



## Using *Sculptor* and *Situs* for simultaneous assembly of atomic components into low-resolution shapes

Stefan Birmanns\*, Mirabela Rusu, Willy Wriggers<sup>1</sup>

University of Texas School of Biomedical Informatics at Houston, 7000 Fannin St. UCT 600, Houston, TX 77030, USA

### ARTICLE INFO

#### Article history:

Available online 13 November 2010

#### Keywords:

Multi-resolution modeling  
Multi-component registration  
Macromolecular assembly  
Cryo-electron microscopy  
Immersive visualization  
Haptic rendering

### ABSTRACT

We describe an integrated software system called *Sculptor* that combines visualization capabilities with molecular modeling algorithms for the analysis of multi-scale data sets. *Sculptor* features extensive special purpose visualization techniques that are based on modern GPU programming and are capable of representing complex molecular assemblies in real-time. The integration of graphics and modeling offers several advantages. The user interface not only eases the usually steep learning curve of pure algorithmic techniques, but it also permits instant analysis and post-processing of results, as well as the integration of results from external software. Here, we implemented an interactive peak-selection strategy that enables the user to explore a preliminary score landscape generated by the *colores* tool of *Situs*. The interactive placement of components, one at a time, is advantageous for low-resolution or ambiguously shaped maps, which are sometimes difficult to interpret by the fully automatic peak selection of *colores*. For the subsequent refinement of the preliminary models resulting from both interactive and automatic peak selection, we have implemented a novel simultaneous multi-body docking in *Sculptor* and *Situs* that softly enforces shape complementarities between components using the normalization of the cross-correlation coefficient. The proposed techniques are freely available in *Situs* version 2.6 and *Sculptor* version 2.0.

Published by Elsevier Inc.

### 1. Introduction

Modeling macromolecular complexes at atomic-level of detail is an important step toward understanding their functional properties. Although X-ray crystallography and NMR can provide such high-resolution structures, alternative biophysical techniques also provide a wealth of intermediate- to low-resolution data. The different experimental and environmental conditions of a cryo-electron microscopy (EM) single particle study (Frank, 2002), or of an electron tomography reconstruction (Medalia et al., 2002), often allow the collection of complementary information. In addition, these low-resolution techniques are often the only way to visualize large molecular complexes in their assembled state.

Over the last few years, a variety of algorithms has been developed for the automatic interpretation of volumetric data, which can be categorized by the scoring function that they aim to optimize. One set of techniques employs the cross-correlation coefficient (CC; Volkman and Hanein, 1999; Roseman, 2000; Chacón and Wriggers, 2002) as the scoring function. These methods can determine the position of a probe molecule inside a volumetric target map in a fully automatic way, by either determining the

placement of a single unit at a time (Rath et al., 2003; Garzón et al., 2007), or multiple models simultaneously (Kawabata, 2008; Lasker et al., 2009; Rusu and Birmanns, 2010).

The docking accuracy of these techniques depends on the shape of the fitted components and on the resolution of the map (Chacón and Wriggers, 2002). Spurious placements can be encountered for low-resolution cryo-EM reconstructions where interior density features such as secondary structure elements remain obscured. A Laplacian filter (Chacón and Wriggers, 2002) could improve the performance of contour-based matches, but the resolution remained a limiting factor. Other limitations that may complicate the docking include resolution anisotropy from preferred particle orientations, stain artifacts in negative stain EM, and conformational heterogeneity eroding features of the reconstruction. In addition, one might have to rely on distant homologues for the basis of the atomic models instead of crystal or NMR structures.

The aim of the present report is to extend the performance of quantitative multi-resolution docking methods for challenging systems where the fitting is ambiguous. We propose an integrated strategy that combines fully automatic algorithms with an interactive peak selection method. Instead of depending exclusively on automated optimization, our method enables visual exploration of a preliminary scoring landscape on a discrete lattice, one component at a time. The user selects local maxima of the docking score based on her expert knowledge of the spatial organization of the

\* Corresponding author. Fax: +1 713 500 3907.

E-mail address: [Stefan.Birmanns@uth.tmc.edu](mailto:Stefan.Birmanns@uth.tmc.edu) (S. Birmanns).

<sup>1</sup> Present address: D.E. Shaw Research, 120 W. 45th St., New York, NY 10036, USA.

atomic models. The preliminary results of the interactive (or automatic) peak search are refined further with a novel off-lattice refinement in which the placement of the component structures optimized simultaneously. Fitting a group of independent components at the same time introduces additional steric constraints, which can increase the docking accuracy, especially in the case of ambiguous shapes or low-resolution data. Fig. 1 shows the different stages of the workflow supported by the software system.

In addition to the above-mentioned algorithmic techniques, the modeling workflow typically requires an exploration and visualization of the data sets. Therefore, in Section 2, we will first present a brief overview of the *Sculptor* software and its main features. Subsequently, we will demonstrate the proposed methods using simulated and experimental maps for which atomic structures exist, such as the 14 Å resolution cryo-EM map of the actomyosin complex (Holmes et al., 2003) and a GroEL reconstruction at 11.5 Å resolution (Ludtke et al., 2001). In Section 3 we will describe the interactive peak selection and construction of preliminary start models for refinement. The results of the peak search are then optimized by the simultaneous multi-body refinement step, as reported in Section 4. We will also compare the results with traditional single-body refinement.

