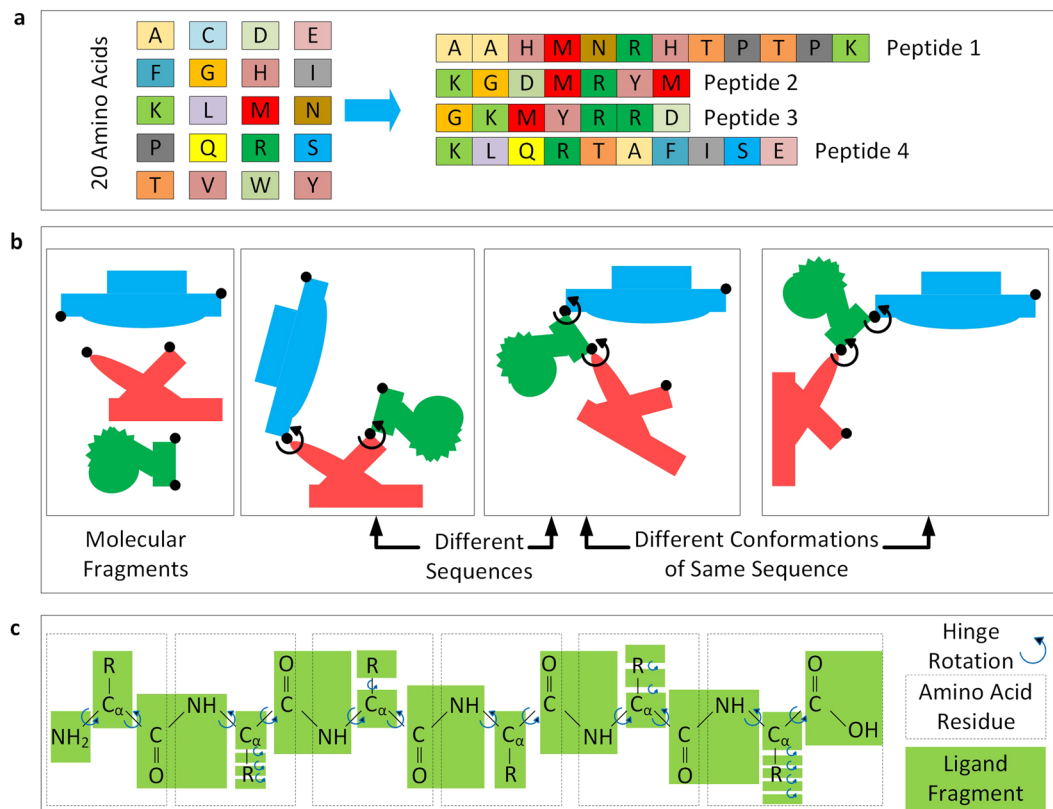
Note that although amino acids are natural building blocks of the peptide chain, they do not reflect the rigid regions of the chain molecule. In fact the ligand rigid fragments are the result of a different segmentation of the chain (Fig. 2c). Each main chain fragment can be connected to one, two or three other pieces, one on the left (closer to the N-terminus of the chain), one on the right (closer to the C-terminus of the chain), and maybe one side chain fragment. Each side chain fragment can be connected to one or two other fragments. Herein, we call these connection spots on each piece the male and female ports. The pieces are designed in such a way that, the intuitive mind can connect them to each other in the proper fashion. In fact, the revolute joints (responsible for the flexibility of the assembled molecule) are embedded in the model by means of male/female pairs, i.e., one protrusion on one piece is paired with a hole on another piece (Fig. 3b top right). For an assembly to reflect a real peptide chain, it is important that the male port of one rigid fragment is connected to only the female port of another rigid fragment and vice versa. Custom setting of shape and size of ports advocates differentiation between correct and incorrect connections and completely eliminates the possibility of any faulty assembly by the user (Fig. 3b bottom right). Measurement marks are embedded on the junction between the two pieces by which the user can read and report the associated torsion angles (Fig. 3b top left) to reconstruct the molecular system *in silico*.



**Figure 1.** The Assemble-And-Match software extracts molecular structure information from protein data bank or uses a custom *ab initio* model to generate CAD models of the receptor and ligand fragments. These are fabricated using a 3D printer. The user then applies their geometric intuition and insight to form high affinity ligand-receptor complexes. Following that, they report the favorable conformations to the software by measuring the torsion angles. The software enables evaluation of various properties of the complex.

**Printing Equipment.** In general, any 3D printing equipment can be used to fabricate the construction pieces. For the examples shown in this paper, we have used Monoprice Select Mini 3D Printer (Monoprice, CA, USA) with heated plate along with 1.75 mm ABS filament (HATCHBOX, CA, USA). The following conditions were used for printing: temperature of the heating bed = 70 °C; temperature in nozzle = 230–240 °C; layer height = 0.2 mm;





