Domain Motions of EF-G Bound to the 70S Ribosome: Insights from a Hand-Shaking between Multi-Resolution Structures

Willy Wriggers,*† Rajendra K. Agrawal,‡§ Devin Lee Drew,* Andrew McCammon,* and Joachim Frank‡¶

*Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California, San Diego, La Jolla, California 92093-0365; †Departments of Cell Biology and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037; ‡Department of Biomedical Sciences, State University of New York at Albany, Empire State Plaza, Albany, New York 12201-0509; and ¶Howard Hughes Medical Institute, Health Research, Inc. at the §Wadsworth Center, Empire State Plaza, Albany, New York 12201-0509 USA

ABSTRACT Molecular modeling and information processing techniques were combined to refine the structure of translocase (EF-G) in the ribosome-bound form against data from cryoelectron microscopy (cryo-EM). We devised a novel multi-scale refinement method based on vector quantization and force-field methods that gives excellent agreement between the flexibly docked structure of GDP · EF-G and the cryo-EM density map at 17 Å resolution. The refinement reveals a dramatic "induced fit" conformational change on the 70S ribosome, mainly involving EF-G's domains III, IV, and V. The rearrangement of EF-G's structurally preserved regions, mediated and guided by flexible linkers, defines the site of interaction with the GTPase-associated center of the ribosome.

INTRODUCTION

Recent research using a diversity of structural probes suggests that many processes in the cell involve dynamic interactions among macromolecules in a highly organized way. Large molecules with multiple highly specific binding sites act both as templates and multiple catalysts in the functional interaction with ligand molecules. Examples are the ribosome, spliceosome, and transcription complexes. Such systems have been called "macromolecular machines" (Alberts, 1998), a term intended to invoke complexity, dynamic behavior, and precision. To fully understand the action of such a machine, one must know all dynamic states of all components in their correct time sequence. The large size and the large number of ligand-binding states and conformations of macromolecular machines make it unlikely that atomic-resolution structures will be available for more than a few. The demands for routine crystallization, data collection, and interpretation of molecules $>1\ M_d$ often complicate solving the atomic structure directly, as exemplified by the considerable effort that has gone into obtaining crystallographic electron-density maps of the ribosome (Ban et al., 1999; Cate et al., 1999; Clemons, Jr. et al., 1999; Tocilj et al., 1999). Atomic resolution, however, enables a detailed understanding of macromolecular interactions.

Within the past decade, cryoelectron microscopy (cryo-EM) has emerged as a powerful technique of three-dimensional (3D) imaging that does not require crystallization of large assemblies. Cryo-EM of single particles (Frank, 1996) poses no restrictions on the conformational range of multicomponent complexes, and is capable of yielding low-resolution density maps that allow crystal structures of components to be fitted and docked. Thus, in principle, an atomic model of a molecular machine in a particular state of processing can be built from its components, provided that the fitting procedure incorporates the constraints of molecular interactions and mechanics. The progress in x-ray crystallography of ribosomes (Ban et al., 1999; Cate et al., 1999; Clemons, Jr. et al., 1999; Tocilj et al., 1999) makes it important and timely to develop ways of studying dynamical states of such assemblies. Although the potential of the "hybrid" approach to molecular structure has been apparent in many recent applications (DeRosier and Harrison, 1997; Stowell et al., 1998), as yet there has been no satisfactory solution to the problem of docking *deviating* atomic structures into cryo-EM densities. In this article we present a general solution using a novel flexible docking procedure based on a vector quantization algorithm and force-field methods. The refinement seeks to preserve the conformation of the x-ray structure locally, while adjusting the position of larger segments based on a small number of global constraints represented by so-called *codebook vectors* and their *Voronoi cells.* We show that application of this technique to elongation factor G (EF-G) leads to a detailed model of major domain rearrangements in the protein and defines its position relative to the ribosome's GTPase-associated center.

METHODS

Vector quantization is a data-clustering technique that has been used since the 1950s for digital signal compression in engineering applications such as digital speech and image processing (Makhoul et al., 1985). The method allows one to approximate the probability density distribution $m(\mathbf{r})$ of 3D data signals, $\mathbf{r} \in \mathcal{R}^3$, using a finite number of so called codebook vectors $\mathbf{w}_i \in \mathcal{R}^3$, $i = 1, \cdots, N$. The discretization of the data (here, biological structural data at variable resolution) is performed by systematic lowering of the distortion error

Received for publication 11 January 2000 and in final form 12 May 2000.

Address reprint requests to Dr. Willy R. Wriggers, Dept. of Molecular Biology TPC6, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. Tel.: 858-784-8823; Fax: 858-784-8688; E-mail: wriggers@scripps.edu.

© 2000 by the Biophysical Society

0006-3495/00/09/1670/09 $2.00

E of the discrete data representation (Makhoul et al., 1985):

$$E = \sum_j m(\mathbf{r}_j)\|\mathbf{r}_j - \mathbf{w}_{i(j)}\|^2, \quad (1)$$

where the index j of a data point $\mathbf{r}_j \in \mathcal{R}^3$ is the atom index of a crystal structure or the grid index of a density map, $m(\mathbf{r}_j)$ is the weight, i.e., the corresponding atom mass or electron density, respectively, and $i(j)$ is the index of the codebook vector $\mathbf{w}_{i(j)}$ closest to $\mathbf{r}_j$ in terms of Euclidean distance. E can be interpreted simply as the mean-squares deviation of the discrete set of codebook vectors from the encoded 3D data. Several algorithms exist that solve the vector quantization problem by systematic updating of the $\mathbf{w}_i$ until the minimum of E is found. As usual in optimization, one is interested in finding a unique distribution $\{\mathbf{w}_j\}$ corresponding to the global minimum of E. We use topology-representing neural networks, which avoid getting trapped in local minima of E by stochastic gradient descent on a smooth energy landscape that slowly converges toward E (Martinetz et al., 1993). Convergence assures that the set $\{\mathbf{w}_j\}$ represents the underlying biological data in a unique, well-defined, and reproducible fashion. The codebook vectors form a set of control points or landmarks that provide information about the shape and density distribution of the encoded biological object. The sole free parameter is the number N of vectors chosen for the representation, which depends on the desired level of detail. By superimposing two structural data sets with their corresponding pairs of codebook vectors one can perform rigid-body docking of single molecule data or construct protein aggregates from their subunits (Wriggers et al., 1998, 1999).

