Journal of Structural Biology **133,** 193–202 (2001)
doi:10.1006/jsbi.2000.4350, available online at http://www.idealibrary.com on IDEAL®

Using *Situs* for Flexible and Rigid-Body Fitting of Multiresolution Single-Molecule Data

Willy Wriggers*,[1] and Stefan Birmanns†

**Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037; and †Central Institute for Applied Mathematics, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*

Received December 18, 2000, and in revised form March 20, 2001; published online June 8, 2001

We describe here a set of multiresolution visualization and docking procedures that we refer to as the *Situs* package. The package was developed to provide an efficient and robust method for the fitting of atomic structures into low-resolution data. The current release was optimized specifically for the visualization and docking of single molecules. A novel 3D graphics viewer, *volslice3d*, was developed for the package to provide an immersive virtual reality environment for measuring and rendering volumetric data sets. The precision of single-molecule, rigid-body docking was tested on simulated (noise-free) low-resolution density maps. For spatial resolutions near 20 Å typically arising in electron microscopy image reconstructions, a docking precision on the order of 1 Å can be achieved. The shape-matching score captured the correct solutions in all 10 trial cases and was sufficiently stringent to yield unique matches in 8 systems. Novel routines were developed for the flexible docking of atomic structures whose shape deviates from the corresponding low-resolution shape. Test calculations on isoforms of actin and lactoferrin demonstrate that the flexible docking faithfully reproduces conformational differences with a precision <2 Å if atomic structures are locally conserved.

Key Words: **docking; volumetric registration; topology representing neural networks; visualization; macromolecular assemblies; induced fit.**

INTRODUCTION

Situs is a program package for the quantitative registration of atomic biomolecular structures with corresponding low-resolution data from electron microscopy (EM) or solution X-ray scattering (Wriggers *et al.*, 1999; Wriggers and Chacón, in press). In EM applications, the package makes use of single-molecule density maps that can be obtained by single-particle image processing or by difference mapping of 3D data exhibiting variable subunit composition. The central algorithm is a topology-representing neural network (Wriggers *et al.*, 1998) that encodes resolution-invariant features within the 3D data sets. The encoding of the data by a discrete set of vectors (termed *vector quantization*) and the precision of the subsequent docking enable the interactive fitting of structures within seconds of compute time on a typical UNIX workstation. Though released only recently, the package has already received considerable attention in the EM community. At the time of this writing 100 users have registered the download, and a growing number of unaffiliated research groups use the software for their published work (Moores *et al.*, 2000; Kikkawa *et al.*, 2000; Llorca *et al.*, 2000).

Four new releases of *Situs* were issued since the creation of version 1.0. Substantial revisions were prompted in part by suggestions and requests from the user community. This article provides an overview of the changes and additions to the package design relative to the earlier article (Wriggers *et al.*, 1999). Improvements were made mainly in the areas of file format conversion, automation of docking, estimation of docking precision, and simulation of low-resolution maps. As before, the software consists of individual, stand-alone programs for format conversion, analysis, visualization, manipulation, vector quantization, and docking of 3D data sets. The freely available ANSI C source code is now portable to a large variety of architectures that run the free GNU "gcc" compiler, including Silicon Graphics, Sun, Hewlett-Packard, and PC/Linux workstations.

One of our development goals is to provide state-of-the-art 3D graphics capabilities with *Situs* that enable users to visualize both volumetric data sets and fitted atomic structures at a quality level suit-

[1] To whom correspondence should be addressed. Fax: (858) 784-8688. E-mail: wriggers@scripps.edu.

1047-8477/01 $35.00

able for journal publications (Wriggers *et al.,* 1999; Moores *et al.,* 2000; Llorca *et al.,* 2000). In the near term, to minimize the programming overhead spent on the visualization, we will continue to take advantage of the existing free molecular graphics program VMD (Humphrey *et al.,* 1996). *Situs* provides routines that interface with VMD to render isocontours of volumetric data in wire-mesh or solid representation. In addition to these existing routines we present here a novel tool, *volslice3d,* for the analysis and rendering of volumetric data. Like VMD, *volslice3d* is stereo-capable and supports tracking devices used in immersive 3D virtual reality (VR) environments.

Since multiresolution structural data are docked indirectly (by means of the vector quantization), several *Situs* users have expressed concerns about the fitting accuracy that can be achieved with the package. To address these concerns we developed a novel tool, *pdblur,* for creating simulated EM maps based on real-space kernel convolution. The lowering of the resolution is useful for validating our fitting algorithms on known atomic structures. There are many additional applications of resolution-lowering in EM (Belnap *et al.,* 1999) that are outside the scope of this article. Here, we use *pdblur* to evaluate the rigid-body docking performance as a function of resolution and other parameters, using 10 trial proteins as test systems. The database of trial systems was also used (Wriggers and Chacón, in press) for the validation of *Situs* routines that had been developed specifically for the fitting of atomic structures to bead-models derived from small-angle X-ray scattering (SAXS) data (Chacón *et al.,* 2000). The choice of identical test systems allowed us here to evaluate the effect of the density distribution (segmented bead model vs smooth volumetric map) on the docking performance. Also, the database contains three structures, catalase (tetramer), nitrito-reductase (trimer), and superoxide dismutase (dimer), that exhibit internal symmetry. The oligomeric symmetry is expected to yield degenerate best fits if the oligomers are docked as single molecules.

Applications of *Situs* are no longer limited to rigid-body docking of single molecules alone. The most intriguing problems in EM may arise in situations where the conformation of the reconstructed 3D density clearly deviates from a known atomic structure. Such conformational deviations may be due to functional motions in the biomolecular system or due to induced-fit when the system becomes part of a larger aggregate. We have recently proposed a novel method for the flexible docking of atomic structures to low-resolution shapes (Wriggers *et al.,* 2000). Resolution-invariant features were used as constraints in molecular dynamics (MD) simulations to bring a deviating atomic structure into register with a single-molecule EM map. Here, we present novel *Situs* routines that can be used to drive such flexible fitting simulations. For the first time, the precision of multiresolution flexible docking was tested on simulated low-resolution data. The results should be of interest in all areas of structural biology where large-scale domain motions of biomolecules are routinely modeled.