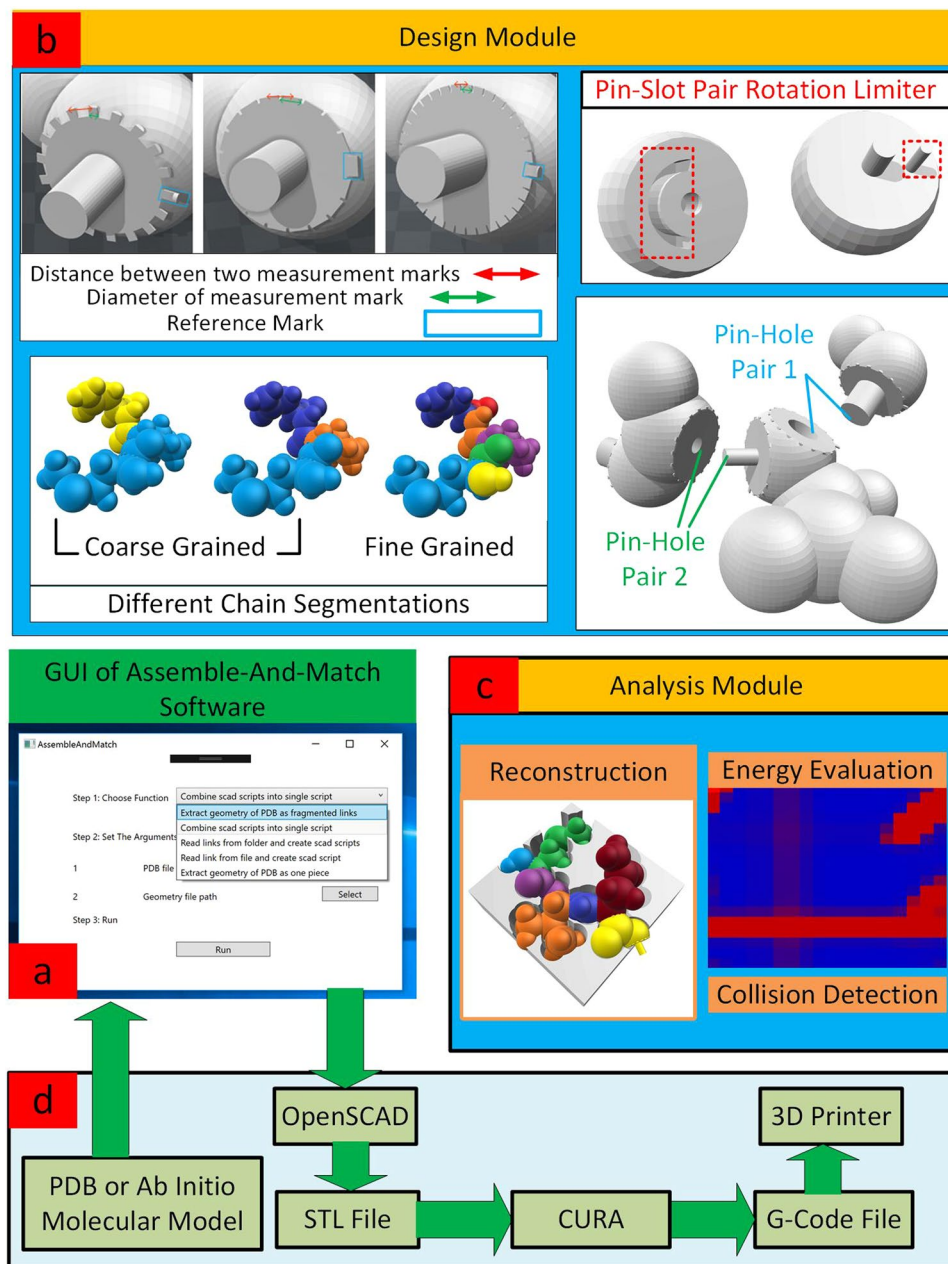
**Figure 2.** (a) Only 20 amino acids can form infinite number of different peptides. (b) Two things can completely determine the final structure of a peptide: the sequence of amino acids and how the torsion angles are set. (c) The ligand rigid fragments are distinct from the amino acid residues.

shell thickness = 0.8 mm; bottom/top thickness = 0.6 mm; fill density = 25%; print speed = 50 mm/s. Also, Kapton tape (TapesMaster, CA, USA) was adhered on top of the printing platform and to help the bottom layer of the print stick well to the bed, ABS was dissolved in acetone and rubbed using cotton to deposit a thin layer of ABS on the Kapton tape.

**Software.** The software developed for the Assemble-And-Match drug design tool consists of two modules, namely design and analysis. The design module delivers a highly customizable arrangement for architecting ligand fragments as well as molecular receptors. The user has control on the location of flexible and rigid regions, meaning that their hands are open to either use the automatic segmentation, so called as “fine fragmentation”, described in the Molecular Model section and used in refs<sup>41–44</sup>, or rather apply their structural biochemistry insight to introduce an alternative segmentation, so called as “coarse fragmentation”. An example of the latter would be the instances where almost rigid molecular domains are present in the system (Fig. 3b bottom left). They can also specify the range of rotation around each joint (Fig. 3b top right). Further, the connection between fragments is customizable in terms of parameters such as joint diameter and depth (Fig. 3b bottom right). As mentioned earlier, this is advantageous for the physical interface by helping the user avoid faulty connections. Depending on the problem at hand and the actual size of macromolecules, the scale can be adjusted before printing. One can also alter the radii of atoms to attain various representations of the molecular system for different applications ranging from skeletal to solvent accessible surface area and beyond. Further, the size and shape of the measurement marks can be adjusted by the user (Fig. 3b top left).

The computational heart of the design module, in conjunction with external database, existing open source software packages and additive manufacturing technology empowers an effective design procedure. PDB files are fed to the software to extract rigid fragments based on user custom settings or alternatively molecular fragments are generated based on a custom *ab initio* model, input by the user. The software then tailors the CAD models of the receptor and the ligand fragments based on the settings chosen by the user. This is conducted by generating SCAD scripts and feeding them to OPENSCAD<sup>45</sup> which is an open source CAD package that can render the generated pieces (i.e., ligand fragments and receptor) and export them as STL files<sup>46</sup>. The generated STL files are then converted through CURA free 3D printing software<sup>47</sup> to G-Code files (needed for the specific 3D printer used in this work) that contain the 3D printing instructions needed by the 3D printer (Fig. 3d).

The analysis module of the software uses the reported measurement by the user to reconstruct the model *in silico* and evaluate the candidate structures discovered via the physical model (Fig. 3c). Although powerful (and in some cases unique) capabilities are offered through the built-in analysis module, the user may choose to employ any other state-of-the-art computational package for their analysis job. This communication is made possible

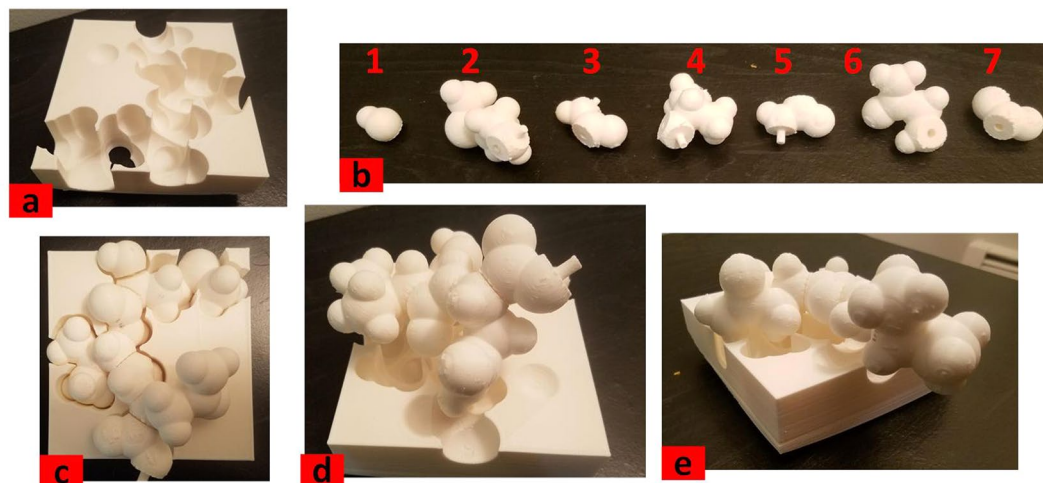


**Figure 3.** (a) A snapshot of Assemble-And-Match software user interface. (b) Highly customizable physical models promoted by the design module of the software. (c) *In silico* reconstruction and evaluation of molecular complexes enabled by the analysis module of the software. (d) Conjunction with external data, software and hardware for an effective customized design process.

