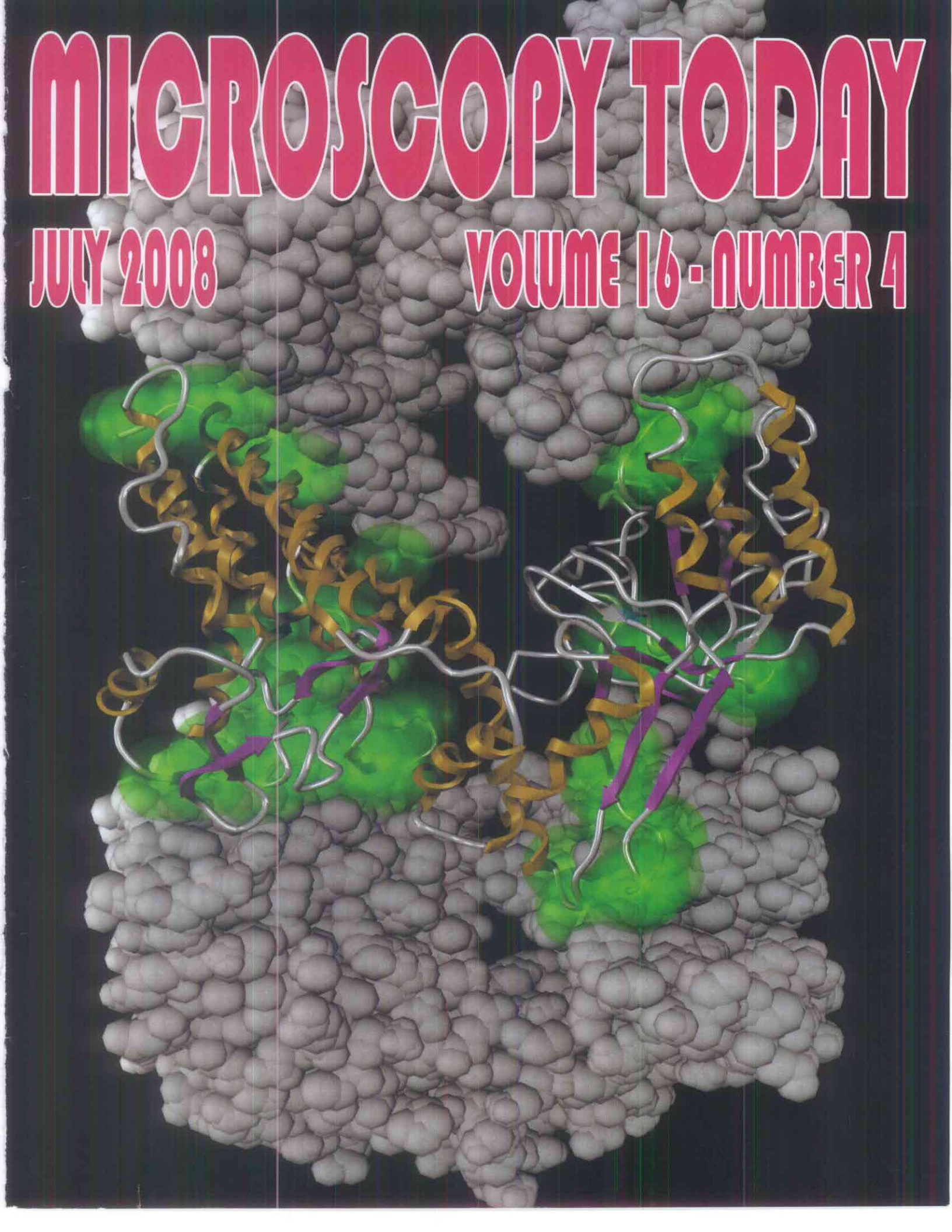
MICROSCOPY TODAY

JULY 2008 VOLUME 16 - NUMBER 4



Beyond the Black Box: Interactive Global Docking of Protein Complexes

Jochen Heyd, Stefan Birmanns
University of Texas at Houston, School of Health Information Sciences, Houston, Texas
Stefan.Birmanns@uth.tmc.edu