## 2. Interactive modeling with *Sculptor*

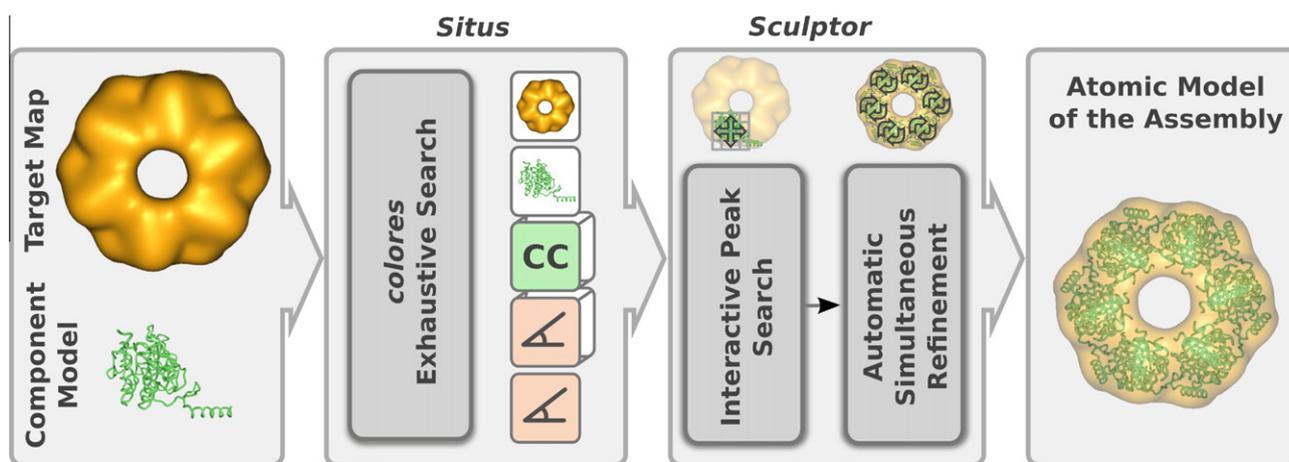
*Sculptor* is a graphical user interface (GUI)-based multi-scale modeling program, which combines various visualization techniques with pattern matching and feature extraction algorithms. The concept originated from the observation that several algorithms of the *Situs* modeling package (Wriggers, 2010) were time efficient enough to be executed in a GUI-based environment, where the outcome can be explored and parameters can be changed immediately. This concept was initially implemented in the form of a software *SenSitus* (Birmanns and Wriggers, 2003), which focused on interactive docking via haptic rendering in a virtual reality environment. Over time, the software transformed into a general purpose visualization and multi-scale modeling tool called *Sculptor*. In this new role, the program was used increasingly for pre- and post-processing of data sets and to analyze the outcome of non-interactive *Situs* algorithms. In addition, *Sculptor* served as a test bed for very efficient docking algorithms, which would benefit directly from the interactivity possible in the GUI environment.

*Sculptor* is organized around a graphical user interface (Fig. 2). The software aims to reduce visual “clutter” by using a single-window approach, where the main user interface elements are arranged around a central 3D graphics area, using non-overlapping frames. Internally, the program forms only a thin layer on top of a general-purpose C++ class library, which implements the algorithms and visualization techniques (Fig. 3). The modular approach shown in Fig. 3 enables the testing of algorithms isolated from the main application program and also permits us to encapsulate the operating system-dependent routines in a single area of the code. Over time, this strategy has also proven to be very effective for introducing new lab members to the project, and it helps to keep the source maintainable.

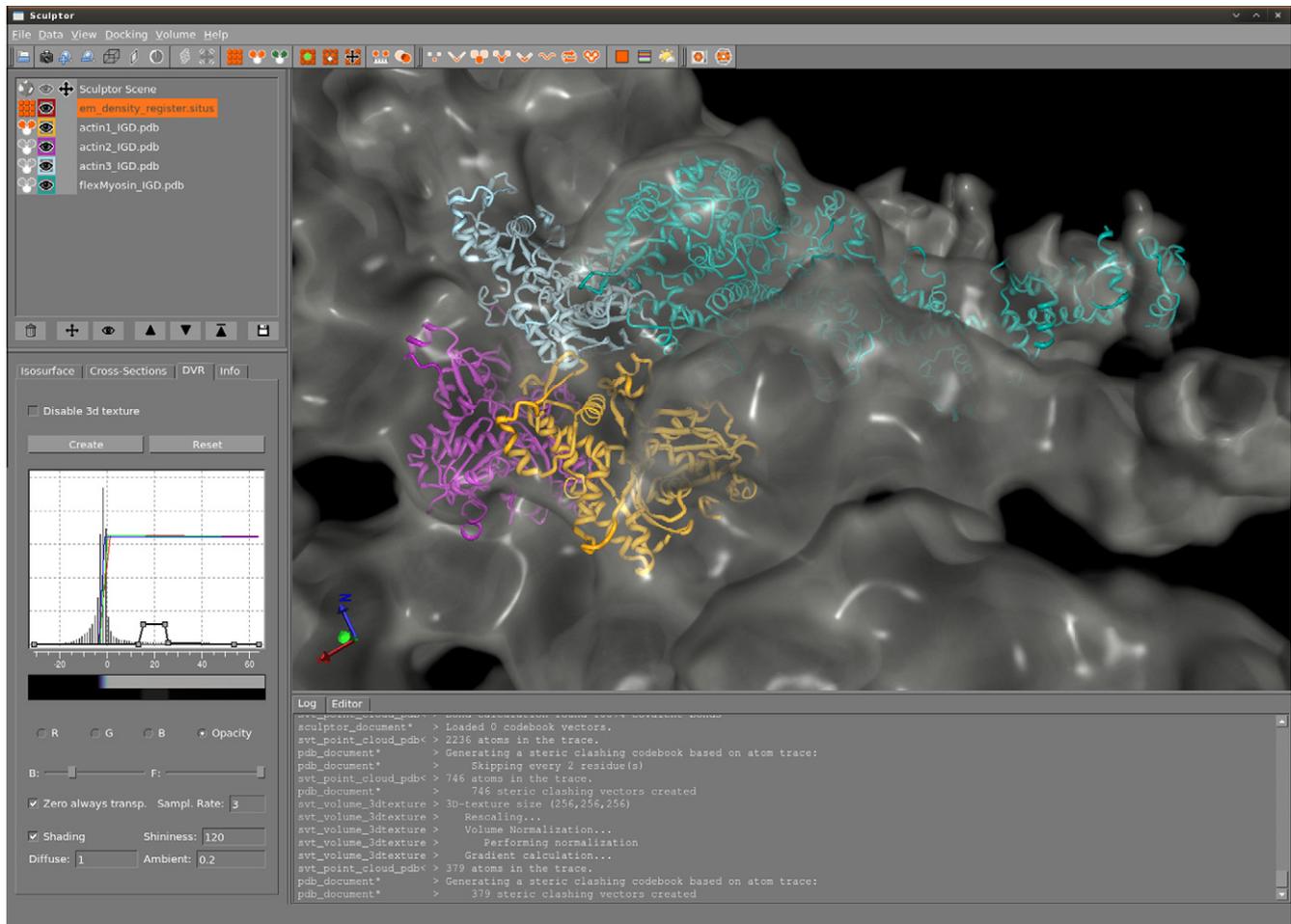
Visualization of atomic models is essential for the interpretation of the 3D shape of molecular systems (Humphrey et al., 1996; Delano, 2002; Goddard and Ferrin, 2007). The use of *Sculptor* for multi-resolution modeling results in unique challenges for the depiction of biomolecular assemblies. In multi-resolution modeling, one routinely combines several atomic models thereby constructing large molecular assemblies. The size of such large systems requires additional efforts to keep the visualization updated in real compute time. We have implemented several innovative rendering techniques that off-load the creation of the geometry to programmable graphics card processors (GPUs). Using GPU acceleration, *Sculptor* is able to display molecular systems with hundreds of thousands of atoms in various rendering styles.