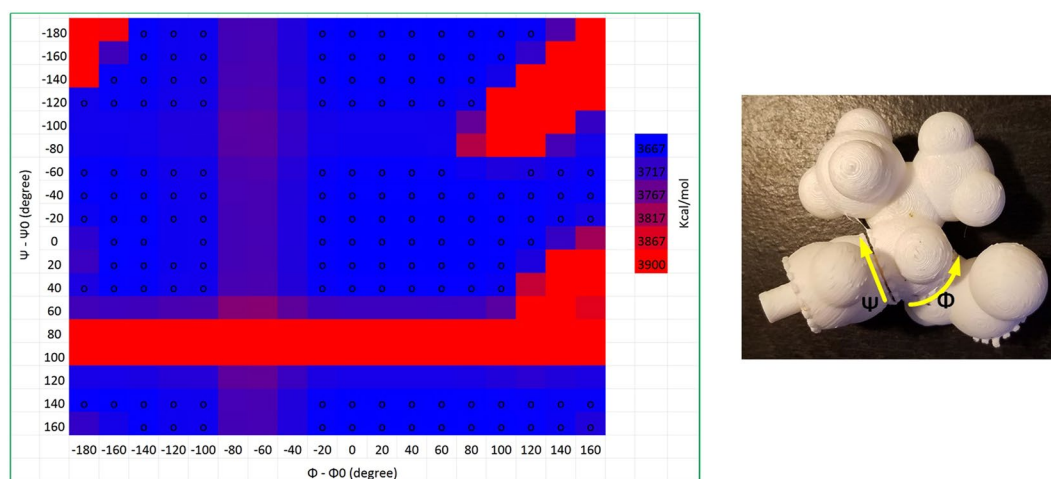
through standard molecular file types such as PDB, mol2, etc. This way, not only none of the powerful features of the computational packages are lost, but also the user can enjoy the automatic and customizable design of CAD models from the molecular database or *ab initio* molecular models.

A user friendly graphical user interface (GUI) has been developed to make it easier for the drug designer to make most of the capabilities offered by the software.

**Physical Interface.** Having the rigid receptor and the ligand fragments tailored through the software and fabricated by the 3D printer, the user, using their geometric intuition, can then select any combination of the ligand pieces, followed by putting them together to form a chain and then try to conform the chain aiming to arrive at a 3D structure that properly fills the binding pocket on the receptor. If they find a good conformation that matches the receptor, then they read the torsion angles via embedded measurement marks (we are planning to make improvements by adding a feature that the angles are read by a phone camera) and then report them along with the sequence to reconstruct the peptide structure in the software to evaluate the properties of the resulting complex. Otherwise, they will continue by trying another assemblage of the rigid fragments until they



**Figure 4.** (a) The semi-perfect template acting as receptor (b) Ligand fragments (c) Perfect match achieved by proper selection of sequence and torsion angles (d) Imperfect match due to off-value torsion angles (e) Imperfect match due to scrambled sequence.



**Figure 5.** Three-component physical model is tested at  $18 \times 18$  combinations of  $\phi$  and  $\psi$  to investigate which ones are physically plausible. Further a heat map of energy is generated using the software.

reach a good match. The insight that the user will attain during this process will help them make better choices in the following trials, both in terms of the selection of rigid fragments and their arrangement and setting of torsion angles.

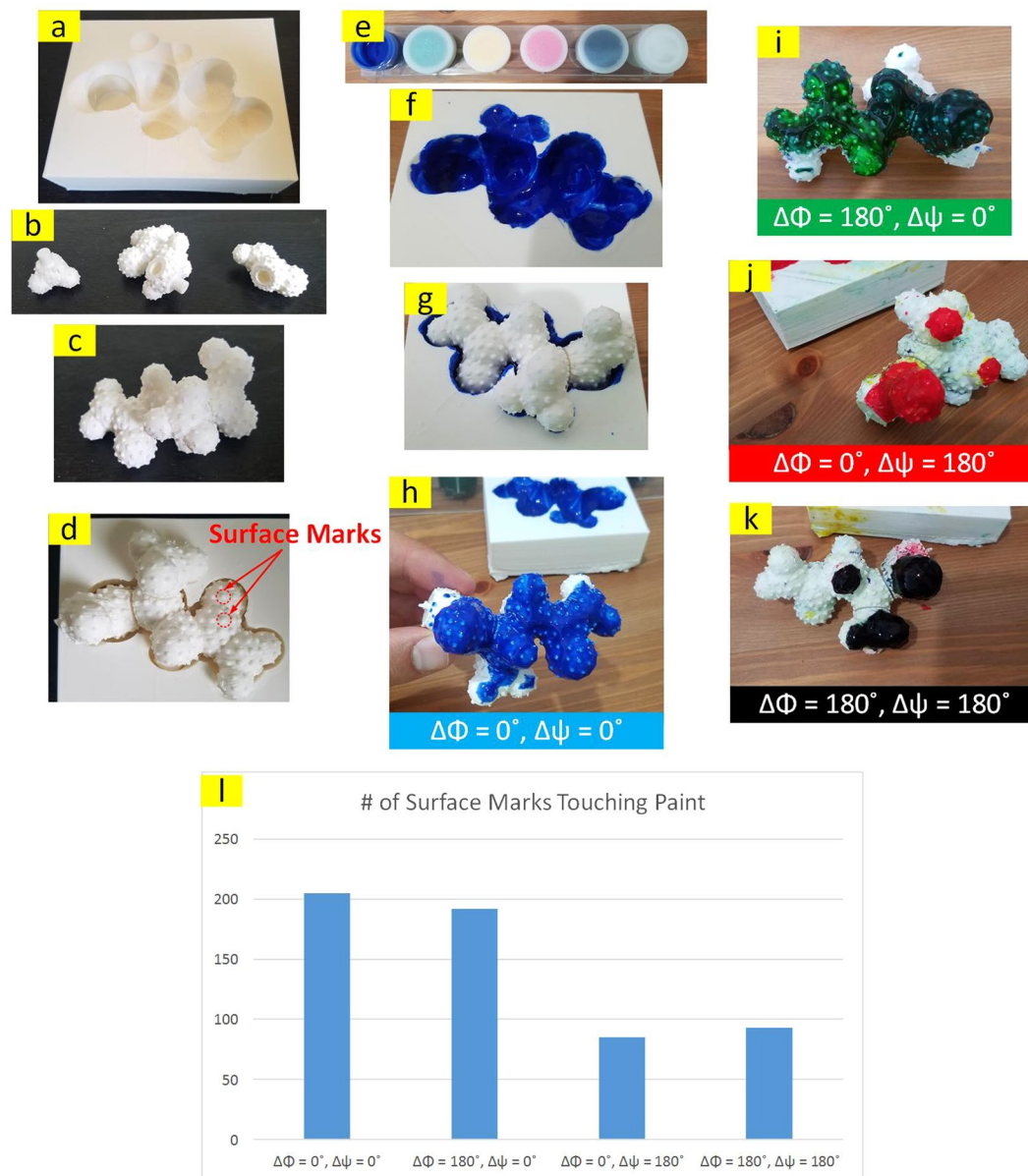
**Data availability.** The datasets analysed during the current study are available in the Protein Data Bank repository, <https://www.rcsb.org/pdb/home/home.do>.