1. Introduction

A key to understanding the function of biological systems is the visualization of their natural state, ideally in a natural environment. At a molecular level, this is challenging. Traditional experimental techniques, like X-ray crystallography, can provide the atomic structure of proteins, but only by removing them from their native surroundings and forcing them into crystals. Over the past decade, microscopy techniques have emerged as alternatives to these traditional structure determination methods, with the advantage of visualizing molecules in a near-native state. Given the current focus of structural biology on interactions between proteins and better understanding of large protein complexes, cryo-electron microscopy (cryo-EM) has become a valuable tool [1]. Both image acquisition techniques and the computational synthesis of 3D volumetric models from micrographs have advanced considerably. 3D reconstructions of large protein complexes or even individual proteins can now be obtained (Figure 1). While cryo-EM thus offers numerous advantages (small sample size, no need to crystallize, no packing effects, *etc.*), its main drawback is its inability to attain atomic resolution. Other techniques, such as cryo-electron tomography (cryo-ET, [2]) or small-angle X-ray scattering (SAXS, [3]) also yield volumetric reconstructions of proteins in near-native states, but with even lower resolution than cryo-EM.

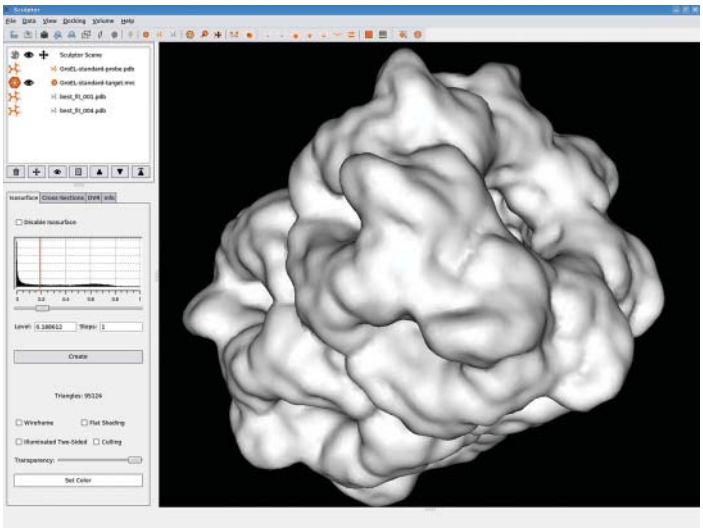


Figure 1: The interface of the Sculptor visualization package, showing a 3D volumetric reconstruction (from cryo-EM) of a GroEL chaperonin.

The failure to achieve atomic resolution with cryo-EM, cryo-ET or SAXS is not as big a stumbling block as one might expect. Often, crystal structures or good homology models are available for individual subunits. These known atomic structures can be docked into experimental volumetric reconstructions, yielding an atomic model of the whole complex. The present report first briefly discusses existing fitting methods and then describes our novel technique.

The current multi-resolution docking approaches generally fall into two categories: 1. Purely user-guided, interactive docking in a visualization program; 2. Exhaustive search based, non-interactive software packages. Each approach has different strengths and weaknesses. The existing user-guided techniques allow a biologist to directly apply his or her knowledge about the system under study. On the other hand, this interactive procedure is highly subjective and the software does not support the user in any way. Manipulating a protein in six dimensions (6D, three translations and three rotations) is non-trivial and the best docking solutions are not necessarily evident. The situation is reversed for exhaustive search based approaches. Here, the 6D search is handled in software, eliminating this tedious process for the user. However, all docking criteria need to be incorporated into a single scoring function, which is used during the exhaustive search. An ideal scoring function would include a large variety of indicators to distinguish correct docking solutions from the rest. In practice, only a single indicator is usually employed, with cross-correlation [4] and feature vector deviations [5] being the most widely used. Besides the scoring function itself, exhaustive search methods encounter a second problem: How are candidate solutions picked? Solely relying on high docking scores is only feasible at relatively high resolutions. At lower resolutions, many almost equivalent solutions exist and the user is then inundated by hundreds of solutions for a protein complex containing only a few proteins.

Our current research focuses on combining these two diametrically opposed docking techniques. Exhaustive search techniques generally only produce a set of candidate solutions but some approaches can easily yield a 3D field of fitting scores. Such a field is amicable to both visualization and further feature extraction. The present report details a first attempt to combine an exhaustive search based scoring field with monomer distance information to allow an intuitive and visual exploration of a given docking problem.