Although we are well connected to many groups in the EM community, we validate our algorithms, in this theoretical paper, solely on simulated EM maps that were derived from existing atomic structures. It is difficult to validate the docking precision using experimental maps, even if corresponding atomic structures are known, due to the limited precision of the alignment and due to unknown conformational differences between the crystal isoform and the imaged specimen. The use of simulated maps reveals the effect of resolution-lowering, but not the effect of noise in a 3D reconstruction, on the docking accuracy. Readers interested in applications of *Situs* to experimental data should look out for forthcoming publications with the groups of Edward Egelman, Elizabeth Wilson-Kubalek, and Seth Darst.

DESIGN OF *Situs* 1.4

The series of procedural steps and the C programs that facilitate and enable the docking of a single-molecule atomic structure into the corresponding low-resolution EM data are shown schematically in Fig. 1. The EM map format conversion tool *convert,* the *Situs*-supporting library files, and the EM data analysis and handling routines are described in the original article. The original handling tools include, e.g., *interpolate* (altering the lattice spacing), *invert* and *subtract* (sign inversion and subtraction of maps), as well as *histovox* (voxel histogram and density rescaling). In the following, we focus only on the novel functionalities and programs. Updated tutorials, available at the *Situs* Web site, demonstrate the application of the current release of the programs in a variety of realistic situations.

To be independent of changes in the map formats developed by laboratories in the EM community, and to facilitate the docking of atomic structures, we continue to maintain the ASCII-based *Situs* map format. The *convert* utility now enables conversion from and to standard formats such as X-PLOR, MRC, CCP4, SPIDER, and ASCII. The *Situs* format enforces a cubic lattice and tracks the coordinate system origin used for the docking. Skewed (non-orthogonal) unit cells in X-PLOR, MRC, and CCP4 map formats are now automatically converted to cubic *Situs* lattices using trilinear interpolation.

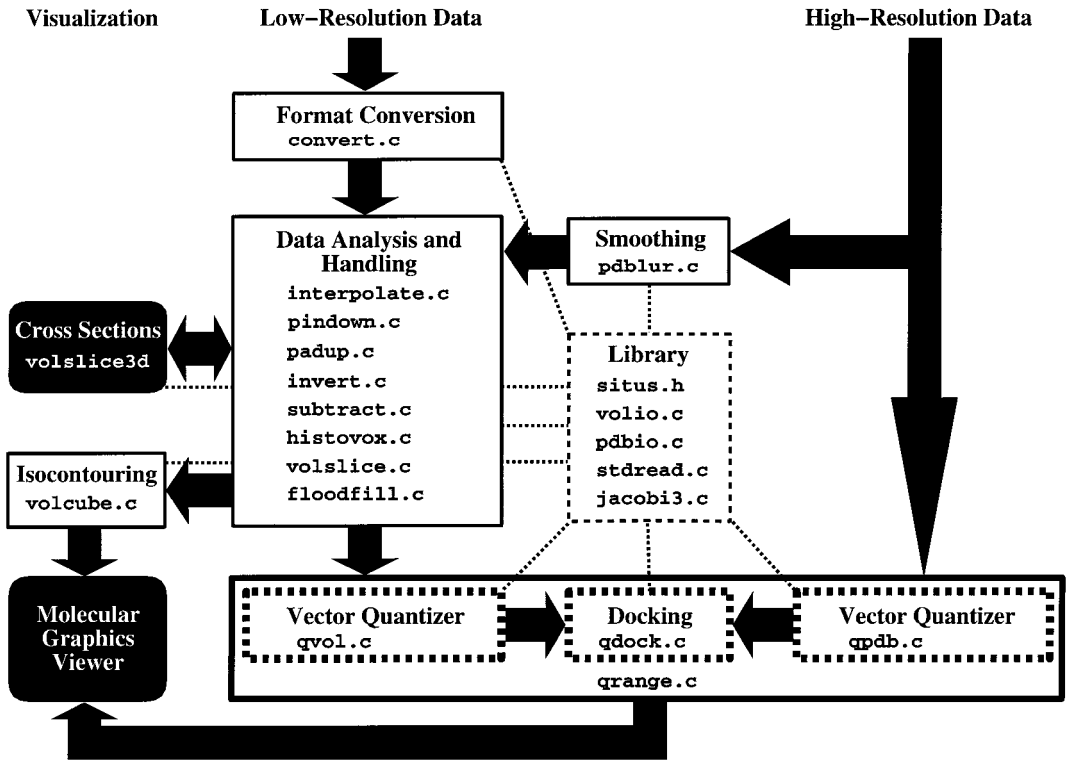
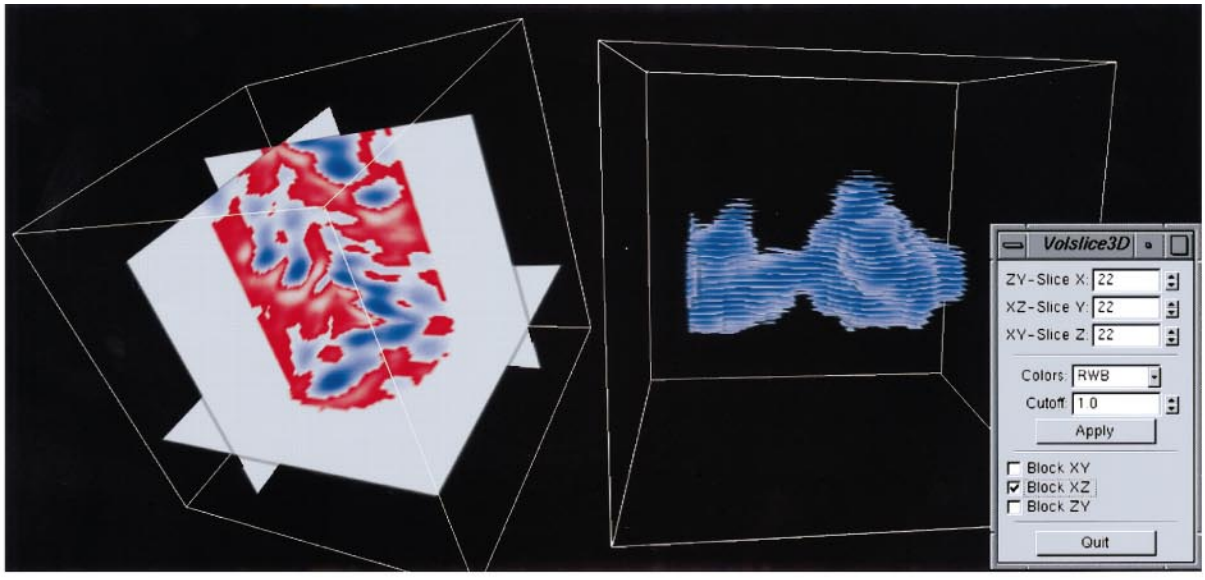