We combine the technique for the first time with molecular mechanics simulations to identify conformational differences between high- (h) and low (l)-resolution biological data. Let us first consider that the codebook vectors $\mathbf{w}_i^{l,h}$ divide each data set (l, h) into a number of subregions

$$V_i^{l,h} = \{\mathbf{q} \in \mathcal{R}^3 \mid \|\mathbf{q} - \mathbf{w}_i^{l,h}\| \le \|\mathbf{q} - \mathbf{w}_k^{l,h}\| k = 1, \cdots, N\}$$
$$i = 1, \cdots, N \quad (2)$$

known in the literature as Voronoi cells. Depending on the number N of vectors chosen, these Voronoi cells characterize all or parts of segmented subdomains and protruding features of the biological object. It can be shown that the codebook vectors $\mathbf{w}_i^{l,h}$ coincide with the centroids of the density distribution $m(\mathbf{r})^{l,h}$ within each cell $V_i^{l,h}$ (Kohonen, 1995). We can exploit this by adding a constraint energy function to the Hamiltonian of a molecular mechanics simulation that penalizes differences between the centroids of each cell V_i^h from the corresponding vectors $\mathbf{w}_i^l$. In the absence of internal motions, the centroids of the high-resolution cells V_i^h are superimposable (Wriggers et al., 1998, 1999) with the corresponding low-resolution vectors $\mathbf{w}_i^l$ by rigid-body fitting. If internal motions are present, a molecular mechanics refinement of the high-resolution structure will optimize the conformation by aligning the centroids of the V_i^h to the $\mathbf{w}_i^l$, thereby bringing the data sets into register.

In principle, the resolution of the fitting can be improved by increasing the number N of codebook vectors. However, there are practical limitations. If there are regions with unaccounted EM density (or unaccounted atomic data) not all vectors converge toward equivalent features in the two data sets if their number N is increased. This poses the challenge of how to take full advantage of the information in the data sets given their experimental limitations. The solution to this problem is to carry out vector quantizations for an entire range of N and, for each N, to select only pairs of corresponding vectors based on perceived local similarity in the vector sets $\{\mathbf{w}_i^{l,h}\}$. This can be done very efficiently by an inspection of the resulting vectors in concert with the encoded structures using a molecular graphics package. It is possible to automate this approach by a systematic evaluation of the distance-matrices of the vectors. However, if parts of the data sets are not accounted for, the modeler ultimately bears responsibility for the assignment of corresponding features. Hence, vector quantization can be seen as a tool that provides point landmarks for flexible registration, to be selected by the modeler. Even if the data correspondence is imperfect, most codebook vectors will still align themselves with clearly identifiable features. Thus, low-confidence regions can be represented by the modeler at low resolution using relatively few selected vectors, and high-confidence regions can be characterized using a larger number of vectors.

The molecular mechanics calculations described in this paper were carried out with X-PLOR (Brünger, 1992) using parameters of the CHARMM all-atom force field (Brooks et al., 1983), versions 22 and 24, for the protein and GDP, respectively, and the TIP3(P) water model (Jorgensen et al., 1983). A dielectric constant $\epsilon = 1$, and a cut-off distance of 12 Å for non-bonded interactions were chosen. The newly modeled regions of EF-G were optimized by energy minimization, followed by refinement in aqueous solution. Powell energy minimization was carried out with using weak Hookean constraints (force constant 0.2 kcal mol^{-1} Å^{-2}) to preserve the side chain positions of the original crystal structure, and with strong Hookean constraints (force constant 20 kcal mol^{-1} Å^{-2}) to preserve the original crystal C_α positions. Subsequently, the biopolymer was immersed in a shell of explicit water molecules of 6 Å thickness, which corresponds to approximately two layers of water molecules. To fully solvate the cleft between domains I and V that clearly is solvent-accessible in the EM map, a 13° rotation of domain V in the direction normal to the domain I interface was applied before the flexible refinement. Following the above minimization and solvation procedures, the system size of the final model was 18,645 atoms (2,575 water molecules, including 21 from the crystal structure).

The constraint energy function that penalizes differences between the centroids of each cell V_i^h from the corresponding vectors $\mathbf{w}_i^l$ was implemented using NOE constraints in X-PLOR (Brünger, 1992). Hookean potentials with force constants 31, 310, 3100, and 31,000 kcal mol^{-1} Å^{-2} were used. Force constants ≥310 kcal mol^{-1} Å^{-2} were sufficient to overwhelm the standard molecular mechanics interactions. Structures produced by using stronger constraints did not differ significantly from those at force constant 310 kcal mol^{-1} Å^{-2}, which are presented in this work.

In addition to energy-minimization, *simulated annealing* calculations of length 170 ps with a maximum temperature 500 K (Wriggers and Schulten, 1998) were carried out. These simulations served as a control to test the stability of the predicted conformations and provide a simple measure of the accuracy of the modeling. After simulated annealing, the completed structure exhibited a C_α rms deviation of 0.9 Å from the initial structure (see Fig. 1 *a*). The observed C_α rms variability among flexibly fitted structures (after simulated annealing) was 2.1 Å.

In addition to the studies on EF-G we have also tested the validity of our approach on pairs of structures of actin and lactoferrin (Wriggers, in preparation). 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. Using four codebook vectors to encode simulated and crystal structures of actin, the method reproduced various conformational changes with an accuracy of 1.4 Å, measured by C_α rms deviation. Our approach assumes that the crystal structure remains locally conserved, hence small-scale changes can not be picked up by the low-resolution fitting. For example, in the case of lactoferrin (PDB entries 1LFG and 1LFH; six codebook vectors), the C_α rms deviation between the two conformations was lowered from initially 6.4 to 2.7 Å—but not further—due to such local rearrangements. In the regions where the lactoferrin structures are conserved (75% of the residues, 1–191, 251–321, and 344–691), the resulting fit was improved (C_α rms deviation 1.7 Å).