## Results and Discussion

We devised a few test cases and demonstrations to investigate different aspects of the functionality of the developed Assemble-And-Match tool.

**Semi-Perfect Template.** The goal of the Assemble-And-Match approach is to provide the drug designer with the proper tool so they can combine their visual and tactile sensory with their insight and experience to find good ligand/receptor matches. In fact, two factors contribute to the quality of the final outcome: efficiency of the tool and how well the tool is used by the user. Thus, it is worth having a test case in which the two factors are decomposed for a more objective evaluation of the tool. For this, we have taken a portion of a random peptide chain structure from PDB (3.5 amino acids starting from the N-terminus of a protein with PDB code of 2MZU) and did the following: (a) Fabricated a “semi-perfect” template for the structure by geometric subtraction of the spherical atoms and their upward sweeps from a rectangular box that was half the size of the



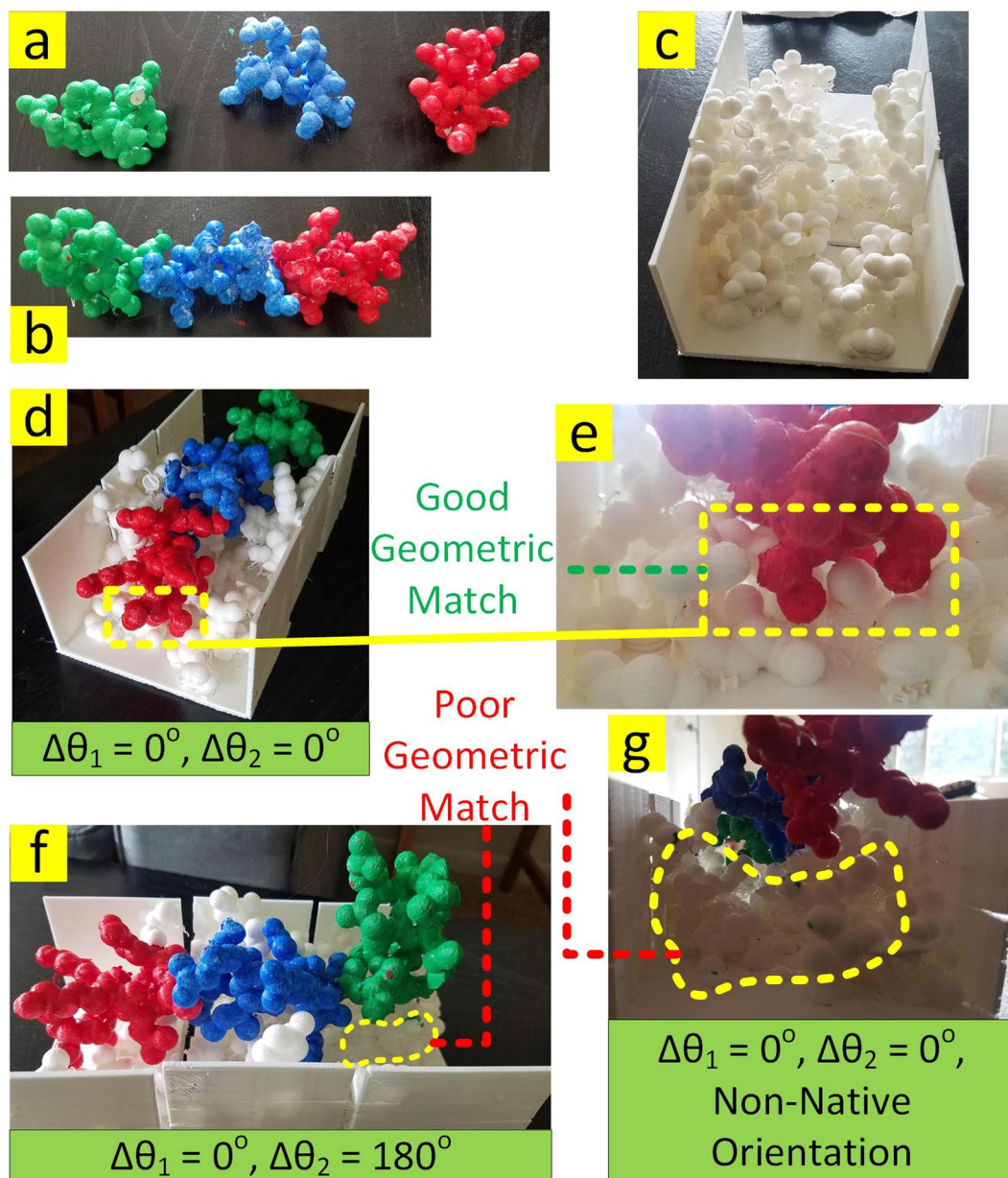


**Figure 6.** (a) 3D printed template of the ligand acting as the receptor. (b) 3D printed ligand fragments with surface marks. (c) Ligand fragments assembled at native dihedral angles extracted from PDB. (d) Geometric match of ligand and receptor. (e) Paint. (f) Painted receptor pocket. (g, h) Ligand touching the painted receptor at native dihedral angles, i.e.  $\Delta\phi = 0^\circ$  and  $\Delta\psi = 0^\circ$ . (i) Ligand touching the painted receptor at  $\Delta\phi = 180^\circ$  and  $\Delta\psi = 0^\circ$ . (j) Ligand touching the painted receptor at  $\Delta\phi = 0^\circ$  and  $\Delta\psi = 180^\circ$ . (k) Ligand touching the painted receptor at  $\Delta\phi = 180^\circ$  and  $\Delta\psi = 180^\circ$ . (l) Number of surface marks touching paint at 4 different conformations.

bounding box of all atom centers. The resulting shape acted as the receptor (Fig. 4a); (b) Segmented the peptide chain to get 7 ligand fragments (Fig. 4b). For simplicity, the side chain torsion angles were assumed to be locked; (c) Assembled the peptide chain with different sequential combinations and different settings of torsion angles and attempted to match the resulting structures with the receptor. With the right sequential order of ligand pieces and torsion angles, the perfect match between the ligand chain and the receptor was obtained (Fig. 4c). Figure 4d depicts a case that the sequential order of ligand pieces is right but the torsion angles are off  $0^\circ$ ,  $60^\circ$ ,  $40^\circ$ ,  $40^\circ$ ,  $80^\circ$  and  $40^\circ$  respectively. Similarly, Fig. 4e shows a case that the order of fragments is scrambled namely,  $1 \rightarrow 2 \rightarrow 5 \rightarrow 6 \rightarrow 7 \rightarrow 3 \rightarrow 4$  instead of  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$ . Notice that for both of these cases, a perfect match could not be achieved.