FIG. 1. Schematic diagram of the single-molecule EM routines of the *Situs* package (version 1.4). Individual terminal-based C program components (on white background) are classified by their functionality. Graphics windows are shown on a black background. The arrows illustrate the flow of data. The procedures are discussed in the text; additional documentation is available at http://situs.scripps.edu and http://www.fz-juelich.de/vislab/virtual/volslice3d.

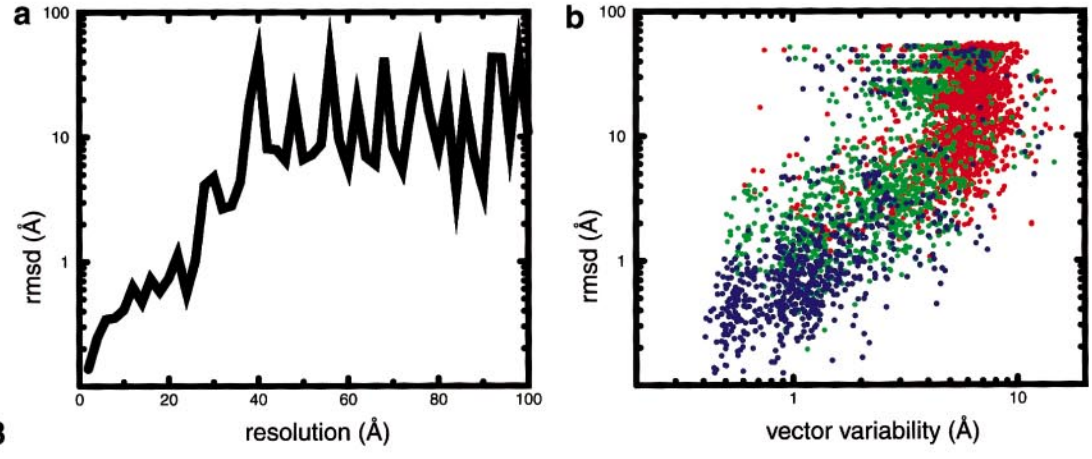
New map manipulation programs distributed in the current release include *padup* and *pdblur*. The *padup* editing utility adds additional x, y, or z margins to existing maps by zero-padding; its functionality is, thus, complementary to the *pindown* (margin-removing) program described earlier. The *pdblur* program creates simulated low-resolution maps from atomic structures by real-space kernel convolution. In *pdblur*, atomic structures are first projected onto a cubic lattice using trilinear interpolation and subsequently convoluted with a point-spread function (kernel). The available kernels include: Epanechnikov: $\max\{0, A(1-\frac{1}{2}(r/R)^2)\}$; "semi-Epanechnikov": $\max\{0, A(1-\frac{1}{2}(r/R)^{1.5})\}$; triangular: $\max\{0, A(1-\frac{1}{2}(r/R))\}$; and Gaussian: $A\exp(-\frac{3}{2}(r/\sigma)^2)$. Input parameters include the kernel amplitude A and the half-max kernel radius R, or the desired spatial resolution (i.e., 2σ, where σ is the R-dependent standard deviation). The user has the choice of correcting the kernel resolution for the point-spread introduced by the lattice projection.

To enable the docking of multiresolution data using a reduced data representation, two vector quantization routines were originally distributed with *Situs: qpdb*, for the quantization of single-molecule atomic resolution data, and *qvol*, for the quantization of single-molecule low-resolution volumetric maps. Vector quantization places a number of so-called *codebook vectors* at characteristic features of 3D density distributions. The vectors form a set of point landmarks that are robust under changes in resolution, identify gross features, and thereby provide information about the shape of a biological object (Wriggers *et al.*, 1998). *Situs* takes advantage of such a reduced representation of 3D data sets. By aligning only the gross features, the method brings multiresolution structures into register within seconds of computation on a typical workstation.