Besides the visualization of atomic models, an efficient depiction of volumetric maps is important for exploring the shape of and the visual agreement between the multi-resolution data sets. Therefore, *Sculptor* offers a variety of volume rendering strategies. The maps can be viewed using the traditional iso-surface technique, which relies on the marching cube algorithm (Lorenson and Cline, 1987) to convert the volume into a triangular mesh. This well-established technique is very efficient and allows an interactive rendering of small to medium-sized data sets, but it cannot directly display different intensity values simultaneously.

As an alternative, *Sculptor* also supports a direct volume rendering technique, which enables the user to assign material properties freely to the intensities. This approach is based on the pre-integrated volume rendering idea (Engel et al., 2001), but was implemented in *Sculptor* using the features of modern programmable graphics cards. Direct volume rendering is very popular in other scientific disciplines, for example, medical visualization, and offers more



**Fig. 1.** Overview of the proposed “Interactive Global Docking” workflow. The multi-resolution data is first analyzed using an exhaustive search approach in *Situs* (Wriggers, 2010): the program *colores* exports the preliminary score landscape (translation function) as a 3D map of maximum CC values. The rotations corresponding to each voxel, determined by an exhaustive search of three Euler angles with user-selected rotational granularity, are exported as a 3D map of indices into a rotation table (green and red boxes). In *Sculptor*, the interactive peak search based on the preliminary scores yields a preliminary model of the entire complex, which is then further optimized using the new multi-body refinement procedure.



**Fig. 2.** Screenshot of the graphical user interface of *Sculptor*. The software is organized around the main 3D rendering panel, and minor panels are grouped into non-overlapping frames. The upper section of the left panel shows a list of the loaded data files along with their type, and their visibility and mobility status (the lower half of the panel changes its content based on the selected document). In the screenshot, the volumetric map is selected and the direct volume rendering tab is shown. The user can freely define the transfer function (mapping of material properties to intensities) based on the histogram of the map. The bottom right section shows the log panel, where the program communicates the status and outcome of commands and algorithms.

flexibility than surface-based representations. Every intensity level of the volumetric data is rendered using specific (and user-controllable) material settings, such as varying colors and transparency levels. This permits a simultaneous rendering of different intensities that are distinguishable by the user. Thus smaller, almost disconnected regions can be shown as part of the main body of the molecule (Fig. 4). The technique can also help to visualize map intensities that differ between parts of a map (see article by Wriggers et al., in this issue) without requiring a matching of densities.

Finally, *Sculptor* also offers a two-dimensional visualization of map cross-sections that can be added to the main scene, and the GUI-based “Map-Explorer” permits direct access to and querying of individual voxels. Although not relevant to the scope of this paper, the software is able to render deformable models in the context of flexible fitting of molecular models (Rusu et al., 2008). *Sculptor* reads and writes volumetric files in the *Situs* format and can import and export feature vectors for the interactive manipulation of a feature vector “skeleton” used in flexible fitting (Wriggers et al., 2004). The feature vectors can also be used for a coarse-grained modeling of structural data sets and for an interactive “point cloud matching” (Birmanns and Wriggers, 2007).

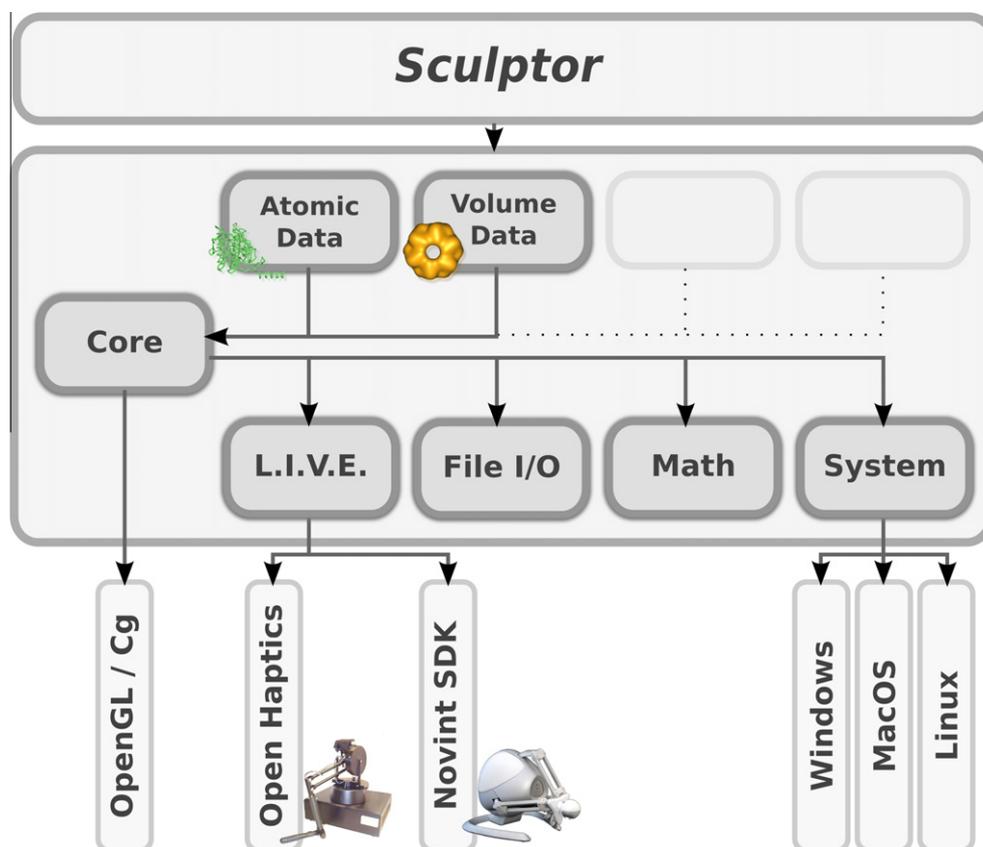
As described above, *Sculptor* was initially designed as a GUI-based extension of the *Situs* modeling package, and it is therefore connected in several ways with the *Situs* command-line tools. In the following section, we present a specifically useful link between

the two suites, where intermediate results of the *colores* program are explored in *Sculptor*, and the peaks of a scoring landscape are identified for further automatic refinement.