**Generating Ramachandran Plot.** One of the inherent features of the Assemble-And-Match tool is the capability of avoiding molecular conformations and assemblies that are highly unfavorable (i.e., those found at very high free energy levels) by simply taking advantage of physical collision. To investigate this property, a Ramachandran plot<sup>48</sup> was generated using the physical model and the developed software (which has the Amber

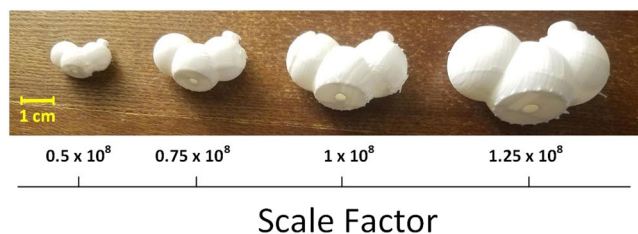




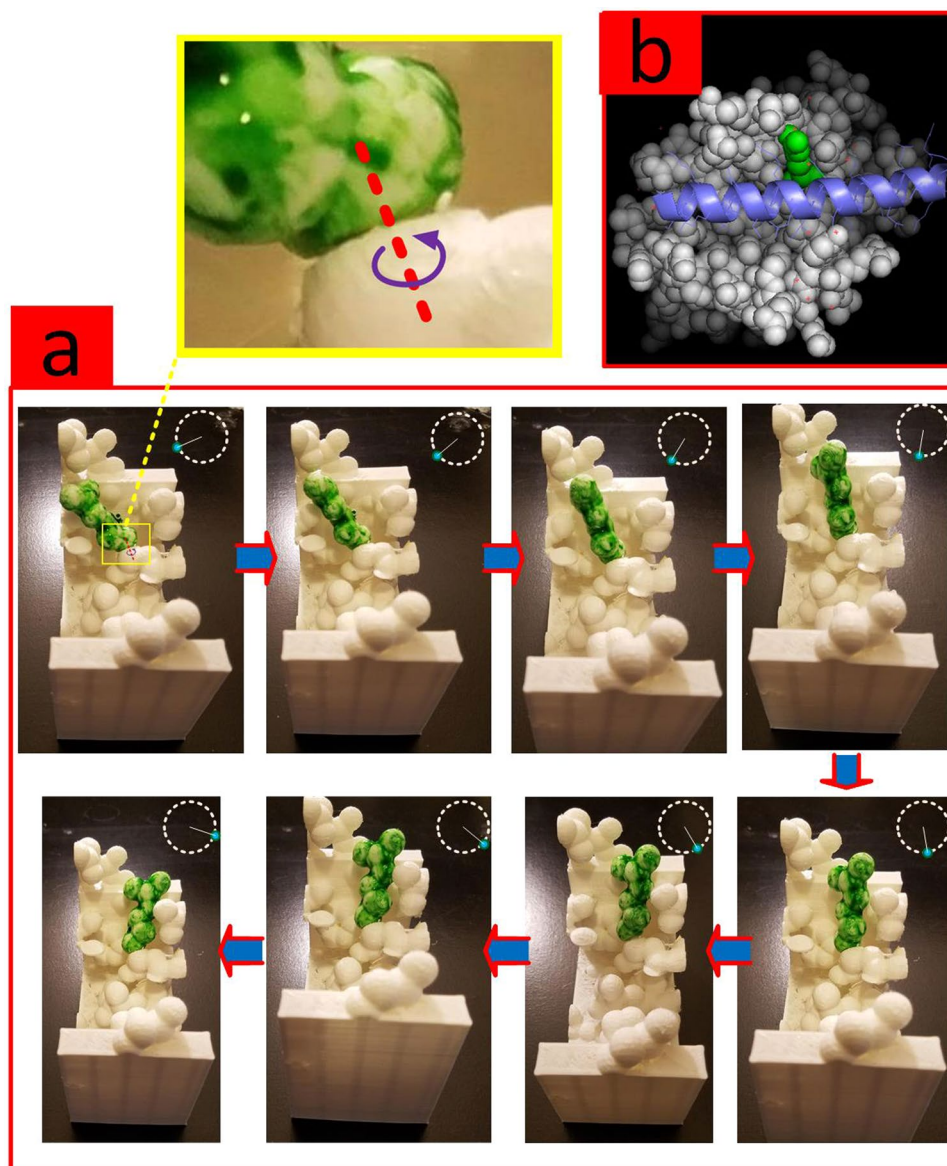
**Figure 7.** Reproducing the ligand/receptor complex found in PDB 3PL7. (a) Coarse fragmentation of the ligand. (b) Ligand chain with connected pieces. (c) Binding site of the receptor segmented into 3 pieces to make 3D printing possible on a small platform. (d) Ligand/receptor at native torsion angles and orientation. (e) Magnified view of d. (f) Non-native setting of one of the torsion angles. (g) Wrong orientation of ligand relative to the receptor.

force field embedded) to see how well the two match. The molecular fragments used for generating the plot were extracted from a protein chain with the PDB code 2MZU. The main chain dihedral angles  $\phi$  and  $\psi$  of the second amino acid (Val) were altered by increments of 20 degrees to form an  $18 \times 18$  mesh. Figure 5 overlaps the heat map of free energy obtained from the software and the physically reachable dihedral angle combinations (denoted with  $\bullet$  marks) obtained manually. A great match is observed even though atom van der Waals radii are scaled 0.7 times to expand the conformation space of the model (Note that there is a trade off between adopting atom radii near their van der Waals values and the size of the conformation space. Our experience shows a scale down of 0.7 yields a pretty handy model. Our selection is also in agreement with the claim made in ref.<sup>16</sup>). Note that the axes of the plot reflect deviations from the original  $\phi$  and  $\psi$  values (extracted from PDB) and not  $\phi$  and  $\psi$  themselves.