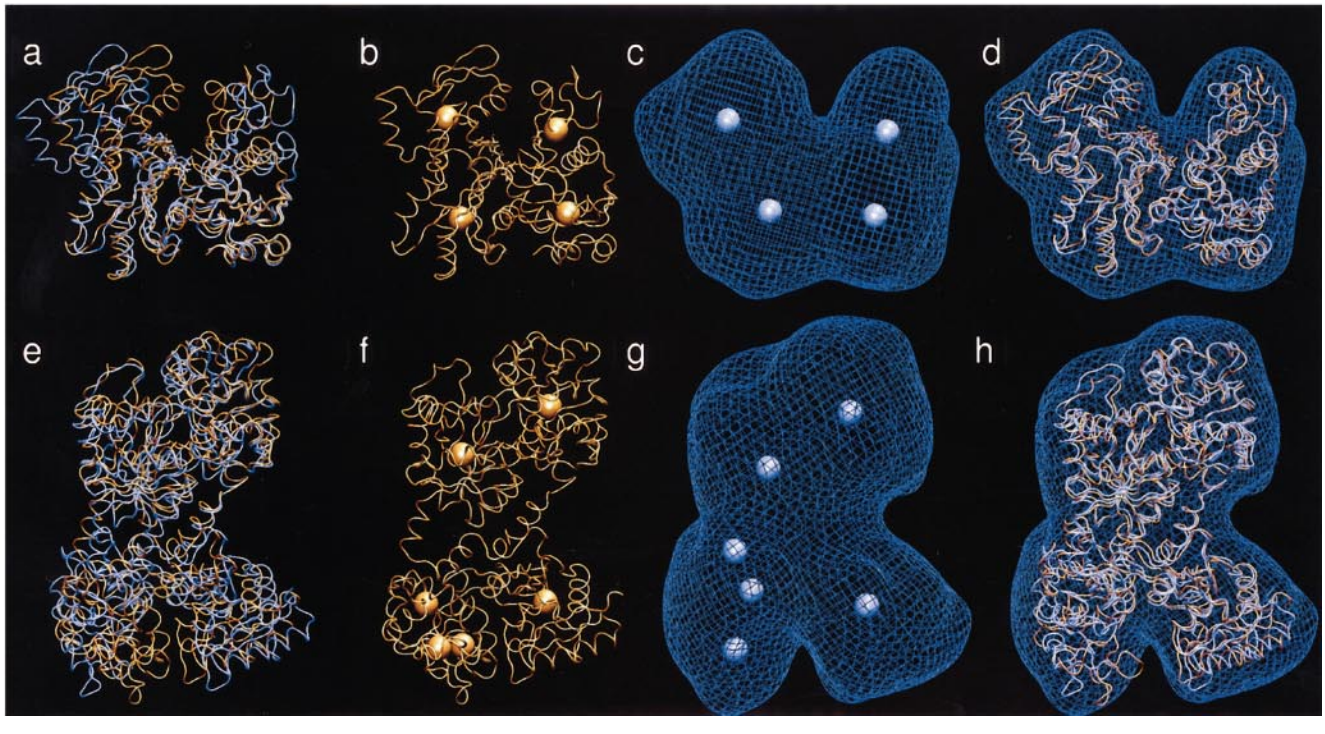
The vector quantization with *qpdb* or *qvol* represents a single-molecule data set by k codebook vectors: $\mathbf{x}_i$ (corresponding to high-resolution data), or $\mathbf{y}_j$ (corresponding to low-resolution data; $i, j = 1, \cdots, k$). An index map $I{:}j \rightarrow i$ defines the k pairs of corresponding vectors. In practical situations, I is not known a priori. The program *qdock* carries out an exhaustive search of the $k!$ possible permutations $(I(1), \cdots, I(k))$ and returns a list of best fits, ranked by the residual rms deviation (rmsd) after least-squares fitting of the vectors $\mathbf{x}_{I(j)}$ to the $\mathbf{y}_j$. Originally, k was restricted to values ≥ 4. In the


Volslice3D
ZY-Slice X: 22
XZ-Slice Y: 22
XY-Slice Z: 22
Colors: RWB
Cutoff: 1.0
Apply
Block XY
Block XZ
Block ZY
Quit

2


a
b
rmsd (Å)
resolution (Å)
vector variability (Å)
100
10
1
0
20
40
60
80

3


a
b
c
d
e
f
g
h

4

current release, *qdock* is able to handle also the special planar case $k = 3$. The ranking of the resulting (k-dependent) fits typically produces a clear prediction of the optimal docking configuration, as will be demonstrated under "Precision of Rigid-Body Docking."

The original alignment strategy using *qvol, qpdb,* and *qdock* required considerable effort. Since the optimum number k of vectors is not known a priori, a user had to explore docking results for various numbers k by exchanging data between the three programs in the UNIX shell. The new program *qrange* consolidates the functionality of the vector quantization and docking routines into a single utility and carries out searches for a range of k, $3 \leq k \leq 9$. A maximum of nine vectors is suitable for more complex single-molecule shapes, although in most cases a small number (≥ 3) are sufficient to determine the six rigid-body degrees of freedom for the docking. We evaluated two k-dependent criteria to select the optimum number k automatically. The first criterion is the codebook vector rmsd of the best fit for a given k. This criterion measures the geometric match of the reduced shape representation as a function of the spatial detail (should be minimal for the optimum k). The second criterion is the average statistical variability of the codebook vectors arising in a predefined number (default: 8) of statistically independent runs of the vector quantization. This criterion measures the k-dependent, shape-sensitive convergence of the vector quantization for a given single-molecule shape (should also be minimal for the optimum k). Under "Precision of Rigid-Body Docking" we evaluate the performance of the two criteria in simulated docking experiments.

Although most of the functionality of *qvol, qpdb,* and *qdock* has been superseded by *qrange,* we continue to develop and disseminate the former three programs. There are application areas where vector quantization and docking may require complementary functionality at a fixed k that is not suitable for implementation in the automatic *qrange* procedure. For example, *qdock* provides a choice of ranking criterion; best fits may be ranked by the correlation coefficient as an alternative to the codebook vector rmsd. This ranking by correlation coefficient is computationally expensive and not suitable for interactive *qrange* applications. The coefficients typically lie within a very narrow numeric range, i.e., fits based on the correlation coefficient alone are often ambiguous (Wriggers *et al.,* 1999). For these reasons users will rarely wish to use *qvol, qpdb,* and *qdock* for rigid-body docking. We note, however, that *qvol* and *qpdb* are required for flexible docking. In particular, distance constraints can be learned and implemented that freeze in essential degrees of freedom and reduce distortion in the flexibly fitted atomic structure (see "Discussion and Future Prospects").

Using low-resolution densities, the isocontouring program *volcube* (Fig. 1) generates wire-frame meshes or solid surfaces of isocontours that can be displayed, together with fitted atomic structures, using the free molecular graphics package VMD (Humphrey *et al.,* 1996). In the following section we describe a novel visualization tool, *volslice3d,* for analysis and visualization of volumetric data. While *volcube* and VMD still have a role in the visualization of fitted structures, *volslice3d* adds attractive new graphics capabilities to the package.

VISUALIZATION OF VOLUMETRIC DATA WITH *volslice3d*

Single-molecule density maps can be extracted from data sets that exhibit variable subunit composition by difference-mapping. The segmentation

FIG. 2. Screen shots of the *volslice3d* utility. Shown on the left side are x, y, and z cross sections of a cofilin-decorated actin filament (RWB mode). Shown on the right side is a block-representation of a thresholded map of troponin C and the corresponding GUI.