2. Interactive Global Docking

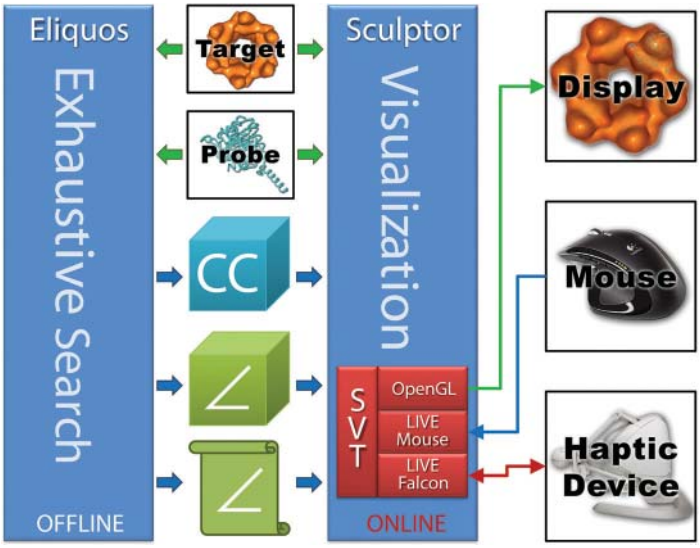


Figure 2: The overall software architecture of the interactive global docking system.

Interactive global docking (IGD) consists of two separate steps: First, an offline exhaustive search is performed in Eliquos, our cross-correlation based exhaustive search software. This typically takes from a few minutes to a few hours, depending on the

size of the system. The exhaustive search produces a vector field containing fitting scores and orientations of the probe structure, *i.e.* a scoring field. This field is read into our Sculptor visualization package and the user interactively explores both the scoring field and any additional docking information generated on-the-fly by Sculptor. Figure 2 shows an outline of the software architecture and the sections below discuss these two steps in more detail.

2.1 Exhaustive search

Eliquos, which performs the off-line exhaustive search, is a general, back box, multi-resolution docking application. It is geared towards fast and accurate docking of large data sets, using a scoring function based on cross-correlation. Several advanced filtering techniques are available to enable high-accuracy docking at both high and low resolutions. Like most exhaustive search docking programs, Eliquos is a black box in the sense that once the user specifies the input parameters, no further interaction occurs and the software will output a set of solutions guided solely by the chosen scoring function. For a detailed description, please see Ref. [4].

Eliquos performs a standard, FFT accelerated, exhaustive evaluation of the scoring function based on the user's choice of unfiltered or filtered (Laplacian or Hessian) cross-correlation. Traditionally, the top scoring candidate solutions are further refined and then written out. In the case of IGD, Eliquos instead generates several files containing the scoring field. These files consist of two volume data sets and one file with an index of sampled orientations. The first volume file contains the scalar component of the scoring field, *i.e.* the score of the best solution at each position. The second volume holds the corresponding orientation in the form of an index. Thus, the Sculptor visualization package can read in and use this scoring field as one of the docking criteria supporting the user's fitting decisions.

2.2 Interactive visualization

The interactive stage of IGD is performed in the Sculptor visualization package. Sculptor (Figure 1) is aimed primarily at working with volumetric data sets and docking of high resolution structures into these 3D reconstructions. It makes extensive use of hardware graphics acceleration to provide high performance visualization of large data sets. Some of the provided features are: volume manipulation tools, isosurface and direct volume rendering, feature-based multi-resolution docking using vector quantization, fast flexible fitting based on interpolation, haptic (force-feedback) rendering during docking, and cross-correlation based refinement of approximately fitted structures.

In addition to pure visual feedback during docking, Sculptor also supports haptic rendering, *i.e.* force-feedback to the user about the quality of the current docking position [6]. Until now, this force feedback was only based on a simplified cross-correlation score due to CPU time constraints. In order to provide smooth forces for the user, the forces need to be updated 1000 times per second. Thus, it is not possible to employ advanced filtering techniques during the cross-correlation calculation. By using a precomputed scoring field, scoring functions of arbitrary complexity can be used and the force calculation simplifies to a table lookup and trivial interpolation. In addition, the orientations contained in the field allow the software to rotate the probe structure into the best orientation at the current point in space. The user thus only has to translate the structure, which greatly simplifies his task. Limiting the interactive search space to 3D also has the advantage that inexpensive haptic devices can be used for the force feedback. Novint, for example, offers the Falcon 3D device priced below $200. While the Falcon was developed for immersive 3D gaming, it is ideally suited for our docking approach.