RESULTS

Fitting EF-G to the ribosome

EF-G acts as a GTP-dependent catalyst of protein synthesis through transient interactions with the ribosome (Kaziro, 1978). During the peptide elongation cycle, in which

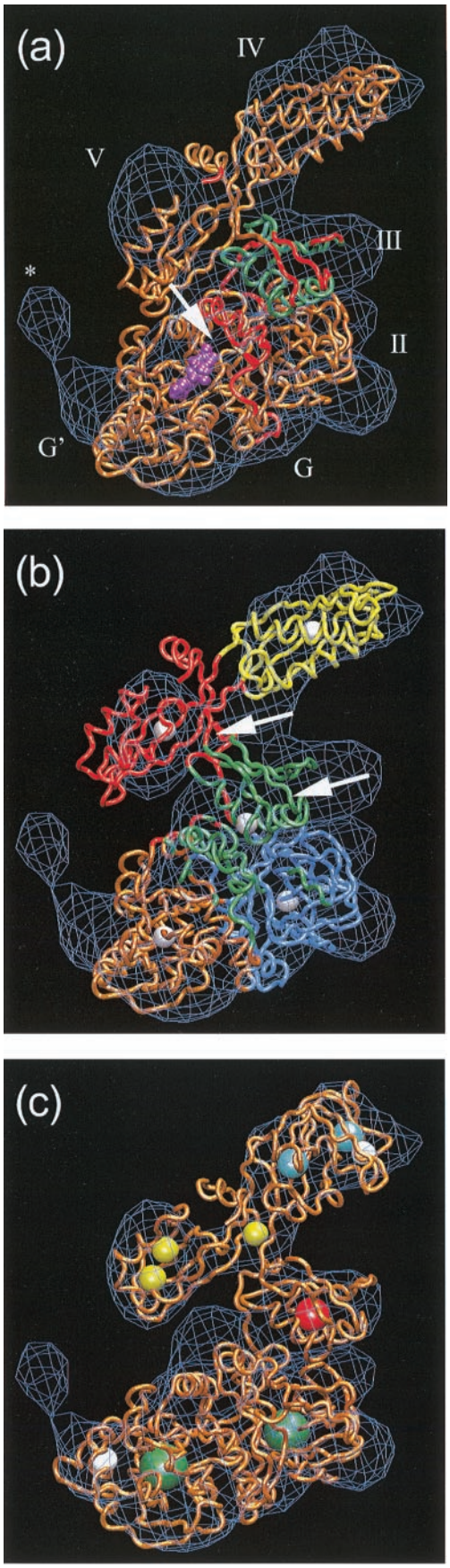
mRNA advances by one codon, EF-G facilitates the translocation of the A- and P-site tRNAs to the P and E sites, respectively. The crystal structures of EF-G in its nucleotide-free form (Ævarsson et al., 1994) and in complex with GDP (Czworkowski et al., 1994) are very similar. The structure is divided into five domains, numbered I–V. Domain I is again divided into the GDP binding subdomain G and an insertion, subdomain G′.

The absence of any discernible nucleotide-dependent conformational change (Agrawal et al., 1999; Czworkowski and Moore, 1997) suggests that GTP hydrolysis induces an observed change in the ribosome (Agrawal et al., 1999) mainly by modulation of EF-G's binding affinity to the ribosome, similar to other G-proteins that perform a nucleotide-dependent activation of target protein binding sites (Bourne et al., 1990; Czworkowski and Moore, 1997). However, there are many indications that the conformation of EF-G bound to the ribosome is not the same as in solution. Significant differences between the conformations of ribosome-bound and free EF-G are indicated by the observation that fusidic acid binds to EF-G only in complex with the ribosome (Willie et al., 1975), and that all known fusidic acid-resistant mutants are mutants of EF-G (Johanson et al., 1996). That EF-G changes its conformation upon ribosome binding is indicated also by an ~1000-fold decrease of the effective dissociation constant of the GMPPCP nucleotide upon EF-G binding to the ribosome (Baca et al., 1976). The stabilizing effect suggests an allosteric mechanism that links ribosome-induced conformational changes with the nucleotide binding pocket. Finally, the visually most intriguing evidence for conformational differences was provided recently in the form of cryo-EM image reconstructions of bound EF-G. These images, which were obtained from difference maps of free ribosomes and of various ribosome–EF-G complexes in GDP and GTP states, are clearly incompatible with the shape of the crystal structure (Agrawal et al., 1998, 1999).

The fitting of the GDP · EF-G crystal structure (Czworkowski et al., 1994) to the cryo-EM maps is further complicated by the fact that 112 of 691 amino acid residues, including the effector region and domain III, are undefined. Coordinates for the disordered domain III were obtained from the nucleotide-free structure (Ævarsson et al., 1994) (residues 409–416, 425–444, 448–470), and fitted as shown in Fig. 1 *a*. To predict the fold of the remaining undefined regions of the polypeptide chain (Fig. 1 *a*) we used Holley/Karplus secondary structure prediction. The backbone of the undefined chain was folded according to the secondary structure assignment (Holley and Karplus, 1989) and fitted into the structure. Subsequently, the side chains were placed after a spin search and the geometry of the constructed peptide chain was optimized. All modeling calculations were performed with Quanta 97 (MSI, 1997). Most of the predicted peptides (residues 1–6, 400–408, 417–424, 445–447, 471–475, 690–691) comprised only a small number of amino acids, and their fitting into the crystal coordinates was unambiguous. Residues 40–67, however, including the "switch 1" γ-phosphate sensing loop, form a long chain and pose a more challenging problem. The switch 1 region is well preserved among G-proteins (Bourne et al., 1990), and the assigned secondary structure was found in good agreement with the secondary structure of the related chain in the factor EF-Tu (Berchtold et al., 1993). The tertiary fold of residues 40–67 resembles that of the equivalent region in EF-Tu as identified by sequence alignment (Bourne et al., 1991). In particular, the conserved (γ-phosphate sensing) Thr-64 in the switch 1 loop was placed into the vicinity of the β-phosphate of GDP. The completed structure (residues 1–691) was further refined using energy minimization and simulated annealing protocols in explicit solvent.

If a rigid-body alignment of domains G, II, and IV is attempted, domains G′, III, and V deviate from corresponding low-resolution features (Fig. 1 *a*). We note here that some of the regions in the difference map will remain unexplained even when allowing conformational changes of the completed x-ray structure. These positive densities appear to be caused by conformational changes in the ribosome (Agrawal et al., 1998, 1999, 2000). For example, the pronounced arc-like extension marked by the asterisk was interpreted as an interaction of the G′ domain with a protrusion at the base of the L7/L12 stalk. Although this extension is shown in Fig. 1, it was removed from the EM map for a more accurate fitting: above a certain threshold level (17% of the maximum density), the arc becomes disconnected. The arc density was isolated and subtracted with programs of the Situs (Wriggers et al., 1999) package (data not shown). The quality of the fit was assessed by the correlation coefficient, as defined by Wriggers and co-workers (1999). [We note that the magnitudes of the coefficient presented here are lower than those reported by Agrawal and co-workers (1999). In their computation of the coefficients the resolution of the atomic model was reduced to that of the cryo-EM density. In the current work, we computed the scalar product of high- and low-resolution data directly by linear interpolation of the mass-weighted atomic coordinates to the eight proximal voxels on the density lattice (Wriggers et al., 1999). This method of computation results in a lower correlation coefficient.] The original difference density map was used for the computation of the correlation coefficient, which measured 0.70 in the first case (Fig. 1 *a*).