FIG. 3. Docking precision tested on simulated EM maps. (a) Troponin C (see Table 1): rmsd of the *Situs*-docked structure from the initial structure used for creating the simulated map as a function of the map resolution. The vector variability criterion (see text) was used to automatically select the optimum vector number k. (b) Scatter plot of the docking rmsd achieved for the 10 trial systems (see Table 1) for a vector number range $3 \leq k \leq 9$ as a function of the statistical vector variability. The color of the scattered data illustrates the resolution of the simulated EM maps: 2–20 Å (blue); 22–50 Å (green); 52–100 Å (red).

FIG. 4. Use of vector quantization for flexible docking. (a) The atomic structure of actin, modeled as described in Wriggers and Schulten (1999), is shown in brown. The target structure (blue) was created by distortion of the atomic coordinates using molecular dynamics simulations and constraints that induce an opening of actin's nucleotide-binding cleft. (b) Four codebook vectors were computed for the initial structure. (c) The resolution of the target structure was lowered to 15 Å, and the resulting 3D density was encoded by four vectors. (d) Flexible docking by the four pairs of codebook vectors, performed in the absence of atomic information of the target, lowers the C_α rmsd between the structures from 4.36 to 1.37 Å. (e) Two crystal structures of lactoferrin, PDB entries 1LFG (brown) and 1LFH (blue), show significant conformational differences (C_α rmsd 6.43 Å), particularly in the flexible N2 domain (lower left). (f) The initial structure was encoded by six vectors (each of the domains C1, C2, and N1 was encoded by a single vector; the flexible N2 domain was encoded by three vectors). (g) The resolution of the target structure (blue) was lowered to 15 Å, and the resulting 3D density was encoded by six vectors as in (f). (h) Flexible docking by the six pairs of codebook vectors lowers the rmsd between the structures to 2.72 Å. The scenes were rendered with *volcube* and with the molecular graphics program VMD (Humphrey *et al.,* 1996).

utility *floodfill* (Fig. 1) has been designed to extract contiguous regions of density from volumetric difference maps. To identify individual single-molecule regions in such difference maps it is necessary to inspect and measure the data sets. Existing commercial 3D visualization packages include functionality to analyze maps, but most of the features provided by such software are unnecessary for the task at hand. Therefore, we decided to make specific 2D and 3D graphics tools freely available, for measuring voxel positions within a map and for visualizing thresholded density levels, to support the use of the above *Situs* utilities.

The graphics program *volslice3d* is an improved version of the *volslice* utility distributed with the *Situs* package. The original *volslice* program displays 2D cross sections ("slices") of 3D data with 1960s era teletype graphics that uses ASCII characters for the rendering of thresholded density levels. This easy to use terminal-based method is intended for older UNIX workstations that are not capable of displaying 3D graphics. Here, we present a more advanced graphics alternative for interactive analysis and visualization of volumetric data. Although its functionality is currently limited to the rendering of cross sections, *volslice3d* serves as a prototype for a future Situs graphics interface.

Figure 2 presents screen shots of the *volslice3d* windows, the main window and the graphical user interface (GUI) window. The user can rotate the scene by dragging the mouse in the main graphics window. To manipulate the 3D view, the program also provides a simple and intuitive GUI. The first three widgets move the active x, y, and z slices through the density map. The positions of the slices are also adjustable with the arrow buttons or with the type-in text fields. The color box specifies the color mode used. Currently, three different color modes are supported: (i) Gray: Density values are mapped to gray levels by linear interpolation. (ii) RWB: The densities are mapped to a red–white–blue gradient. White corresponds to a density value of zero. Red corresponds to negative, and blue to positive, density values. (iii) RYG: The densities are mapped to a red–yellow–green gradient by linear interpolation. In all color modes, densities below a threshold (cutoff) value are not visible.

To change the color mode or the cutoff value *volslice3d* needs to recalculate the texture maps for the cross sections. This process takes only seconds on a typical workstation and is started with the "Apply" button. It is also possible to display all slices simultaneously in order to create a 3D volumetric effect (Fig. 2). This "block" mode is more expensive to compute. Therefore, the blocks can be switched on or off separately for each x, y, and z direction. The default settings for the display are set up during the startup in the UNIX shell with a *.svtrc* configuration file.

At Forschungszentrum Jülich, Germany, *volslice3d* is used in a VR environment. A headtracking system, stereo shutter glasses, and a dual-side projection system (see e.g., Krüger *et al.*, 1995) create an immersive graphics environment (van Dam *et al.*, 2000) and provide a more intuitive interaction with the 3D data. The latest development is the so-called *ad-box*, a special interface for VR devices. It is possible to use devices with multiple potentiometers and buttons as control units for the visualization. As a first demonstration of the ad-box, we designed a cube-shaped input device with perpendicular rods for each *x*-, *y*-, and *z*-axis. The rods are internally connected to potentiometers. When a rod is moved, the corresponding slice moves likewise in the 3D environment. The movement of the entire cube is tracked, so when the user rotates the cube, the *volslice3d* scene will rotate in the VR environment.

To run the *volslice3d* program the user must have access to a modern UNIX workstation (or PC/Linux) with hardware accelerated 3D graphics (e.g., a Silicon Graphics O2). The program offers stereo support (with shutter glasses such as CrystalEyes) and headtracking support (with the Polhemus Fastrak system). Binary executables for a variety of platforms and documentation of the *volslice3d* graphics and VR capabilities are available at http://www.fz-juelich.de/vislab/virtual/volslice3d.

PRECISION OF RIGID-BODY DOCKING

Using *pdblur* we convoluted the atomic structures of 10 trial proteins (Table I) with Gaussian kernels whose resolution was varied from 2 to 100 Å (in 2-Å increments), to measure the precision and reliability of the fitting. For each system the resolution-dependent rmsd of the docked from the initial structure was evaluated. Figure 3a shows the resolution-dependent rmsd of troponin C as a typical example. The docking rmsd initially increases with higher resolution value, but the precision of the docking is more than one order of magnitude better than the nominal resolution of the simulated map. The slow change in docking rmsd ends at a critical resolution value, at which the docking breaks down (docking rmsd >10 Å) due to a catastrophic mismatch. At higher resolution values, i.e., beyond the range defined by the critical value, the docking is unstable and fluctuates between mismatches and matches that can be demarcated roughly at the 10-Å rmsd level.