**Quantitative Evaluation of the Geometric Match Quality.** Finding the best (geometric) match between the ligand and receptor is key in structure based drug design. Thus, it is critical for a physical model to be capable of reflecting the match quality in a comparative fashion to differentiate between promising and poor molecular pairs. To benchmark our tool against this criterion, we extracted the starting section of a protein with



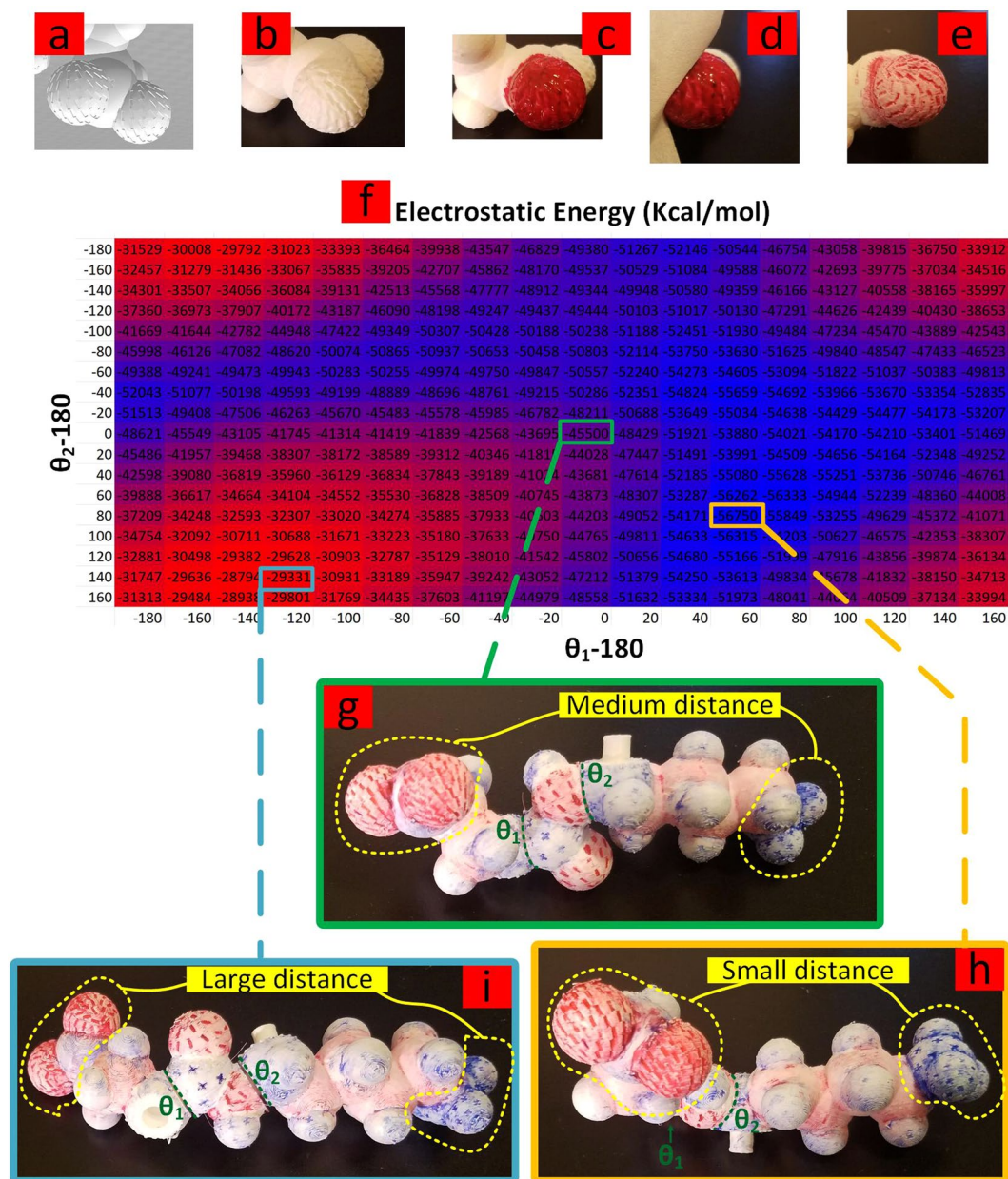
**Figure 8.** Scale factor is adjustable by the user based on their needs. Here, the same ligand fragment is printed with 4 different scale factors. The value of the scale factor reflects the ratio of a distance in the physical molecular model to the corresponding distance in the real molecule.



**Figure 9.** (a) Side chain mobility of amino acid R139 (in green) of the target through rotation about the covalent bond between  $\alpha$  and  $\beta$  carbons is shown. (b) The ligand/receptor complex in PyMOL<sup>51</sup>. Receptor, R139 and ligand are shown in white, green and blue respectively.

PDB code 1L2Y to form a 3-piece ligand. Approximately equally distanced surface marks were embedded on the spherical atoms as a means of representing surface area in a discrete fashion. We also made the semi-perfect template (as described above) as the receptor for this ligand (such that perfect match occurs between ligand and

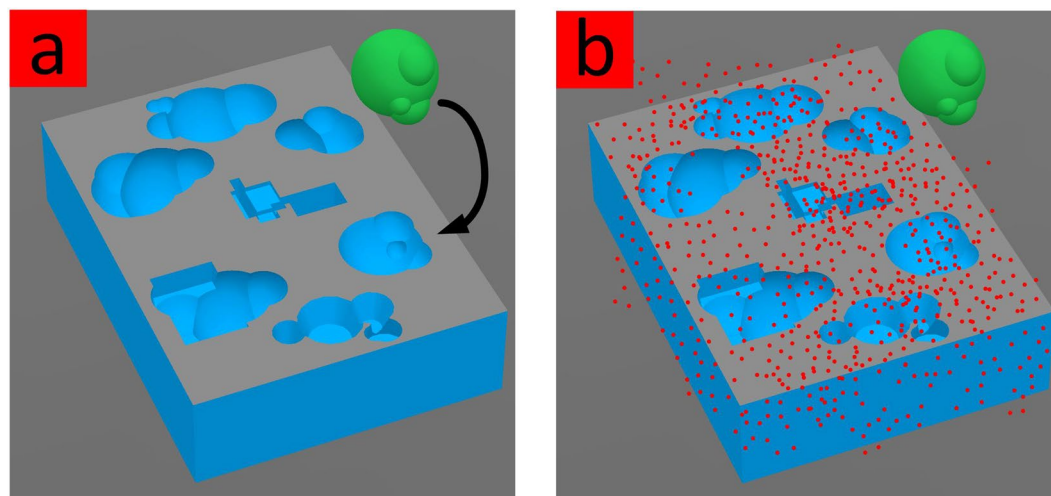




**Figure 10.** (a) Automatically embedded charge marks on the atoms' surfaces in the CAD model. (b) Charged marks in the 3D printed model. (c) Paint used to color a partially negatively charged atom. (d) Wiping paint from the surfaces. (e) Charged marks remained colored after wiping the paint. (f) The heat map of electrostatic energy generated for a 3-piece chain consisting of one negatively charged side chain (GLU) and one positively charged side chain (LYS), at  $pH=7$ , flexible via two main chain angles of  $\theta_1$  and  $\theta_2$ ; Different levels of electrostatic energy for relatively (g) medium, (h) small, and (i) large distances of the two regions in which the opposite charges are concentrated.

receptor at the PDB torsion angles). This was followed by measuring and comparing the contact area between the ligand and the receptor at different conformations (i.e., different selections of the two torsion angles). For that the binding site on the receptor was painted with a water based color and then each conformer of the ligand was placed inside the pocket. For the sake of having a reasonable comparison, the third fragment's position and orientation were preserved as much as physically possible. Needless to say, the parts of the ligand that were in contact with the receptor's pocket got painted. The surface area that was painted was measured by counting the number of surface marks that were colored (Fig. 6). It is observed from Fig. 6 that the native conformation (i.e.,  $\Delta\phi=0^\circ$  and  $\Delta\psi=0^\circ$ ) has resulted in the largest contact area which is in agreement with our expectations.