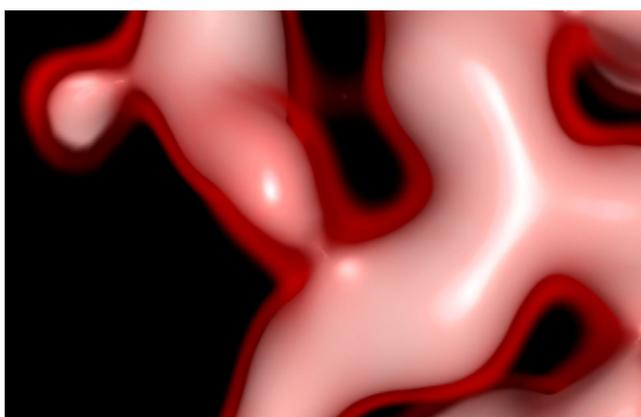
### 3. Interactive peak search, one component at a time

The CC is an established scoring function for multi-resolution modeling and has been used for a decade in various software tools (Volkman and Hanein, 1999; Roseman, 2000; Chacón and Wriggers, 2002). The maximization of the CC in rigid-body docking can be separated into three translational and three rotational degrees of freedom. The translational search can then be greatly accelerated by exploiting the Fourier convolution theorem. The main idea is to scan all rotations explicitly with a user-definable rotational granularity while rapidly computing the 3D translation function on the discrete lattice of the EM map using the Fast Fourier Transform (Wriggers and Chacón, 2001). As part of the overall docking, the *colores* program (Chacón and Wriggers, 2002) generates an intermediate 3D “map” of CC scores and corresponding angles by saving the maximum CC and corresponding angles for each voxel. This preliminary score “landscape” is sufficiently well sampled to enable the identification of peaks.

This component-wise peak-search process would be straightforward if it could be based on the score values alone. But in



**Fig. 3.** Software architecture of *Sculptor*. The main application program comprises only a thin software layer on top of a C++ class library, which is subdivided into modules. Low-level modules like “System”, which connects the library with the different operating systems, form the basis for higher-level modules. The “Library for Input devices in Virtual Environments” (LIVE) provides routines to dynamically load device drivers for special input devices like the SensAble 6DOF or Novint Falcon force feedback controllers. “Core” implements a general purpose 3D scenegraph architecture, which is extended by high-level modules that provide nodes to visualize specific data types. *Sculptor* uses the visualization nodes for atomic models and volumetric maps, although additional high-level modules exist for other applications.



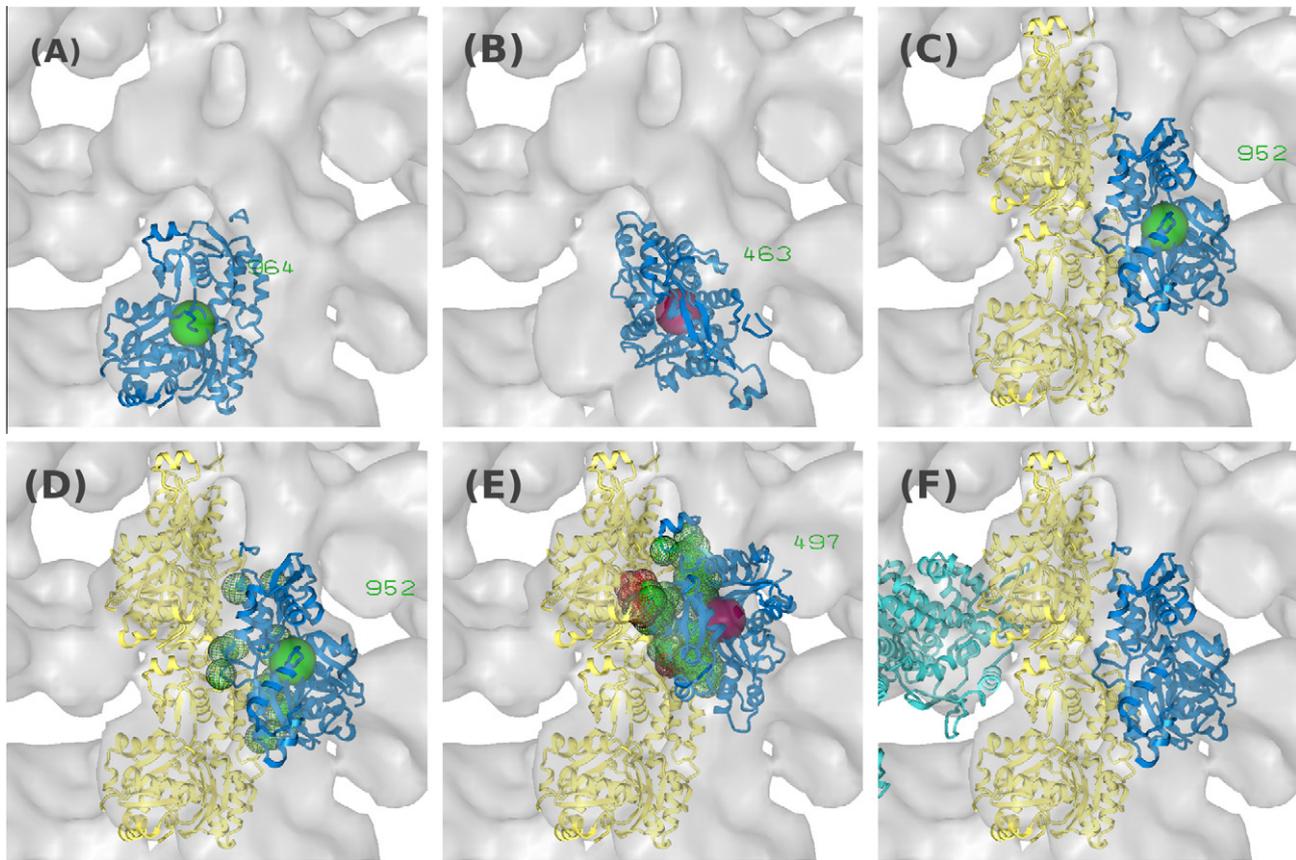
**Fig. 4.** Direct volume rendering permits the visualization of soft boundaries in volumetric maps. Shown here is an interactive visualization of the DegP protease-chaperone in *Sculptor* (EMDB 1504).

practice other criteria have to be optimized simultaneously, for example, the proximity of other docking solutions. (In the automated peak detection of *colores*, a relatively small user-defined number of candidate peaks are explored, for which both score values and a simple proximity criterion – based on map resolution – must be satisfied). The following sections describe an interactive peak search technique implemented in *Sculptor*, and provide examples of its application in practical situations.