Table I demonstrates that the viable resolution range for the docking is system-dependent due to

TABLE I
Docking Characteristics of Simulated EM Maps

Protein used for generating bead model	PDB entry	Var. criterion		RMSD criterion		Oligomeric symmetry	Degeneracy of best fit[c]
		Range[a] (Å)	Accuracy[b] (Å)	Range[a] (Å)	Accuracy[b] (Å)		
Catalase	7cat	84	0.4	4	—	4×	4×
β-4-Integrin	1qg3	54	0.6	54	1.0	1×	1×
Chymotrypsinogen A	2cga	40	1.5	60	2.2	1×	1×
Myoglobin	1mbn	50	1.0	48	1.0	1×	2×
Nitrito-reductase	2nrd	30	0.7	18	—	3×	6×
Ovalbumin	1ova	44	1.3	44	0.8	1×	1×
Spermadhesin	1spp	48	1.1	46	0.8	1×	1×
Superoxide dismutase	1xso	72	0.5	44	0.9	2×	2×
Troponin C	1top	36	0.7	38	0.7	1×	1×
αβ-Tubulin	1tub	38	0.8	34	0.8	1×	1×

Note. For docking rmsd evaluations, the fitted structure with the lowest deviation from the target structure was selected among any degenerate fits.

[a] The stated viable resolution range is the maximum resolution with docking rmsd <10 Å.

[b] The stated accuracy value is the docking rmsd at 20-Å resolution.

[c] The degeneracy is the number of optimum fits (within viable resolution range) that were found to cluster near the optimum score (see text).

the different sizes of the proteins studied. Also, the limiting resolution depends in some cases on the criterion used for the selection of the optimum vector number k. If the vector variability criterion (see "Design of Situs 1.4") was used, the docking was stable at least to 30 Å, a resolution that can be achieved in most experimental EM image reconstructions. The docking precision (measured at 20-Å resolution) was on the order of 1 Å in all 10 trial systems. For 8 trial systems the choice of selection criterion did not alter significantly the precision and the limiting resolution. However, the use of the vector rmsd criterion resulted in an abridged resolution range in 2 cases (catalase: 4 Å; nitrito-reductase: 18 Å; see Table I). In these two cases it would be problematic to predict the correct match with the vector rmsd criterion. Instead, the variability criterion is more suitable.

We investigated further how the codebook vector variability correlates with the actual docking rmsd achieved, not only for the optimum k but for the full range $3 \leq k \leq 9$. Figure 3b illustrates the docking rmsd (10 trial systems, resolution 2–100 Å, $3 \leq k \leq 9$) as a function of the observed vector variability. Clearly, one would expect a high correlation between the variability and the actual rmsd if the variability were a good criterion for selecting k. Indeed, the correlation coefficient (as computed by standard linear regression analysis) is 0.52, and a correlation is clearly visible in the double-logarithmic scatter plot (Fig. 3b). Moreover, we have color-coded three zones of resolution in the plot: high (2–20 Å), medium (22–50 Å), and low (52–100 Å). One would expect that the variability separates out the high-resolution cases where the docking is well behaved. Clearly, the high-resolution cases that correspond to experimentally accessible resolutions cluster at low vector variability (and low docking rmsd values), whereas the medium- and low-resolution clusters are found at progressively higher variability and rmsd values. Note that not all high-resolution cases exhibit low variability and docking rmsd, but such mismatches are avoided if the optimum k is selected based on the vector variability minimum.

We have also carried out a similar statistical analysis for the vector rmsd criterion (not shown). The coefficient for the correlation between the vector rmsd and the docking rmsd was 0.50. The slightly reduced coefficient (compared to the vector variability case) again indicates that the vector rmsd is a less stringent criterion for selecting k than the vector variability.

The highest-scoring fits are degenerate in certain cases. This degeneracy (Table I) closely follows the number of symmetry-related oligomeric subunits in the trial systems. Only in two cases did the algorithm return spurious fits that were not symmetry-related (myoglobin and nitrito-reductase). In these systems each correct (or symmetry-related) high-scoring fit was paired with an additional incorrect fit that was indistinguishable by means of the codebook vector docking score. A detailed inspection of the program output revealed that this ambiguity arises in cases where $k = 3$ vectors are chosen automatically. In the case of the nitrito-reductase trimer, the three vectors form an equilateral triangle, and the 3!

= 6 possible docking conformations are degenerate by default. In the case of myoglobin, the vectors form an isosceles triangle whose two-fold symmetry axis is responsible for the twofold degeneracy of the best fit.

An exclusion of the special planar case $k = 3$, i.e., the narrowing of the vector range to $4 \leq k \leq 9$, restored the unequivocalness of the prediction for myoglobin and nitrito-reductase and restored the suitability of the rmsd criterion for catalase and nitrito-reductase (limiting resolution >70 Å). However, the use of $k = 3$ vectors is important for superoxide dismutase, myoglobin, and chymotrypsinogen A, since the increased unequivocalness for $4 \leq k \leq 9$ came at the price of a significantly (>15 Å) reduced limiting resolution in these three cases (data not shown). Thus, the narrowing of the search range to $4 \leq k \leq 9$ vectors cannot be generally recommended. We solved the problem of maximizing both unequivocalness and resolution range by allowing the user to make an informed decision to exclude $k = 3$ if ambiguities arise in this special case.

In summary, our validation analysis demonstrated that the vector-based shape-based docking can be performed by nonexpert users reliably for resolutions up to 30 Å without requiring knowledge of the details of the algorithm. We recommended the use of the variability criterion for selecting the optimum number k of vectors and the use of a range $3 \leq k \leq 9$. The user is informed of the positional and rotational accuracy based on a statistical analysis and alerted to the possibility of ambiguous matches.