During the interactive stage of IGD, Sculptor not only uses the scoring field generated by Eliquos, but also generates additional, on-the-fly, docking information which the user can draw on to find ideal docking positions. This additional information is based on steric interactions between candidate solutions for different monomers in the protein complex. Whenever a new potential solution is found by the user, Sculptor calculates a distance map which contains the distance from any point in space to the closest existing solution. When the probe structure is moved, this map is consulted to quickly determine the existence of a steric clash or a good protein-protein contact.

3. IGD in practice

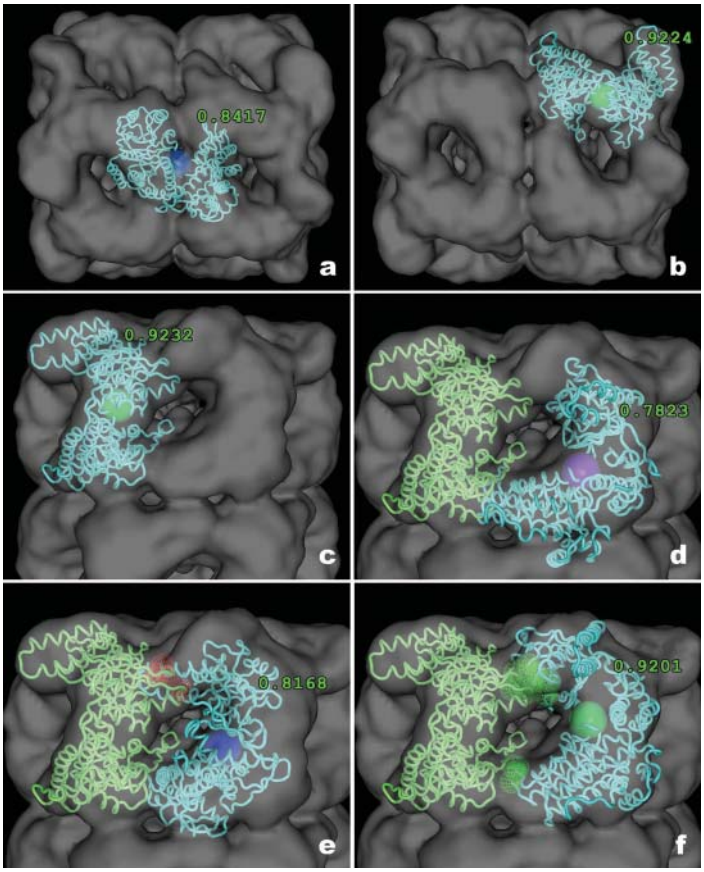


Figure 3: Stages of interactive global docking. The figure shows direct screen captures from the Sculptor visualization package. Please see text for details.

Let us now focus on a specific example and examine the steps performed by both the user and the software. As a test system, we chose the GroEL chaperonin (PDB entry 1GRL). This protein is a homo-14mer with a monomer weight of approximately 47 kDa and numerous experimental cryo-EM data sets are available. Here, we are using a 11.5 Å reconstruction [7]. First, an exhaustive search is performed in Eliquos and the scoring field is saved to a set of files. In the present case, the field was generated via standard cross-correlation, without the use of any filters. This scoring field is then loaded in Sculptor and the IGD is initialized. Figure 3a shows the initial display with the probe structure in the center of the experimental volume. The volume data set has been turned transparent