### 3.1. Technical implementation of the interactive peak search

The purpose of the interactive peak search in *Sculptor* is to empower the user to freely explore an unrestricted number of possible peak positions in translational space (as an alternative to the automated peak filter used by *colores*). Components are fitted one at a time. The exploration of the peaks (by mapping rigid-body degrees of freedom onto fitted structures) is efficient for on-line use, unlike the exhaustive search that requires a slower, off-line calculation with *colores*. The interactive peak search comprises one step in the overall “Interactive Global Docking” workflow (Fig. 1) that uses exhaustive search data in an interactive fashion, combining it with steric information generated on-the-fly while the user examines the solution space.

As mentioned above, a cross-correlation based exhaustive search cannot be carried out in real-time because it takes several minutes to hours to complete, depending on the search granularity and map size. Therefore, the scoring data is generated in advance by the *colores* program of the software package *Situs*. To enable the interactive exploration of peaks a new option “-sculptor” was introduced in *colores* in *Situs* version 2.5 that writes intermediate scoring data to *Sculptor*-readable files. Specifically, the program exports (i) the target map in compatible indexing, (ii) a preliminary score “master file” as a 3D map of maximum CC values, and (iii) the rotations corresponding to each voxel as a 3D map of indices into a rotation table. *Sculptor* then reads the relevant files and allows the user to interactively explore fits, with the help of both visual and haptic (force) feedback from the software. Once a good candidate solution is identified for one of the subunits, steric interaction data



**Fig. 5.** Different stages of interactive peak search shown as direct screen captures from *Sculptor*: docking of subunits into the 14 Å cryo-EM map of the actomyosin complex (Holmes et al., 2003). (A) and (B) The user translates the G-actin monomer as desired inside the map, while the optimal rotation at each position is updated in real time. The color of the central sphere and the numeric value reflect the (normalized) CC value. (C)–(E) Two G-actin monomers are docked (yellow ribbons) and fixed, while a third G-actin monomer is translated inside the map. (E) and (F) Contacts are shown between the mobile G-actin monomer and the fixed monomers (favorable contacts are displayed as green spheres and steric clashes in red). (F) The myosin S1 fragment docked inside the map of the complex.

is generated and displayed on the fly, warning the user of potential clashes and highlighting good shape complementarity (Heyd and Birmanns, 2008). If a haptic device is available, the CC values and steric interactions can be interpreted as a potential function from which to compute forces that guide the user (Heyd and Birmanns, 2009).

However, the docking solutions generated by the peak search are only approximate. The preliminary solutions are biased by the particular order of the placement. They are also subject to the translational and rotational search granularity. They will be refined further via automated local optimization (Section 4).

### 3.2. Practical applications of the interactive peak search

We demonstrate the interactive peak search procedure using a specific experimental example, the actomyosin complex (Holmes et al., 2003). The 14 Å resolution map comprises (filamentous) F-actin decorated with myosin fragment S1. To generate a model of the filament, we used two atomic structures: a G-actin monomer (Holmes et al., 2003) and a flexed model of myosin S1, as described by Wriggers (2010). The docking was compared with two reference structures: the F-actin model (Holmes et al., 2003) alone or in complex with the flexed myosin S1. Our test was realistic in the sense that the subunits did not have perfectly fitting interfaces: G-actin and myosin S1 subunits used for the docking were all determined in isolation and not in the complex visualized by the cryo-EM map. Although the complex has never been solved at atomic resolution, the reference structures were deemed sufficiently reliable for this

test because the F-actin model was independently refined against X-ray fiber diffraction data (Holmes et al., 2003), and the myosin S1 was fitted to an individual S1 fragment after subtracting the F-actin (Wriggers, 2010).

Within *Sculptor*, the user can move the probe structure using the mouse or a supported haptic (force feedback) device. As the probe structure is moved around, it automatically rotates into the most favorable orientation (Fig. 5A and 5B show the G-actin monomer as the scoring landscape is explored). In addition, the (globally normalized) score at the present position is displayed graphically through color changes of the central sphere, as well as numerically. A suitable candidate position for the first monomer can thus be located, on the basis of both the global docking score and the user's knowledge of the system. If a haptic device is available, forces will guide the user to promising docking locations. The chosen location is saved as a solution (Fig. 5C–E shows two such solutions in yellow ribbon representations), and the probe is moved to search for the next candidate position. At this stage, the additional steric information generated by *Sculptor* is displayed. Both shape complementarity (Fig. 5D) and steric clashes (Fig. 5E) can be visualized.

The *colores* tool was employed for the preliminary exhaustive search fitting of the two atomic components to the 14 Å-resolution map. The rotational granularity was 9° and the translational granularity (voxel spacing) was 2.75 Å. The preliminary model generated by the interactive peak search shown in Fig. 5 exhibited a root mean square deviation (RMSD) of 5.1 Å from F-actin alone and a RMSD of 3.8 Å from the complete actomyosin complex

**Table 1**

Comparison of various docking approaches for the actomyosin complex and the chaperonin GroEL (subunits include 12 G-actin monomers/12 myosin S1 for the actomyosin complex, and 14 monomers for GroEL). Root mean square deviation (RMSD) from the references and cross-correlation coefficients (CC) are shown. The CC values of actomyosin correspond to the map shown in Fig. 6 and are systematically lower than those of the GroEL due to filament end effects. The interactive peak search model is equivalent to *colores* before Powell optimization (Chacón and Wriggers, 2002). In the single-body refinement each fragment is fitted independently (equivalent to *colores* after Powell optimization), while in the multi-body refinement all fragments were simultaneously optimized.