**An Existing Ligand/Receptor Example.** An existing ligand/receptor example, with the PDB code of 3PL7, was used as a benchmark for Assemble-And-Match tool to examine its capability of effectively representing the real world molecular complexes. The ligand was coarsely fragmented into 3 pieces where each piece contained



**Figure 11.** (a) Instantaneous recognition of the potential match by the cognitive mind. (b) Computer screening through extensive sampling.

about 7 amino acids (Fig. 7a,b). Also, the binding site on the receptor was located and then segmented into 3 thirds to make the 3D printing possible with the small 3D printing platform available for this project (Fig. 7c). Despite the fact that the hydrogen atoms were not included in the model, the atom radii were scaled 0.7 times and some printer errors were present for this example, it is observed that for proper selection of the two torsion angles (between pieces 1 and 2 as well as pieces 2 and 3), and correct orientation of ligand and receptor, a reasonable geometric match will be attained (Fig. 7d,e). This is while non-native setting of even one torsion angle (Fig. 7f) or choosing the wrong orientation between ligand and receptor (Fig. 7g) leads to poor geometric matching.

**Size Customization.** Depending on the size of the receptor molecule and the number of fragments that construct the ligand, the option of adjusting the scale factor is provided to the user. Figure 8 demonstrates the same ligand fragment printed in 4 different scales. As a result, unlike the prebuilt models (e.g., commercial ball and stick models), in which, hand-manipulation of the model becomes substantially difficult when molecular size increases over a certain extent, using Assemble-And-Match, even very large molecules can still be handled conveniently.

**Mobility of Receptor's Side Chains.** Enabled by the software, the side chains of the target molecule can be optionally mobile. Figure 9 illustrates the mobility of a side chain from the target molecule in the ligand/receptor complex with PDB code 3PL7.

**Charged-Charged Interactions.** To provide a visual guide for electrostatic interactions to the user, when manipulating the physical model, the software can embed charge marks (+ or -) on the spheres that represent atoms, in a way that the population of the charge marks reflects the partial charge of the corresponding atom. After printing and following a simple painting/wiping step performed by the user, using red and blue colors, representing negative and positive charges respectively (this step can be eliminated if multi-material 3D printing is employed), the physical model now helps the user get a qualitative sense of favorable and unfavorable electrostatic energies. Figure 10, shows a 3-piece chain consisting of one negatively charged side chain (GLU) and one positively charged side chain (LYS), at  $pH=7$ , flexible via two main chain angles of  $\theta_1$  and  $\theta_2$ . The heat map of electrostatic energy has been generated for an  $18 \times 18$  mesh of the two angles. The correlation between the distance of the two regions in which the opposite charges are concentrated, namely the protonated/deprotonated parts of the two side chains (indicated by yellow dotted lines in Fig. 10), and its effect on the overall electrostatic energy can be observed through the 3 conformations shown in Fig. 10. It is worth emphasizing however, that, such qualitative measure is integrated by accurate computations provided by the software at any reported conformation to yield the best results. As a manifestation of the easy connection with other software, the molecular coordinates of this example were first generated via Avogadro software<sup>49</sup>, and then fed to Assemble-And-Match software using mol2 data file type.

Also, enabled by the software, the printed models can visually guide the user towards proper formation of hydrogen bonds, in terms of distance and direction.

### Limitations and Future Directions

Despite several automated features of the tool, there are certain cases in which users' manual action is still required, some of which are subject to automation in the following versions. These include but are not limited to reporting (in the software) the peptide sequence and the torsion angles as well as the spatial pose of the final conformation of the ligand with respect to the receptor to establish instantaneous communication between physical and *in silico* models. Special mechanisms and visual tools are currently under development to address these



areas. Moreover, we are working on a novel analog computing method for evaluating the quality of matches which can make a breakthrough in the area of shape complementarity, in general and structure-based drug design, in particular.

Assemble-And-Match serves as a powerful tool for students and instructors for better comprehension and teaching of biochemistry concepts and for researchers to enjoy a tangible interface while benefiting from accurate *in silico* computations. Although, the educational aspect might be manifested more strikingly in the current version, the prospective tool is expected to provide more efficient screening processes by exploiting the cognitive abilities of humans. Inspired by Foldit<sup>50</sup>, a crowd-sourcing computer game which enables its players to contribute to solving the protein folding problem, we envision that Assemble-And-Match can become a crowd-sourcing physical game for drug discovery purposes. Perhaps, the exaggerated example illustrated in Fig. 11 can show how the idea of combining human cognition with computer power can be seeded towards establishing more effective drug design methods in the future.

## Conclusion

We reported the design and validation of a hybrid tool for rational structure-based drug design, Assemble-And-Match, consisting of computational and physical modules. The fact that the physical model is highly tangible and interactive and yet can be measured and evaluated using software has empowered the combination of human sensory and intuition with the *in silico* computational power for solving the drug design problem. The developed tool is highly detailed and customizable. More so, its plug-and-play feature supports rapid investigation of various molecular combinations, which in fact is ideal for structure based drug design applications. The conducted tests on the first version of our design promise a tool that can be used, as is, by the experts in the field of drug design, for instance, by providing a more convenient way of communication between experimental and computational scientists through a model that is physically tangible and yet, offers accurate computations. Further, the tool has a great potential to be enhanced such that it can also be used by non-experts in a way that it promotes democratization of drug design.