An overview of the hand-shaking method between high- and low-resolution structural data is given in Fig. 1, *b* and *c*. The terminology and algorithmic details are described in Methods. First, we carried out a vector quantization of the volumetric data set using five vectors. Then the crystal structure was partitioned with an equal number of codebook vectors, and the Voronoi cells were computed. Next, the centroids of the high-resolution Voronoi cells were forced to coincide with the corresponding vectors from the low-



resolution quantization. This optimization took place in a realistic molecular mechanics simulation including modeling of the solvent.

In the case of five codebook vectors the Voronoi cells agree roughly with the five structural domains of the protein (Fig. 1 *b*). The resulting structure conforms to a certain degree to the shape and density distribution of the low-resolution map. The higher correlation coefficient of 0.76 indicates that the procedure produced a significantly better fit than rigid docking. Inspection of the overlap between the data sets, however, suggests that the fitting could be improved even further if the tip of domain G′ is oriented toward the arc-like extension of the density, and if domain III and the bridge of density between domains IV and III are moved to the corresponding features in the map (Fig. 1, *b*, *arrows*).

As described in Methods, the resolution of the fitting can be improved by increasing the number N of codebook vectors, but care must be taken if densities are not fully accounted for, i.e., if not all vectors converge toward equivalent features in the two data sets. The agreement between corresponding high- and low-resolution codebook vectors of EF-G worsened, as judged by their rms deviations 5.2 Å and 8.7 Å, when going from five to six vectors, respectively (rigid-body docking). Visual inspection, too, suggested that a docking with $N > 5$ vectors would produce unrealistic results. Nevertheless, it is clear from the fitting shown in Fig. 1 *b* that certain features of the data sets agree more than others. To take full advantage of the information in the data sets, given their experimental limitations, we have per-

FIGURE 1 Rigid-body and flexible docking of EF-G. (*a*) The completed structure of EF-G (tube representation), fitted into the EM density map (Agrawal et al., 1999) with the Situs package (Wriggers et al., 1998, 1999) using four codebook vectors (not shown). The colors code for the original crystal coordinates (*brown*) of EF-G in complex with GDP (Czworkowski et al., 1994), coordinates of domain III (*green*) extracted from the nucleotide-free structure (Ævarsson et al., 1994), and coordinates added by secondary structure prediction (*red*). The position of EF-G's structural domains is indicated. The arrow points to a gap between Thr-64 in the switch 1 loop (*red*) and the nearby beta phosphate of GDP. The EM density (*blue*) in Figs. 1 and 2 is contoured at 14% of the maximum value. The asterisk denotes a region of positive density that can be attributed to a conformational change in the ribosome (Agrawal et al., 1998, 1999). EF-G is rotated by 180° about the long axis relative to the conventional (Agrawal et al., 1998) orientation to visualize the GDP nucleotide (*purple*). (*b*) The flexibly fitted model of EF-G using five pairs of codebook vectors $\mathbf{w}_i^{l,h}$ (*gray spheres*) after the centroids of the Voronoi cells V_i^h (Eq. 2, shown in color) were forced to coincide with the $\mathbf{w}_i^l$ (see text). The arrows point to notable structural differences. (*c*) The final optimized protein structure (*brown tube representation*). Also represented by spheres are the positions of 10 codebook vectors $\mathbf{w}_i^l$ after the centroids of the Voronoi cells V_i^h (not shown) were forced to coincide with the $\mathbf{w}_i^l$. The volume of the spheres codes for the size of the Voronoi cells, and their color codes for the number N of codebook vectors in the corresponding vector quantizations: $N = 5$ (*green*), $N = 10$ (*red*), $N = 15$ (*cyan*), $N = 18$ (*yellow*), and $N = 24$ (*white*). Figs. 1 and 2 were created with Situs and with the molecular graphics program VMD (Humphrey et al., 1996).

formed an optimization based on 10 codebook vectors selected by visual inspection of five different vector quantizations, $5 \geq N \geq 24$. Fig. 1 *c* shows the result of this selective docking based on local similarity in the data sets (see Methods). The 10 vectors (and the corresponding number *N*) were chosen as follows: the major parts of domains I, II, and III were encoded by a single vector each. The bridge of density between domains IV and III was encoded by a single vector. The major parts of the domains IV and V (deviating in Fig. 1 *a*) were encoded by two vectors. Finally, the protruding tips of domains G′ and IV were each encoded by a single vector.

An improvement of the fit of all domains relative to the docking shown in Fig. 1, *a* and *b* is clearly discernible by eye (Fig. 1 *c*). The correlation coefficient of this fit is 0.77. Although the increase over the initial value (0.70) is small, the coefficient was not used as a criterion for the docking and it is reassuring that improved alignment (as judged by this optimization) results in an improvement of the correlation-based similarity measure. We note that this measure (unlike the rms deviation of atomic coordinates) is nonspecific in terms of structural correspondence. Earlier work has shown that the fit of the codebook vectors is a more useful parameter to compare multi-resolution data because correlations are relatively insensitive to changes in the superposition (Wriggers et al., 1999). The correlation coefficient values are therefore expected to be within a narrow numeric range.