Reference model	Actomyosin complex		GroEL (emd-1080 PDB 1XCK)	
	F-actin/actomyosin RMSD (Å)	CC	RMSD (Å)	CC
Reference model		0.576/0.703		0.946
Interactive peak search	5.1/3.8	0.537/0.672	4.3	0.881
Single-body refinement	2.6/2.1	0.569/0.698	1.7	0.945
Multi-body refinement	1.8/1.4	0.583/0.709	1.3	0.950

(Table 1). This preliminary model (still subject to search granularity) is sufficiently close to the reference structure to allow the refinement procedures of Section 4 to converge.

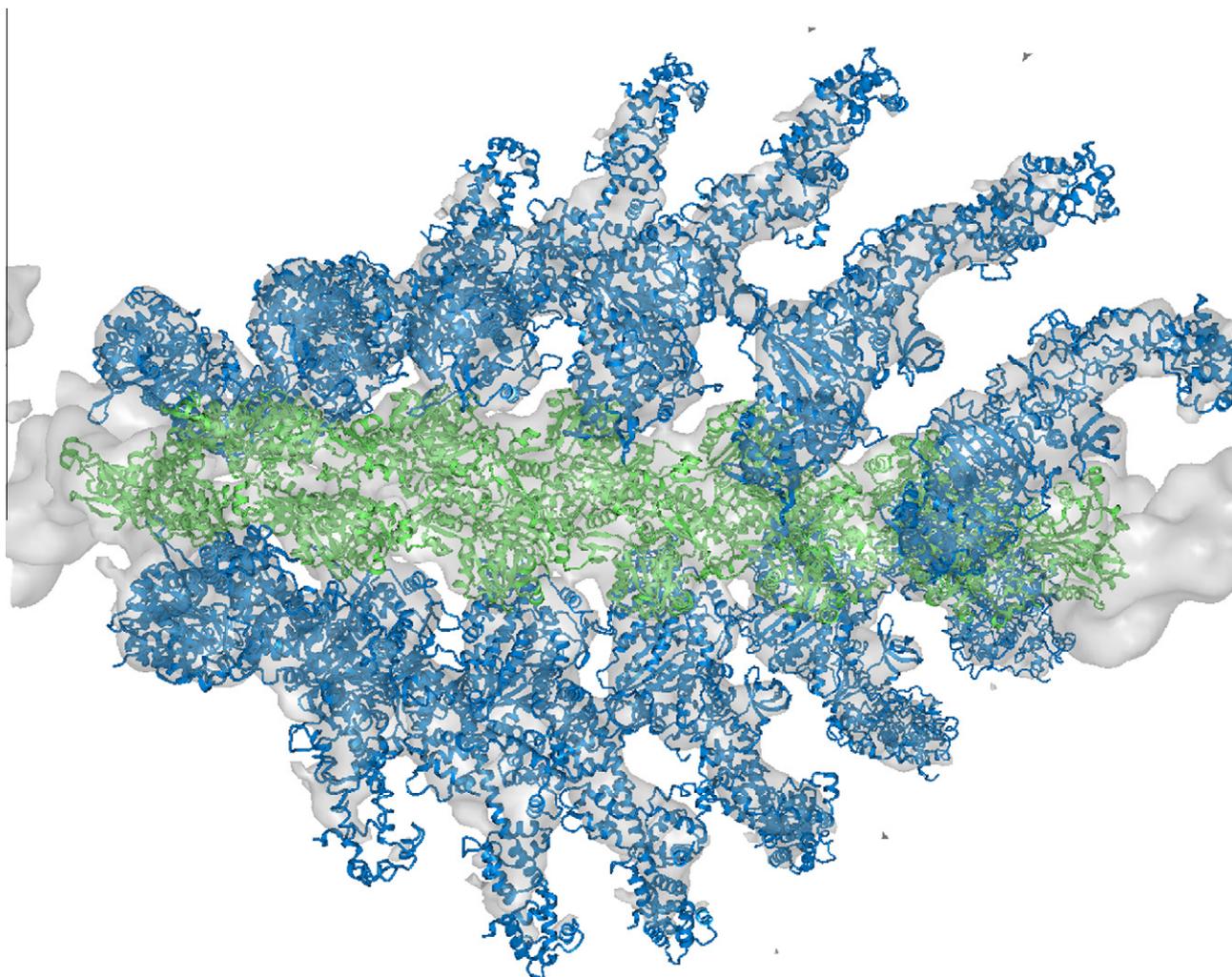
Similar results were obtained for a 11.5 Å-resolution reconstruction of GroEL (EMDB-1080, Ranson et al. (2001), Tagari et al. (2002)), where multiple copies of one monomer, extracted from

the crystal structure of the wild type GroEL (PDB 1XCK, Berman et al. (2002)), were docked into the map of the 14-mer. The preliminary model obtained after interactive peak search exhibited an RMSD of 4.3 Å from the reference (Table 1).

#### 4. Multi-body refinement

After exhaustive search and interactive or automatic peak search, an off-lattice refinement is required to optimize the position and the orientation of the preliminary scoring peaks. This is typically achieved with a local optimization that explores the off-lattice scoring landscape in the vicinity of an on-lattice peak. For example, the “Fit In Map” tool in Chimera (Pettersen et al., 2004) employs a steepest ascent local optimization. The programs previously developed by us, such as *colores* of *Situs* (Wriggers, 2010) and the “Refinement” tool in *Sculptor* (Birmanns and Wriggers, 2007), employ conjugate gradient ascent optimization (Powell, 1964). Similar to the gradient ascent, the Powell method uses the gradient information but employs a directional search to accelerate the convergence.

The traditional local refinement approaches have one factor in common: they optimize only the six rigid-body degrees of freedom of individual fragments, one at a time. However, a typical docking scenario often involves a larger number  $N > 1$  components that



**Fig. 6.** The actomyosin complex: Multi-body refinement of 12 G-actin monomers/12 myosin S1 inside the 14 Å-resolution map.

sterically interact with each other. Here, we introduce a novel multi-body refinement procedure that applies the Powell conjugate gradient optimization simultaneously to all components, to maximize the CC as a function of  $6N$  rigid-body degrees of freedom. The technique was implemented in *Sculptor* and in the *collage* tool of *Situs*. Starting from a good preliminary configuration, such as that provided by the above exhaustive search and interactive peak search, it is reasonable to expect that the algorithm quickly converges to the nearest maximum, despite the higher dimensionality of the search space compared to the traditional single-body refinement.