## References

1. Kaul, P. N. Drug discovery: past, present and future. In *Progress in drug research*, 9–105 (Springer, 1998).
2. Todd, A., Anderson, R. & Groundwater, P. W. Rational drug design-identifying and characterising a target. *The pharmaceutical journal*. **283**, 19–20 (2009).
3. Anderson, A. C. The process of structure-based drug design. *Chemistry & biology* **10**, 787–797 (2003).
4. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery* **3**, 935–949 (2004).
5. Horton, N. & Lewis, M. Calculation of the free energy of association for protein complexes. *Protein Science* **1**, 169–181 (1992).
6. Lybrand, T. P. Ligand-protein docking and rational drug design. *Current opinion in structural biology* **5**, 224–228 (1995).
7. Gardner, A. Scripps physical model service <http://models.scripps.edu/> (2017).
8. Roth, E., Nickel, A. & Herman, T. Molecular models (2005). <https://www.google.com/patents/US20050227213>. US Patent App. 11/089,430.
9. Rahman, M. M. *et al.* Use of computer in drug design and drug discovery: A review. *International Journal of Pharmaceutical and Life Sciences* **1** (2012).
10. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Bioinformatics* **47**, 409–443 (2002).
11. Murray, C. W., Auton, T. R. & Eldridge, M. D. Empirical scoring functions. ii. the testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of bayesian regression to improve the quality of the model. *Journal of computer-aided molecular design* **12**, 503–519 (1998).
12. Antony, J., Grimme, S., Liakos, D. G. & Neese, F. Protein–ligand interaction energies with dispersion corrected density functional theory and high-level wave function based methods. *The Journal of Physical Chemistry A* **115**, 11210–11220 (2011).
13. Raha, K. & Merz, K. M. A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. *Journal of the American Chemical Society* **126**, 1020–1021 (2004).
14. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology* **295**, 337–356 (2000).
15. Schneider, G. & Fechner, U. Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery* **4**, 649–663 (2005).
16. Chakraborty, P. & Zuckermann, R. N. Coarse-grained, foldable, physical model of the polypeptide chain. *Proceedings of the National Academy of Sciences* **110**, 13368–13373 (2013).
17. Khatib, F. *et al.* Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature structural & molecular biology* **18**, 1175–1177 (2011).
18. Eiben, C. B. *et al.* Increased diels-alderase activity through backbone remodeling guided by foldit players. *Nature biotechnology* **30**, 190–192 (2012).
19. Von Ahn, L. Human computation. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, 1–2 (IEEE, 2008).
20. Moravec, H. When will computer hardware match the human brain. *Journal of evolution and technology* **1**, 10 (1998).
21. Norrby, M., Grebner, C., Eriksson, J. & Bostrom, J. Molecular rift: virtual reality for drug designers. *Journal of chemical information and modeling* **55**, 2475–2484 (2015).
22. Zonta, N., Grimstead, I. J., Avis, N. J. & Brancale, A. Accessible haptic technology for drug design applications. *Journal of molecular modeling* **15**, 193–196 (2009).
23. Gillet, A., Sanner, M., Stoffler, D. & Olson, A. Tangible interfaces for structural molecular biology. *Structure* **13**, 483–491 (2005).
24. Center for biomolecular modeling, milwaukee school of engineering <http://cbm.msoc.edu/> (2017).
25. Dori, Y. J. & Barak, M. Virtual and physical molecular modeling: Fostering model perception and spatial understanding. *Educational Technology & Society* **4**, 61–74 (2001).
26. Dori, Y. J. Cooperative studyware development of organic chemistry module by experts, teachers, and students. *Journal of Science Education and Technology* **4**, 163–170 (1995).
27. LABQUIP. Labquip <http://www.labquip.clara.net/> Protein Kits. (2017).
28. Chakraborty, P. Physical biomodeling: a new field enabled by 3-d printing in biomodeling. *arXiv preprint arXiv:1502.04687* (2015).
29. Amin, R. *et al.* 3D-printed microfluidic devices. *Biofabrication* **8**(2), 022001, <https://doi.org/10.1088/1758-5090/8/2/022001> (2016).

30. Amin, R., Joshi, A. & Tasoglu, S. Commercialization of 3D-printed microfluidic devices. *Journal of 3D Printing in Medicine* **1**(2), 85–89, <https://doi.org/10.2217/3dp-2016-0010> (2017).
31. Stone-Sundberg, J., Kaminsky, W., Snyder, T. & Moeck, P. 3d printed models of small and large molecules, structures and morphologies of crystals, as well as their anisotropic physical properties. *Crystal Research and Technology* **50**, 432–441 (2015).
32. Cooper, A. K. & Oliver-Hoyo, M. Creating 3d physical models to probe student understanding of macromolecular structure. *Biochemistry and Molecular Biology Education* (2017).
33. Wai, J., Lubinski, D. & Benbow, C. P. Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* **101**, 817 (2009).
34. Wu, H.-K. & Shah, P. Exploring visuospatial thinking in chemistry learning. *Science education* **88**, 465–492 (2004).
35. Preece, D., Williams, S. B., Lam, R. & Weller, R. Let's get physical: advantages of a physical model over 3d computer models and textbooks in learning imaging anatomy. *Anatomical sciences education* **6**, 216–224 (2013).
36. Harris, M. A. *et al.* A combination of hand-held models and computer imaging programs helps students answer oral questions about molecular structure and function: a controlled investigation of student learning. *CBE-Life Sciences Education* **8**, 29–43 (2009).
37. Zuckermann, R., Chakraborty, P. & Derisi, J. Peptptides: Interactive models of polypeptide chains. Tech. Rep., LBNL (Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States)) (2014).
38. Chakraborty, P. A computational framework for interacting with physical molecular models of the polypeptide chain (2014).
39. Martinez, X., Ferey, N., Vezien, J.-M. & Bourdot, P. Virtual structure reconstruction and energy estimation of a peptide from a physical tangible interface. In *Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR)*, 2015 IEEE 1st International Workshop on, 41–42 (IEEE, 2015).
40. Davenport, J. *et al.* A self-assisting protein folding model for teaching structural molecular biology. *Structure* **25**, 671–678 (2017).
41. Tavousi, P., Behandish, M., Ilies, H. T. & Kazerounian, K. Protoloid ii: Enhanced model and implementation for kinetostatic protein folding. *Journal of Nanotechnology in Engineering and Medicine* **6**, 034601 (2015).
42. Tavousi, P., Kazerounian, K. & Ilies, H. Synthesizing functional mechanisms from a link soup. *Journal of Mechanical Design* **138**, 062303 (2016).
43. Tavousi, P., Behandish, M., Kazerounian, K. & Ilies, H. T. An improved free energy formulation and implementation for kinetostatic protein folding simulation. ASME Paper No. DETC2013-12671, <https://doi.org/10.1115/DETC2013-12671> (2013).
44. Behandish, M., Tavousi, P., Ilies, H. T. & Kazerounian, K. GPU-Accelerated Computation of Solvation Free Energy for Kinetostatic Protein Folding Simulation. In ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. V02AT02A009-V02AT02A009). *American Society of Mechanical Engineers*. <https://doi.org/10.1115/DETC2013-12675> (2013).
45. Kintel, M. & Wolf, C. Openscad. GNU General Public License, p GNU General Public License (2014).
46. Roscoe, L. *et al.* Stereolithography interface specification. *America-3D Systems Inc* **27** (1988).
47. Ultimaker. Cura. Open source software <https://ultimaker.com/en/products/cura-software> (2017).
48. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology* **7**, 95–99 (1963).
49. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8 (2015).
50. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics* **4**, 17 (2012).
51. Cooper, S. *et al.* The challenge of designing scientific discovery games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games, FDG' 10*, 40–47, <https://doi.org/10.1145/1822348.1822354> (ACM, New York, NY, USA, 2010).

## Author Contributions

Pouya Tavousi: Contributed to the development of idea and writing of the manuscript. Reza Amin: Contributed to the development of idea and writing of the manuscript. Sina Shahbazmohamadi: Contributed to the development of idea and writing of the manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018