FLEXIBLE DOCKING

The reduced representation of data sets by positional markers inspired us to characterize conformational differences between high- and low-resolution biological data by combining vector quantization with molecular mechanics simulations (Wriggers *et al.*, 2000). The codebook vectors $\mathbf{x}_i$ and $\mathbf{y}_j$ divide each of the two compared 3D data sets into a number of subregions

$$V_i^x = \{\mathbf{q} \in \mathcal{R}^3 \| \|\mathbf{q} - \mathbf{x}_i\| \leq \|\mathbf{q} - \mathbf{x}_k\| k = 1, \cdots, k\} \qquad i = 1, \cdots, k \tag{1}$$

(similarly for V_j^y), known in the literature as Voronoi cells. The codebook vectors $\mathbf{x}_i$ and $\mathbf{y}_j$ coincide with the centroids of the data density distribution within each cell V_i^x and V_j^y. We may exploit this by adding a constraint energy function to the Hamiltonian of an MD simulation that penalizes differences between the centroids of each cell $V_{I(j)}^x$ from the corresponding vectors $\mathbf{y}_j$. If internal motions are present, the MD refinement of the high-resolution structure will optimize the conformation by aligning the centroids of the $V_{I(j)}^x$ to the $\mathbf{y}_j$, thereby bringing the data sets into register. The method was first applied in the flexible fitting of a crystallographic conformation of ribosomal elongation-factor G to a difference-density map of the ribosome-bound conformation at 17-Å resolution (Wriggers *et al.*, 2000).

We demonstrate here the validity of the flexible docking approach on pairs of structures of actin and lactoferrin (Fig. 4). The development of the *pdblur* utility enabled us to create simulated low-resolution maps for the demonstration. The resolution of a target atomic structure was lowered by convolution with a Gaussian kernel to 15 Å. Subsequently, the alternate structure (in a different conformation) was flexibly fitted to the low-resolution density in the absence of any atomic information about the target structure. Figure 4 presents the actin (Figs. 4a–4d) and lactoferrin (Figs. 4e–4h) demonstration results. Using four codebook vectors to encode structures of actin, the flexible docking reproduces a 4.4-Å rmsd conformational change with an accuracy of 1.4 Å. It is assumed that the tertiary fold of the actin structure remains locally conserved. Hence small localized changes in actin remain undetected by the low-resolution fitting. In the case of lactoferrin (PDB entries 1LFG and 1LFH; six codebook vectors), the 6.4-Å conformational difference between the two crystallographic conformations is reproduced with an accuracy of 2.7 Å. The accuracy is limited by localized differences between the crystal structures: In the regions where the lactoferrin structures are conserved (75% of the residues, Nos. 1–191, 251–321, and 344–691), the resulting fit is improved (rmsd 1.7 Å).

The minimizations of the MD Hamiltonian were carried out with X-PLOR (Brünger, 1992) using the default parameters of the CHARMM united-atom force field (Brooks *et al.*, 1983), version 19. The constraint energy function that penalizes differences between the centroids of each cell $V_{I(j)}^x$ from the corresponding vectors $\mathbf{y}_j$ was implemented using NOE constraints (Brünger, 1992). The constrained structures were subject to 1000 steps of Powell energy minimization. Hookean potentials with force constants of 3100 kcal mol^{-1} Å^{-2} were employed, as described (Wriggers *et al.*, 2000). The details of the computational setup for the lactoferrin calculation and the X-PLOR/CNS-compatible procedures have been made available in a tutorial at http://situs.scripps.edu/tutorial_flex.html.

DISCUSSION AND FUTURE PROSPECTS

The lack of prior knowledge about the mutual correspondence of features complicates the shape-

based registration of 3D data sets from various biophysical sources. By testing *Situs* systematically against simulated low-resolution data we addressed four questions of interest in applications of rigid-body docking algorithms: (i) whether the docking is reliable for a range of resolutions typically encountered in EM, (ii) whether correct solutions are inadvertently missed due to the reduced search space, (iii) whether incorrect (spurious) fits give rise to ambiguities, and (iv) whether we can identify a suitable criterion for an automatic selection of the level of spatial detail in the data representations.

Our results indicate that the method is reliable for resolution values up to 30 Å on idealized (noise-free) data sets, and the fitting accuracy is on the order of 1 Å in terms of atomic rmsd. This is likely a lower bound of the fitting inaccuracy that can be expected when using experimental maps, but nevertheless the results are very satisfying, given *Situs*' indirect comparison of multiresolution data by vector quantization. It is clear that realistic EM reconstructions may differ from the idealized low-resolution maps used in this study. The point-spread function introduced by the microscope is non-Gaussian, and there may be noise and disorder effects due to image-processing that alter the shape of the averaged 3D image reconstruction. A more detailed assessment of the fitting performance must wait until it can be demonstrated on a practical example.

In our 10 trial systems none of the correct solutions were missed as a result of the reduced spatial detail of the vector-based data representation. We observed spurious fits in 2 cases involving the special planar case of $k = 3$ vectors. Unless the sides of the triangle defined by three vectors differ significantly in their lengths, it is clear that such a simplified data representation may give rise to ambiguities, i.e., a twofold degeneracy in the case of an isosceles triangle and a sixfold degeneracy in the case of an equilateral triangle. In the current release the *qrange* utility alerts the user if such ambiguities arise for $k = 3$.

In the original article we proposed two selection criteria for selecting the optimum level of spatial detail in the reduced codebook vector-based representations. It came as a surprise that the vector variability appears to be a more stringent criterion for selecting the optimum number k than the vector rmsd when judged by the actual rmsd achieved after docking of the atomic structures. We note that a low codebook vector rmsd does not necessarily guarantee a low docking rmsd, due to the effect of matching ambiguities and due to noise introduced by the vector variability. Low vector variability, however, guarantees that the vector positions are unique and reproducible, which reduces the risk of inadvertent mismatches that restrict the range of resolutions suitable for the docking. We observed that the vector variability was minimal for integer multiples of the numbers of oligomeric subunits in the cases of catalase and nitrito-reductase (data not shown). This indicates that the variability reflects the shape and symmetry of the data at hand and thereby constitutes a measure of the quality of a given reduced representation. We also note that the analysis of vector variability provides a statistical estimate of the fitting accuracy. In the current release, *qrange* and *qdock* predict the positional and angular fitting accuracy based on the observed statistical vector variabilities.