To demonstrate the performance of multi-body refinement, we compare it against the performance of single-body refinement for the actomyosin and GroEL test cases in Table 1. The traditional single-body optimization, in which each fragment is individually refined, improved the RMSD of 3.8 Å in the actomyosin model to 2.1 Å in 41 min.<sup>2</sup> Likewise, the GroEL model was improved from 4.3 Å to 1.7 Å RMSD in 6 min 43s<sup>2</sup>. Both actomyosin and GroEL models benefited from the new multi-body refinement as their RMSD was reduced further to 1.4 Å or less. The 24-units of the actomyosin complex were refined in 2 h 20 min while the 14 subunits of GroEL only required 16 min 17 s for the optimization<sup>2</sup>. The final actomyosin model is shown in Fig. 6. The CC values of the fits in Table 1 generally mirror the RMSD results. In other words, lower RMSD corresponds to higher CC scores, with the interesting exception of the reference structures themselves. (The small remaining RMSD after multi-body refinement thus appears to be caused by over fitting of the CC – we do not necessarily expect the reference structures to exhibit the highest CC because of their different biophysical origins from those of the cryo-EM maps).

Our validations on experimental maps show that the additional spatial constraints from multi-body refinement improve the fit of subunits in the map without requiring any special filters or masks. We expected that such a refinement would be particularly beneficial for low-resolution maps where all densities can be accounted for, such as in symmetric oligomers. To test this hypothesis, we created simulated low-resolution maps (Belnap et al., 1999) from various known oligomeric structures (PDBs: 1NIC – trimer (Adman et al., 1995), 1QQW – pentamer (Ko et al., 2000), 1XMV – hexamer (Xing and Bell, 2004), 1TYQ – heptamer (Nolen et al., 2004)). The deviation of monomers from the oligomer crystal structure was computed as a function of resolution of the simulated maps (data not shown). For single-body refinement with standard CC, the models maintained the correct architecture down to 9–15 Å resolution, below which the monomers drifted toward the center of the map where the higher densities generate a high correlation. This breakdown of the traditional CC criterion due to blurring of interior map detail is well known (Chacón and Wriggers, 2002) and has prompted the development of filters and masks for single-body refinement in the past (Wriggers and Chacón, 2001). In contrast, in multi-body refinement, the models maintained the correct architecture (sub-Ångstrom accuracy placement in the simulated maps), even at resolutions as low as 50 Å, because the fragments had nowhere to go. Although this test used idealized data (perfectly simulated EM maps, and perfectly complementary subunits), the results do validate the multi-body refinement approach further.

## 5. Discussion and conclusions

We have described a novel, combined software system for the interactive exploration, interpretation, and analysis of multi-resolution data. Building on the classic *Situs* program package, *Sculptor* combines intermediate results of the exhaustive, quantitative

docking tool *colores* with a direct visual interpretation of the results. The interactive peak search and the new simultaneous multi-body optimization of the results guarantee a high level of flexibility and enable expert participation in the modeling, while still ensuring that the output is based on reproducible quantitative scoring functions.

The main difference between the proposed workflow (Fig. 1) and traditional docking with *colores* is the replacement of the automated peak search with an interactive one and the added possibility of performing multi-body refinement.

The experimental examples, actomyosin and GroEL, demonstrate that it is straightforward to select local optima of the CC score with the interactive peak search method in *Sculptor*. The interactive approach is beneficial in situations where information about the approximate position of a subunit is available from biophysics, i.e., where the correct candidate solution is obvious to an expert user. We expect that the interactive approach will work well in challenging situations where many degenerate solutions exist (such as in symmetric structures or multi-body docking at low resolution), or when correct fits are obscured by spurious maxima of the scoring function.

We proposed a novel multi-body refinement technique that aims to prevent steric clashes and guards against individual fragments wandering off into high densities that are already occupied by a different fragment. The spatial constraints, implicitly introduced by the normalization of the CC during simultaneous refinement of all components, help to identify better quality models. Our validations on both simulated and experimental test cases showed that multi-body refinement is more beneficial than single-body optimization. Based on the tests with simulated maps, we expect this benefit to become prominent especially at very low resolution. The test results suggest that a standard CC, without filters and masks, can be used as a scoring function for multi-body refinement at low-resolution if densities are well accounted for by the fitted atomic structures.

Although the runtimes are significantly higher in the case of the multi-body refinement, they are still in an acceptable range from minutes up to a few hours for large systems with a high number of subunits. Complementing the standalone *collage* tool of *Situs*, the progress can be monitored visually in *Sculptor* if desired. The inspection afforded by *Sculptor* allows the user to stop the process anytime, to fine tune parameters or to test if a sufficient level of convergence has been reached.

While the test results have been very promising, they do not rule out possible limitations of the proposed methods. First, it is possible that the  $6N$ -dimensional search exhibits a smaller zone of convergence than the  $N$  separate 6D searches (e.g. due to the higher dimensionality of the search space and dependencies between fragments). The size of the zone of convergence is not critical for off-lattice peak refinement (which starts close to the maximum) but it might pose a problem for sub-optimal models that are farther from the peak maxima. Second, some of the presented tests may have been overly idealized. In the GroEL and simulated map tests the subunits (extracted from the crystal structure of the complex) exhibit perfect shape complementarity. It is likely that in realistic docking problems, the atomic structure of the complex is unknown and that fragments assemble by induced fit. Flexible loops or secondary structure elements at the fragment interface might lower the score of the correct solution due to inadvertent steric clashes, requiring a flexible protein–protein docking (Noy and Goldblum, 2010). These scenarios and possible limitations depend heavily on the experimental system and will be explored in future work.

A possible extension of the current approach, which we plan to investigate in the future, is the integration of symmetry in the multi-body refinement procedure. It is expected that refining

<sup>2</sup> Runtime measured on a Intel 2 Duo processor E8400 @3.0 GHz.

symmetrical copies simultaneously is beneficial as it reduces the degrees of freedom of the search space.

We look forward to receiving feedback on the software performance from the user community. *Sculptor* is freely available at <http://sculptor.biomachina.org> for Windows, Macintosh and Linux operating systems. We have prepared convenient distribution packages for the different operating systems and have uploaded various tutorials and manual texts online. The features described in this manuscript are included in *Sculptor* versions 2.0 and later. The programs *colores* and *collage* are part of the *Situs* program package, available at <http://situs.biomachina.org>. The export of the *Sculptor*-compatible peak search files is implemented in *Situs* version 2.5 and later, and the multi-body refinement in *collage* is implemented in *Situs* version 2.6 and later.