to make the probe structure visible. The user now activates IGD and starts moving the probe structure, using either the mouse or a haptic device (Figures 3b and 3c). The user is only responsible for translating the center of the monomer, the optimal orientation at the present position is determined by the scoring field. So, as the probe structure is moved around the experimental data set, it automatically rotates into the most favorable orientation. In addition, the (globally normalized) score at the present position is displayed graphically through color changes of the central sphere, as well as numerically. The user then locates a suitable candidate position for the first docked monomer, taking the global docking score and his or her knowledge of the system into account. This location is saved as a solution (Figure 3d, green structure) and the probe is moved in search for the next candidate location. At this stage, the additional steric information generated by Sculptor comes into play. The user can visualize steric clashes between the current probe structure and all previously docked solutions (Figure 3e) as well as good protein contacts (Figure 3f). Once all constituent proteins are approximately placed, their positions can be further adjusted by either manual or automatic refinement. For example, Figure 4 shows a closer view of the fit resulting from automatic Laplacian refinement of all subunits. No steric clashes are evident and the protein interfaces are highlighted by green spheres, signifying good contacts.

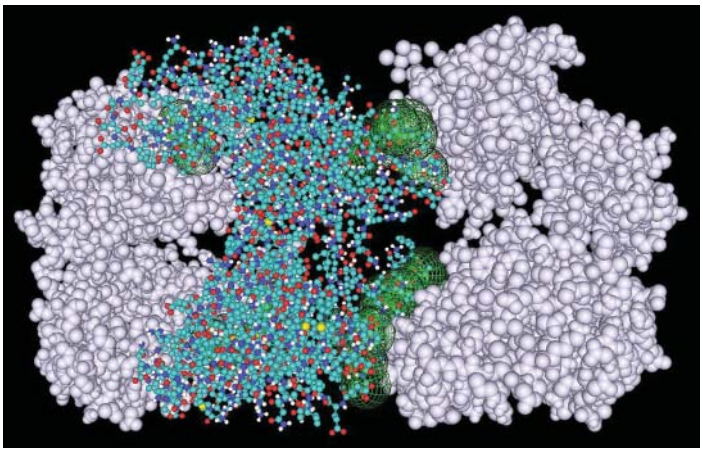


Figure 4: Results of automated refinement of the interactively docked structures. No steric clashes occur and the protein-protein interfaces are highlighted by green spheres.

4. Conclusions

The interactive global docking approach, presented here, combines the best features of non-interactive exhaustive search techniques and purely interactive visualization methods. It provides the user with visual feedback about global docking scores, steric clashes, and good protein-protein interfaces. In addition, haptic rendering can be employed to further enhance the interaction with the user. The additional information supplied during IGD allows biologists to not only rely on their personal knowledge of the system but also draw on objective, software-generated, fitting information. The currently available indicators represent only the first steps for incorporating more information into multi-resolution docking procedures. In the future, we plan on including contact information from mutation experiments, distances from NMR or FRET quenching measurements, as well as improving the existing scoring functions. ■

Availability

Both Eliquos and Sculptor are freely available on our website (http://www.biomachina.org).

Acknowledgments

This work was supported by NIH Grant R01-GM62968 as well as a training fellowship from the National Library of Medicine to the Keck Center for Interdisciplinary Bioscience Training of the Gulf Coast Consortia (NLM Grant No. 5T15LM-07093) to J. H.

References

[1] Frank, J. 2002. Single-particle imaging of macromolecules by cryo-electron microscopy. *Ann. Rev. Biophys. Biomol. Struct.* 31:303–319.

[2] Lučić, V., F. Förster, and W. Baumeister. 2005. Structural Studies by Electron Tomography: From Cells to Molecules. *Annu. Rev. Biochem.* 74:833–865.

[3] Petoukhov, M. V., and D. I. Svergun. 2007. Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Op. Struct. Biol.* 17:562–571.

[4] Heyd, J., M. F. Schmid, W. Wriggers, and S. Birmanns. 2008. Hessian filtering for docking of atomic structures into 3D cryo-electron microscopy reconstructions. *Biophys. J.* submitted.

[5] Birmanns, S., and W. Wriggers. 2007. Multi-resolution anchor-point registration of biomolecular assemblies and their components. *J. Struct. Biol.* 157:271–280.

[6] Birmanns, S., and W. Wriggers. 2003. Interactive fitting augmented by force-feedback and virtual reality. *J. Struct. Biol.* 144:123–131.

[7] Ludtke, S. J., J. Jakana, J. Song, D. T. Chuang, and W. Chiu. 2001. An 11.5 A Single Particle Reconstruction of GroEL Using EMAN. *J. Mol. Biol.* 314:241–250.