Domain movements

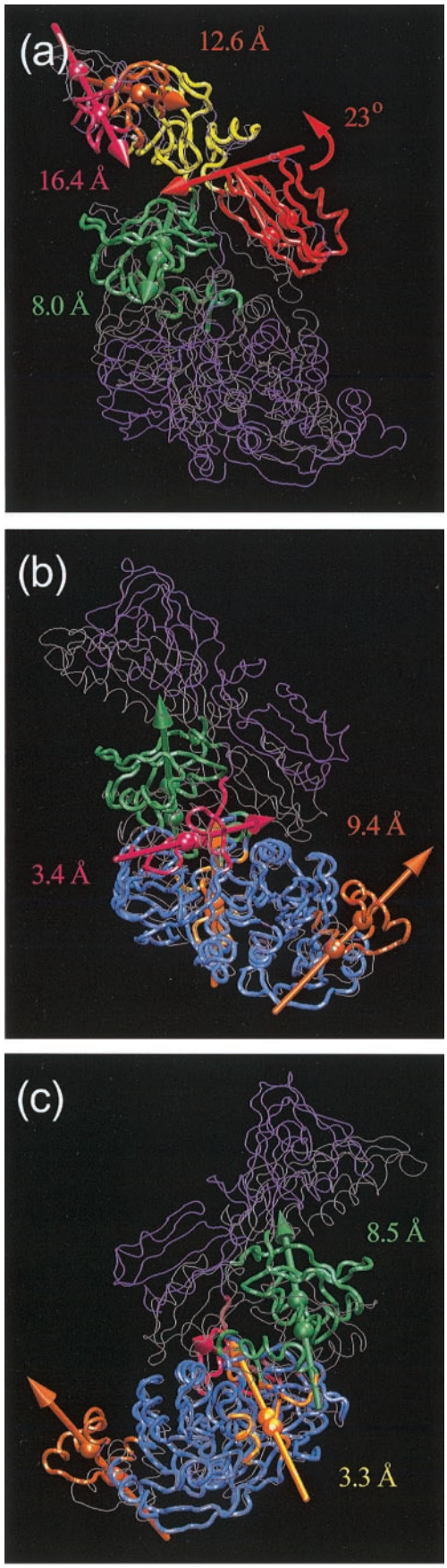
The resulting optimized structure of EF-G in the ribosome-bound form (Fig. 1 *c*) exhibits a C_α rms deviation of 6.9 Å from the completed crystal structure (Fig. 1 *a*). The large conformational change that can be attributed to ribosome binding is not evenly distributed; one can identify nine relatively rigid (C_α rms deviation 1.2–2.5 Å) segments that undergo hinge- and shear-type motions relative to one another. These rigid segments are connected by flexible linkers that enable the global adjustment of EF-G's structure to the low-resolution density.

Fig. 2 visualizes the essential movements of adjacent rigid segments. Essential rigid-body movements are characterized by the location of the effective rotation (hinge) axis if this axis intersects the protein. A hinge axis was found only in one case by this criterion. In all other cases the rotation of the segment is negligible compared to its COM translation. This type of movement typically involves COM displacements parallel to the interface of adjacent segments, as shown in Fig. 2, and can be classified as shear-type movement (Gerstein et al., 1994; Gerstein and Krebs, 1998).

In Fig. 2 *a* the two EF-G structures are superimposed by the largest found rigid region of domain IV. The two small segments at the tip of domain IV undergo relatively large displacements (>12 Å) according to the shape of the domain in the EM difference map that was the basis for the fitting. In the crystal structure, domains III, IV, and V appear as distinct structural subunits, mutually connected by one (III and IV) or two (IV and V) random coils. Based on the structures of EF-Tu in different states one would expect conserved intra-domain structures (Berchtold et al., 1993). This is indeed the case for domains III and V, but the motions within domain IV are significant. Due to the deviation from the crystal structure, details of the two small segments at the tip of domain IV in our model are uncertain and require further experimental investigation.

In contrast to the fitting of domain IV our results indicate that the internal structure of domains III and V remains close to the crystallographic conformation. A segmented, "domain" architecture is a prerequisite for proteins to undergo hinge motions (Gerstein et al., 1994; Gerstein and Krebs, 1998). It therefore comes as no surprise that the motion of domain V relative to domain IV can be described as a 23° rotation about the hinge axis shown that intersects the two connecting coils at residues 605 and 675 (Fig. 2 *a*). The two connections appear to create this axis by acting as mechanical hinges similar to the hinges that constrain the motion of a swinging door (Hayward, 1999). Domain III, in contrast, undergoes an 8.0 Å shear movement accommodated by the single coil following sheet 3_3, which acts as a flexible linker.

The largest rigid segment is formed by the core of domains I and II (Fig. 2, *b* and *c*). The tip of the adjacent domain G′ is displaced by 9.4 Å to form an arc-like connection with a protrusion at the base of the L7/L12 stalk of the ribosome (Fig. 1). The shear-type movement is accommodated by a flexible bending of helices $A_{G'}$ and $C_{G'}$ (Czworkowski et al., 1994) near residues 215 and 245. Such an elastic bending of α-helices is not uncommon and has also been observed, *e.g.*, in aspartate aminotransferase and glyceraldehyde-3-phosphate (McPhalen et al., 1992; Skarzynski and Wonacott, 1988). Due to the absence of packing constraints, the large, shear-type displacement of domain III by 8.5 Å requires only a small elastic response of the protein that appears to be mediated by the (modeled) effector region. Residues in the vicinity of the effector region appear as a separate preserved region and are seen to move to a lesser extent (3.3 Å). This layered architecture of the moving rigid segments is typical also for other proteins that undergo shear, such as citrate synthase and trp repressor (Lesk and Chothia, 1984; Lawson et al., 1988; Gerstein et al., 1994). The modeled loop that bridges domains II and III forms the sole covalent link between the relatively flexible domains III, IV, and V and the relatively conserved domains I and II. As shown in Fig. 2 *b*, the strain produced by the movements of the domains III, IV, and V is communicated via the loop to the exposed face of domain II. In particular, a segment comprising the $1_1/1_2$ sheet-loop-sheet and the loop following sheet 13_2 (Czworkowski et al., 1994) is pulled 3.4 Å toward the connecting loop.



Interaction with the ribosome

The precise positioning of EF-G's structural domains within the cryo-EM density enabled us to narrow down 17 stretches of amino acid residues of EF-G in close (~10 Å) proximity to the surface of the 15 Å resolution density map (Malhotra et al., 1998) of the ribosome (Fig. 3 *a*). These residues from all five EF-G domains are marked in Fig. 3 *b*. Here we focus on the effector region of EF-G, which is within the part undefined by x-ray crystallography that was constructed by modeling, and its interaction with the GTPase-associated center of the ribosome. According to the flexible docking of EF-G into the cryo-EM map, residues 56–60 of the reconstructed region are found in close contact with the ribosome envelope. However, the homology with the related factor EF-Tu suggests that residues 56–62, which correspond to the well-conserved residues 54–60 in the switch 1 region of EF-Tu (Berchtold et al., 1993; Ævarsson, 1995), should form EF-G's main interaction site with the ribosome. The excellent agreement between these independently derived locations cross-validates the results of fitting and homology modeling. Thus the fitting method proves to be extremely useful in locating the crucial portion of the 50S subunit of the ribosome that interacts with the effector region of the elongation factor. Also, it is reassuring that cryo-EM combined with modeling can provide results with an accuracy of two to three amino acid residues.