## Acknowledgments

We thank Jochen Heyd, Manuel Wahle, Maik Boltes, and Herwig Zilken for contributions to earlier versions of *Sculptor*. We also thank Teresa Ruiz, Michael Radermacher, Tanvir Shaikh, Alexander Freiberg, Gkhan Tolun, and Tracy Nixon for their extensive testing of the software and their suggestions for fine-tuning the graphical interface. This work was supported in part by grants from National Institutes of Health (R01GM62968), the Human Frontier Science Program (RGP0026/2003), the Gillson-Longenbaugh Foundation, and by startup funds from the University of Texas Health Science Center at Houston.

## References

- Adman, E.T., Godden, J.W., Turley, S., 1995. The structure of copper-nitrite reductase from *Achromobacter cycloclastes* at five pH values, with NO<sub>2</sub>-bound and with type II copper depleted. *J. Biol. Chem.* 270, 27458–27474.
- Belnap, D.M., Kumar, A., Folk, J.T., Smith, T.J., Baker, T.S., 1999. Low-resolution density maps from atomic models: How stepping “back” can be a step “forward”. *J. Struct. Biol.* 125, 166–175.
- Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C., 2002. The protein data bank. *Acta Cryst. D* 58, 899–907.
- Birmanns, S., Wriggers, W., 2003. Interactive fitting augmented by force-feedback and virtual reality. *J. Struct. Biol.* 144, 123–131.
- Birmanns, S., Wriggers, W., 2007. Multi-resolution anchor-point registration of biomolecular assemblies and their components. *J. Struct. Biol.* 157, 271–280.
- Chacón, P., Wriggers, W., 2002. Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* 317, 375–384.
- Delano, W.L., 2002. The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA, USA. <http://www.pymol.org>.
- Engel, K., Kraus, M., Ertl, T., 2001. High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. In: *HWWS '01: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware*, pp. 9–16.
- Frank, J., 2002. Single-particle imaging of macromolecules by cryo-electron microscopy. *Ann. Rev. Biophys. Biomol. Struct.* 31, 303–319.
- Garzón, J.I., Kovacs, J., Abagyan, R., Chacón, P., 2007. ADP\_EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics* 23 (4), 427–433.
- Goddard, T.D., Ferrin, T.E., 2007. Visualization software for molecular assemblies. *Curr. Opin. Struct. Biol.* 17, 587–595.
- Heyd, J., Birmanns, S., 2008. Beyond the black box: interactive global docking of protein complexes. *Microsc. Today* 16 (4).
- Heyd, J., Birmanns, S., 2009. Immersive structural biology: a new approach to hybrid modeling of macromolecular assemblies. *Virtual Reality* 13, 245–255.
- Holmes, K.C., Angert, I., Kull, F.J., Jahn, W., Schröder, R.R., 2003. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* 425, 423–427.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. *J. Mol. Graphics* 14, 33–38.
- Kawabata, T., 2008. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys. J.* 95 (10), 4643–4658.
- Ko, T.P., Safo, M.K., Musayev, F.N., Di Salvo, M.L., Wang, C., Wu, S.H., Abraham, D.J., 2000. Structure of human erythrocyte catalase. *Acta Cryst. D* 56 (Pt 2), 241–245.
- Lasker, K., Dror, O., Shatsky, M., Nussinov, R., Wolfson, H.J., 2009. EMatch: discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-EM maps. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 4 (1), 28–39.
- Lorensen, W.E., Cline, H.E., 1987. Marching cubes: a high resolution 3D surface construction algorithm. In: *SIGGRAPH '87: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, vol. 21, pp. 163–169.
- Ludtke, S.J., Jakana, J., Song, J.L., Chuang, D.T., Chiu, W., 2001. A 11.5 Å single particle reconstruction of GroEL using EMAN. *J. Mol. Biol.* 314, 253–262.
- Medalia, O., Weber, I., Frangakis, A.S., Nicastro, D., Gerisch, G., Baumeister, W., 2002. Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* 298, 1209–1213.
- Nolen, B.J., Littlefield, R.S., Pollard, T.D., 2004. Crystal structures of actin-related protein 2/3 complex with bound ATP or ADP. *Proc. Natl. Acad. Sci. USA* 101, 15627–15632.
- Noy, E., Goldblum, A., 2010. Flexible protein–protein docking based on Best-First search algorithm. *J. Comp. Chem.* 31, 1929–1943.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comp. Chem.* 25 (13), 1605–1612.
- Powell, M.J.D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* 7 (2), 155–162.
- Ranson, N.A., Farr, G.W., Roseman, A.M., Gowen, B., Fenton, W.A., Horwich, A.L., Saibil, H.R., 2001. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 107, 869–879.
- Rath, B.K., Hegerl, R., Leith, A., Shaikh, T.R., Wagenknecht, T., Frank, J., 2003. Fast 3D motif search of EM density maps using a locally normalized cross-correlation function. *J. Struct. Biol.* 144, 95–103.
- Roseman, A.M., 2000. Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Cryst. D* 56, 1332–1340.
- Rusu, M., Birmanns, S., 2010. Evolutionary tabu search strategies for the simultaneous registration of multiple atomic structures in cryo-EM reconstructions. *J. Struct. Biol.* 170, 164–171.
- Rusu, M., Birmanns, S., Wriggers, W., 2008. Biomolecular pleiomorphism probed by spatial interpolation of coarse models. *Bioinformatics* 24, 2460–2466.
- Tagari, M., Newman, R., Chagoyen, M., Carazo, J.M., Henrick, K., 2002. New electron microscopy database and deposition system. *Trends Biochem. Sci.* 27, 589.
- Volkman, N., Hanein, D., 1999. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J. Struct. Biol.* 125, 176–184.
- Wriggers, W., 2010. Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* 2, 21–27.
- Wriggers, W., Alamo, L., Padrón, R., 2011. Matching structural densities from different biophysical origins with gain and bias. *J. Struct. Biol.* 173, 445–450.
- Wriggers, W., Chacón, P., 2001. Modeling tricks and fitting techniques for multiresolution structures. *Structure* 9, 779–788.
- Wriggers, W., Chacón, P., Kovacs, J.A., Tama, F., Birmanns, S., 2004. Topology representing neural networks reconcile biomolecular shape, structure, and dynamics. *Neurocomputing* 56, 365–379.
- Xing, X., Bell, C.E., 2004. Crystal structures of *Escherichia coli* RecA in complex with MgADP and MnAMP-PNP. *Biochemistry* 43, 16142–16152.