We have recently developed *Situs* utilities for the docking of atomic structures to SAXS bead models (Wriggers and Chacón, in press). The docking to EM maps compared to the docking to bead models differs in the range of viable resolution and in the onset of ambiguous results. Most of the trial systems required a bead diameter <20 Å for accurate docking, which was smaller than the corresponding resolution range values in Table I. Also, spurious fits using bead models were found in one additional (third) case: chymotrypsinogen A. The higher accuracy and unequivocalness of EM-based docking can be attributed to the smoothness of the low-resolution density (as opposed to the segmented SAXS beads). The comparison suggests that the smoothness of EM maps yields vector positions that are more insensitive to resolution changes.

In the two demonstrations of flexible docking the procedures faithfully reproduced conformational differences with a precision of <2 Å if atomic structures are locally conserved. Whether this assumption holds depends on the nature of the conformational difference between two isoforms, which (in realistic situations) is not known a priori. However, it has been shown that about 60% of protein domain rearrangements documented in the PDB are hinge-bending motions where structural domains remain largely intact (Gerstein *et al.*, 1994; Gerstein and Krebs, 1998). It is plausible, at least for hinge-type domain motions, that the low-resolution flexible fitting approach visualizes conformational changes with a precision of single amino acid residues.

The identification of spatial features by vector quantization is sensitive to noise and shape changes due to experimental limitations. One of the open questions in flexible docking is how to maintain the stereochemical quality of a fitted structure, since any overfitting to noise-induced vector displacements would compromise the quality of the atomic model. We have addressed this question in the current release of *Situs* by providing functionality for

learning and "freezing" distances between codebook vectors. Intervector distances are constrained in *qvol* using the SHAKE algorithm (van Gunsteren and Berendsen, 1977; Ryckaert *et al.*, 1977). The resulting vector skeletons (distance-constrained vectors) can be used to eliminate the degrees of freedom that are deemed inessential for the flexible docking. The skeleton-based fitting approach, which is currently in the "beta stage" of development, provides additional robustness against the effects of noise and experimental uncertainty. A detailed account of the effect of intervector constraints on the stereochemical quality of flexibly fitted atomic structures is in preparation.

Skeleton-based fitting has already been applied in the rigid-body fitting of subunits into heterogeneous aggregates (Galkin *et al.*, 2001). Since the single-molecule docking algorithms have now reached a certain level of maturity, we will concentrate our future development efforts mainly on the challenging problem of docking atomic subunits into low-resolution maps of entire biomolecular aggregates.

The current implementation of the *Situs* package is available at http://situs.scripps.edu. The *volslice3d* graphics utility is available at http://www.fz-juelich.de/vislab/virtual/volslice3d.

We thank Pablo Chacón for providing the trial systems and for discussions regarding the modeling and docking of SAXS data. The ongoing development of *Situs* is supported by NIH Grants P41-RR-12255-02 and 1R01-GM62968-01.

REFERENCES

Belnap, D. M., Kumar, A., Folk, J. T., Smith, T. J., and 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.

Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4,** 187–217.

Brünger, A. (1992) X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University.

Chacón, P., Díaz, J. F., Morán, F., and Andreu, J. M. (2000) Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *J. Mol. Biol.* **299,** 1289–1302.

Galkin, V. E., Orlova, A., Lukoyanova, N., Wriggers, W., and Egelman, E. H. ADF stabilizes an existing state of F-actin and can change the tilt of F-actin subunits. *J. Cell Biol.* **153,** 77–86.

Gerstein, M., and Krebs, W. (1998) A database of macromolecular motions. *Nucleic Acids Res.* **26,** 4280–4290.

Gerstein, M., Lesk, A. M., and Chothia, C. (1994) Structural mechanisms for domain movements in proteins. *Biochemistry* **33,** 6739–6749.

Humphrey, W. F., Dalke, A., and Schulten, K. (1996) VMD—Visual molecular dynamics. *J. Mol. Graphics* **14,** 33–38.

Kikkawa, M., Okada, Y., and Hirokawa, N. (2000) 15 Å resolution model of the monomeric kinesin motor, KIF1A. *Cell* **100,** 241–252.

Krüger, W., Bohn, C.-A., Fröhlich, B., Schüth, H., Strauss, W., and Wesche, G. (1995) The responsive workbench: A virtual work environment. *IEEE Comp.* **28,** 42–48.

Llorca, O., Martin-Benito, J., Ritco-Vosonvici, M., Grantham, J., Hynes, G. M., Willison, K. R., Carrascosa, J. L., and Valpuesta, J. (2000) Eukariotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J.* **19,** 5971–5979.

Moores, C. A., Keep, N. H., and Kendrick-Jones, J. (2000) Structure of the utrophin actin-binding domain bound to F-actin reveals binding by an induced fit mechanism. *J. Mol. Biol.* **297,** 465–480.

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comp. Phys.* **23,** 327–341.

van Dam, A., Forsberg, A. S., Laidlaw, D. H., LaViola, J. J., and Simpson, R. M. (2000) Immersive VR for scientific visualization: A progress report. *IEEE Comp. Graph. Appl.* **20,** 26–52.

van Gunsteren, W. F., and Berendsen, H. J. C. (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* **34,** 1311–1327.

Wriggers, W., and Chacón, P. Using *Situs* for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *J. Appl. Cryst.*, in press.

Wriggers, W., and Schulten, K. (1999) Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. *Proteins Struct. Function Genet.* **35,** 262–273.

Wriggers, W., Milligan, R. A., Schulten, K., and McCammon, J. A. (1998) Self-organizing neural networks bridge the biomolecular resolution gap. *J. Mol. Biol.* **284,** 1247–1254.

Wriggers, W., Milligan, R. A., and McCammon, J. A. (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125,** 185–195.

Wriggers, W., Agrawal, R. K., Drew, D. L., McCammon, J. A., and Frank, J. (2000) Domain motions of EF-G bound to the 70S ribosome: Insights from a hand-shaking between multi-resolution structures. *Biophys. J.* **79,** 1670–1678.