By making use of the data from hydroxyl radical probing (Wilson and Noller, 1998), and fits of EF-G's x-ray structure (Agrawal et al., 1998, 1999), the sites of contact between EF-G and two EF-G associated 23S rRNA fragments that were solved to atomic resolution could be identified: 1) the α-sarcin stem-loop (Correll et al., 1998), and 2) a

FIGURE 2 Major domain rearrangements in EF-G. Rigid regions (*colored*) in EF-G were identified with Hingefind (Wriggers and Schulten, 1997) by comparison of the optimized (Fig. 1 *c*) with the original (Fig. 1 *a*) conformation of the protein. The *partition* subroutine of Hingefind compares two known conformations of a protein and identifies connected regions that exhibit preserved packing within a specified tolerance of positional fluctuations. We extracted rigid elements by comparing EF-G's structure after refinement to the completed crystal structure at 4.0 Å tolerance in the position of α-carbons. (*a*) Back view. The structures are superimposed by a large segment of domain IV (63 aa, *yellow*). Visualized are the essential movements of domain III (91 aa, *green*), of domain V (60 aa, *red*), and of two small domain IV segments (27 aa, *magenta*; 22 aa, *brown*). (*b*) Back view, and (*c*) front view of the structures superimposed by the largest found rigid region (301 aa, *blue*). This rigid core is flanked by four regions that roughly correspond to the tip of domain G′ (29 aa, *brown*), domain III (91 aa, *green*), the vicinity of the modeled effector loop (21 aa, *orange*), and the vicinity of the modeled domain II/III connecting loop (28 aa, *magenta*). The remainder of the optimized protein structure is shown as a purple backbone trace. Also shown in gray is the original conformation of the protein. The essential rigid-body motions of neighboring regions are visualized either by their displacement axes or by their (right-handed) hinge rotation axis (including perpendicular tubes that connect the hinge axis to the segment centroids). The segment centroids in either conformation are shown as spheres in the color of the segment.

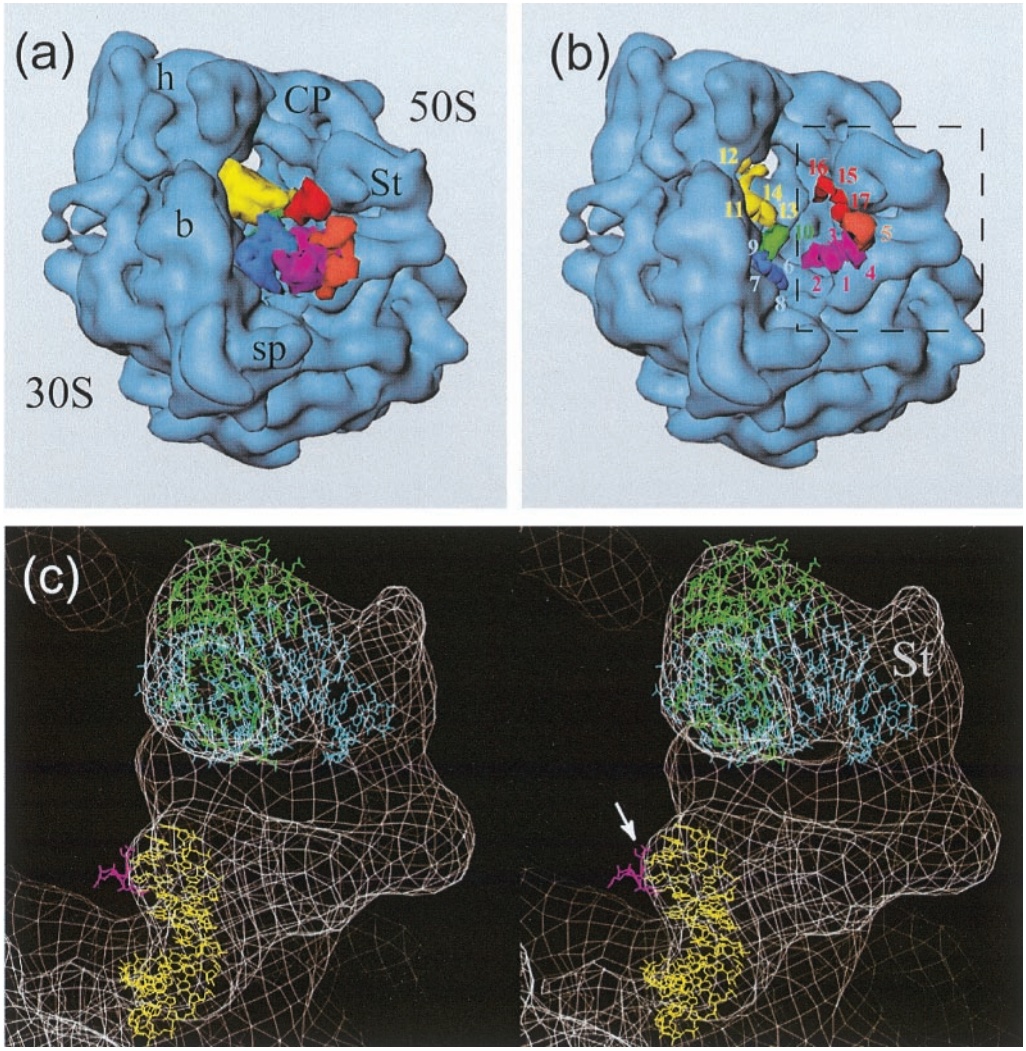


FIGURE 3 EF-G interactions with the ribosome (Malhotra et al., 1998). Landmarks of the 30S subunit: h, head; b, body; sp, spur. Landmarks of the 50S subunit: CP, central protuberance; St, L7/L12 stalk. (*a*) Flexibly docked EF-G (Fig. 1 *c*), domains shown in color: G (*magenta*), G′ (*brown*), II (*blue*), III (*green*), IV (*yellow*), V (*red*). (*b*) seventeen contact regions within 10 Å of the ribosome surface (*Thermus thermophilus* (Czworkowski et al., 1994) residue numbering): (1) 20–23, (2) 53–61, (3) 111–116, (4) 143–156, (5) 212–236, (6) 319–325, (7) 336–340, (8) 351–356, (9) 377–382, (10) 423–445, (11) 498–506, (12) 529–535, (13) 545–551, (14) 567–584, (15) 613–624, (16) 628–639, (17) 658–664. The stalk portion of the ribosome (*boxed in panel b*) is enlarged. (*c*) Stereo view of the interaction (*arrow*) of five crucial amino acids (res. 56–60, *magenta*) of the newly modeled effector region of EF-G with the stalk base of the ribosome (*wire mesh*). Shown are (see text): the α-sarcin stem loop (*yellow*), protein L11 (*green*), and a 58-nucleotide segment of 23S RNA (*blue*).

58-nucleotide long RNA fragment complexed with L11 protein (Conn et al., 1999; Wimberly et al., 1999). An earlier placement of the alpha-sarcin stem-loop (Agrawal et al., 2000) is identical to the placement in the 5 Å resolution x-ray map by Ban and co-workers (1999), which became available later. However, the placement of the structure of the L11-23S RNA complex (Wimberly et al., 1999) by Ban and co-workers, based on the RNA guided fitting in the x-ray map, is fully supported by the density in the cryo-EM map (see Agrawal et al., 2000; Gabashvili et al., 2000) and hydroxyl radical probing data (Wilson and Noller, 1998). These placements show that the amino acid residues 56–62 of the newly built stretch of the G domain of EF-G make contact with the tip of the alpha-sarcin stem loop (Fig. 3 *c*).

DISCUSSION

The number of degrees of freedom in an atomic resolution structure far exceeds the number of independent pieces of information in a low-resolution density map. Hence, there are multiple solutions to the problem of distributing a collection of atoms to match a low-resolution density. We note, however, that our fitting does not involve all atomic degrees of freedom. Rather, the method relies on additional infor-

mation and assumptions that make it feasible to obtain a realistic model of the docked protein. First, the molecular dynamics simulation method maintains the stereochemical quality of the atomic model during the optimization. Second, the minimization of the conformational energy seeks to preserve the initial crystal structure of the protein on the local level. This assumption is based on the observation that most protein conformational changes involve global motions of rigid domains that have their internal structure preserved (Gerstein et al., 1994; Gerstein and Krebs, 1998). Finally, the global refinement against the cryo-EM data is achieved by using only a small number of refinement parameters (the codebook vectors). One can count the number of independent pieces of information available by dividing the volume of EF-G by the volume of a resolution element $(16.5\ \text{Å})^3$. Clearly, the number of independent parameters used in the fitting should not exceed that number. EF-G measures $\sim$100 Å $\times$ 50 Å $\times$ 25 Å, comprising $\sim$36 resolution elements. In our docking, we have used 5–25 codebook vectors. This is below the number of resolution elements so we did not overfit the data.

The comparison of the detailed conformational changes with the control simulations (see Methods) and other studies (Wriggers et al., 1998) show that the fitting of the crystal structure defines the positions of amino acid residues with a precision about one order of magnitude above the nominal resolution of the cryo-EM data if local deviations from the initial crystal structure are small. Although the quality of the fitting is below the quality of crystal structures (where atom positions are typically visualized with sub-Ångstrom accuracy), we expect that our approach produces realistic models that can be useful for further computational or experimental studies. The detailed knowledge of the conformational changes provides functional insights that are not available by inspection or by manual fitting. Specifically, in the application to EF-G, the method predicts a functional role of the loops connecting the rigid segments. The II/III connecting loop is disordered in the crystal structure and in the cryo-EM map, *i.e.*, it may act as an entropic spring that communicates changes from the nucleotide-binding site to the flexible domains III, IV, and V. The functional role of this loop and of other linkers, such as the "door double-hinge" 605–675, can be probed by insertion/deletion mutagenesis experiments.

The novel flexible refinement method puts induced-fit conformational changes in large aggregates on a plausible quantitative footing. The application of multi-scale molecular dynamics enabled us to narrow down the contact points between EF-G and the ribosome, and specifically the important contacts in the GTPase-associated center. We note that the main advancement presented here is the direct, flexible fitting of EF-G to the 17 Å cryo-EM map. One can now begin to speculate how GTP hydrolysis on EF-G is allosterically coupled to ribosome binding and translocation. Comparison with the switch 1 regions in other G-proteins and motor proteins (Vale, 1996) suggests that the exchange of GDP by GTP induces a 3–5-Å shift of the effector contact point (residues 56–62) toward the nucleotide.

The fitting software developed will be distributed as part of the Situs package (Wriggers, in preparation; URL http://www.scripps.edu/mb/wriggers/situs/tutorial_flex.html).

W. W. acknowledges the LJIS Interdisciplinary Training Program and The Burroughs Wellcome Fund for fellowship support. D. L. D. acknowledges support by a Harold Urey fellowship and by the ARCS foundation.

This work was supported in part by grants from The National Institutes of Health and NSF (to J.A.M.), and National Institutes of Health Grants 1RO1GM55440 and 1R37GM29169 (to J.F.).

REFERENCES

Ævarsson, A. 1995. Structure-based sequence alignment of elongation factors Tu and G with related GTPases involved in translation. *J. Mol. Evol.* 41:1096–1104.

Ævarsson, A., E. Brazhnikov, M. Garber, J. Zheltonosova, Y. Chirgadze, S. al-Karadaghi, L. A. Svensson, and A. Liljas. 1994. Three-dimensional structure of the ribosomal translocase: elongation factor G from *Thermus thermophilus. EMBO J.* 13:3669–3677.

Agrawal, R. K., A. B. Heagle, and J. Frank. 2000. Studies of elongation factor G-dependent tRNA translocation by three-dimensional cryo-electron microscopy. *In* The Ribosome: Structure, Function, Antibiotics, and Cellular Interaction. R. A. Garrett, S. R. Donthwaite, A. Liljas, A. T. Matheson, P. B. Moore, and H. F. Noller, editors. ASM Press, Washington, DC. 53–62.

Agrawal, R. K., A. B. Heagle, P. Penczek, R. A. Grassucci, and J. Frank. 1999. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nat. Struct. Biol.* 6:643–647.

Agrawal, R. K., P. Penczek, R. A. Grassucci, and J. Frank. 1998. Visualization of elongation factor G on the *Escherichia coli* 70S ribosome: the mechanism of translocation. *Proc. Natl. Acad. Sci. USA.* 95:6134–6138.

Alberts, B. 1998. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell.* 92:291–294.

Baca, O. G., M. S. Rohrbach, and J. W. Bodley. 1976. Equilibrium measurements of the interactions of guanine nucleotides with *Escherichia coli* elongation factor G and the ribosome. *Biochemistry.* 15: 4570–4574.

Ban, N., P. Nissen, J. Hansen, M. Capel, P. B. Moore, and T. A. Steitz. 1999. Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature.* 400:841–847.

Berchtold, H., L. Reshetnikova, C. O. Reiser, N. K. Schirmer, M. Sprinzl, and R. Hilgenfeld. 1993. Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature.* 365:126–132.

Bourne, H. R., D. A. Sanders, and F. McCormick. 1990. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature.* 348:125–132.

Bourne, H. R., D. A. Sanders, and F. McCormick. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature.* 349:117–127.

Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. 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.

Cate, J. H., M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, and H. F. Noller. 1999. X-ray crystal structures of 70S ribosome functional complexes. *Science.* 285:2095–2104.

Clemons, Jr., W. M., J. L. C. May, B. T. Wimberly, J. P. McCutcheon, M. S. Capel, and V. Ramakrishnan. 1999. Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature.* 400:833–840.

Conn, G. L., D. E. Draper, E. E. Lattman, and A. G. Gittis. 1999. Crystal structure of a conserved ribosomal protein RNA complex. *Science.* 284:1171–1174.

Correll, C. C., A. Munishkin, Y. L. Chan, Z. Ren, I. G. Wool, and T. A. Steitz. 1998. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. USA.* 95: 13436–13441.

Czworkowski, J., and P. B. Moore. 1997. The conformational properties of elongation factor G and the mechanism of translocation. *Biochemistry.* 36:10327–10334.

Czworkowski, J., J. Wang, T. A. Steitz, and P. B. Moore. 1994. The crystal structure of elongation factor G complexed with GDP, at 2.7 Å resolution. *EMBO J.* 13:3661–3668.

DeRosier, D. J., and S. C. Harrison. 1997. Macromolecular assemblages: sizing things up. *Curr. Opinion Struct. Biol.* 7:237–238.

Frank, J. 1996. Three-Dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press, San Diego.

Gabashvili, I. S., R. K. Agrawal, C. M. T. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank, and P. Penczek. 2000. Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell.* 100:537–549.

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

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

Hayward, S. 1999. Structural principles governing domain motions in proteins. *Proteins: Struct., Funct., Genet.* 36:425–435.

Holley, L. H., and M. Karplus. 1989. Protein secondary structure prediction with a neural network. *Proc. Natl. Acad. Sci. USA.* 86:152–156.

Humphrey, W. F., A. Dalke, and K. Schulten. 1996. VMD—Visual Molecular Dynamics. *J. Mol. Graphics.* 14:33–38.

Johanson, U., A. Ævarsson, A. Liljas, and D. Hughes. 1996. The dynamic structure of EF-G studied by fusidic acid resistance and internal revertants. *J. Mol. Biol.* 258:420–432.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Kaziro, Y. 1978. The role of guanosine 5′-triphosphate in polypeptide chain elongation. *Biochim. Biophys. Acta.* 505:95–127.

Kohonen, T. 1995. Self-Organizing Maps. 2nd Ed. Springer-Verlag, Berlin.

Lawson, C. L., R. G. Zhang, R. W. Schevitz, Z. Otwinowski, A. Joachimiak, and P. B. Sigler. 1988. Flexibility of the DNA-binding domains of trp repressor. *Proteins: Struct., Funct., Genet.* 3:18–31.

Lesk, A. M., and C. Chothia. 1984. Mechanisms of domain closure in proteins. *J. Mol. Biol.* 174:175–191.

Makhoul, J., S. Roucos, and H. Gish. 1985. Vector quantization in speech coding. *Proc. IEEE.* 73:1551–1588.

Malhotra, A., P. Penczek, R. K. Agrawal, I. S. Gabashvili, R. A. Grassucci, R. Jünemann, N. Burkhardt, K. H. Nierhaus, and J. Frank. 1998. *Escherichia coli* 70 S ribosome at 15 Å resolution by cryo-electron microscopy: Localization of fMet-tRNA$_f^{met}$ and fitting of L1 protein. *J. Mol. Biol.* 280:103–116.

Martinetz, T. M., S. G. Berkovich, and K. Schulten. 1993. "Neural gas" for vector quantization and its application to time-series prediction. *IEEE Trans. Neural Networks.* 4:558–569.

McPhalen, C. A., M. G. Vincent, D. Picot, J. N. Jansonius, A. M. Lesk, and C. Chothia. 1992. Domain closure in mitochondrial aspartate aminotransferase. *J. Mol. Biol.* 227:197–213.

MSI. 1997. QUANTA 97. Molecular Simulations Inc., San Diego, California.

Skarzynski, T., and A. J. Wonacott. 1988. Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus. J. Mol. Biol.* 203:1097–1118.

Stowell, M. H. B., A. Miyazawa, and N. Unwin. 1998. Macromolecular structure determination by electron microscopy: new advances and recent results. *Curr. Opin. Struct. Biol.* 8:595–600.

Tocilj, A., F. Schlünzen, D. Janell, M. Glühmann, H. A. S. Hansen, J. Harms, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. 1999. The small ribosomal subunit from *Thermus thermophilus* at 4.5 Å resolution: pattern findings and the identification of a functional site. *Proc. Natl. Acad. Sci. USA.* 96:14252–14257.

Vale, R. D. 1996. Switches, latches, and amplifiers: common themes of G-proteins and molecular motors. *J. Cell Biol.* 135:291–302.

Willie, G. R., N. Richman, W. O. Godtfredsen, and J. W. Bodley. 1975. Some characteristics of and structural requirements for the interaction of 24,25-dihydrofusidic acid with ribosome-elongation factor G complexes. *Biochemistry.* 14:1713–1718.

Wilson, K. S., and H. F. Noller. 1998. Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. *Cell.* 92:131–139.

Wimberly, B. T., R. Guymon, J. P. McCutcheon, S. W. White, and V. Ramakrishnan. 1999. A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell.* 97:491–502.

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

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

Wriggers, W., and K. Schulten. 1997. Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. *Proteins: Struct., Funct., Genet.* 29:1–14.

Wriggers, W., and K. Schulten. 1998. Nucleotide-dependent movements of the kinesin motor domain predicted by simulated annealing. *Biophys. J.* 75:646–661.