Quasicontinuum Representations of Atomic-Scale Mechanics: From Proteins to Dislocations

Rob Phillips[1], Markus Dittrich[2], and Klaus Schulten[2]

[1]California Institute of Technology, Pasadena, California 91125 and [2]University of Illinois, Urbana, Illinois 61801; e-mail: Phillips@aero.caltech.edu

Key Words   proteins, coarse-graining, atomistic simulation

■ **Abstract**   Computation is one of the centerpieces of both the physical and bi-ological sciences. A key thrust in computational science is the explicit mechanistic simulation of the spatiotemporal evolution of materials ranging from macromolecules to intermetallic alloys. However, our ability to simulate such systems is in the end always limited in both the spatial extent of the systems that are considered, as well as the duration of the time that can be simulated. As a result, a variety of efforts have been put forth that aim to finesse these challenges in both space and time through new techniques in which constraint is exploited to reduce the overall computational burden. The aim of this review is to describe in general terms some of the key ideas that have been set forth in both the materials and biological setting and to speculate on future developments along these lines. We begin by developing general ideas on the exploitation of constraint as a systematic tool for degree of freedom thinning. These ideas are then applied to case studies ranging from the plastic deformation of solids to the interactions of proteins and DNA.

UNIVERSALITY, SPECIFICITY, AND THE ROLE OF COMPUTATION

The Tension Between Universality and Specificity

One of the key threads running through much of modern physics is the search for those features of the world around us that are universal. Indeed, the notion of universality is one of the central tenets of statistical physics and posits that in certain cases only a minimum of information about the system (i.e., the dimensionality of space and the order parameter) must be in hand in order to characterize that system. The study of phase transitions has exploited such universality at every turn. However, the notion of universality can be construed more broadly to reflect our ability to construct scaling descriptions of material response. This thread will be elaborated below, but as an example to set ideas we note that the scaling of

the radius of gyration of a polymer blob with the number of monomers can be thought of as a universal feature emerging from entropic elasticity. Similar results are commonplace in materials science as well.

By way of contrast, there are a number of settings within which we search for features of a system that are system specific. Such questions range from the relatively mundane issue of what makes one system elastically different from another to the full complexity of the structure-function relationship of certain proteins. As is described below, one important example of such specificity concerns the detailed mechanisms that give rise to dislocation nucleation in solids. From the biological setting, a second case study is that of DNA-protein interactions in which special sequences within the overall DNA sequence serve as ports for the docking of proteins. Part of the argument to be made in our review is that despite the difference in outlook between the search for universal and specific insights, multiscale methods have begun to serve as the basis of linking these viewpoints.

Universality and Specificity in Materials Science

The tension between universality and specificity transcends any particular field of analysis. In the context of materials science, there are a number of scaling laws that reflect the existence of universal features of materials. For example, the famed Hall-Petch laws posit a relation between the strength of a material, σ_y, and the mean grain size, d, of the form $\sigma_y \propto d^{-1/2}$. In this instance, the fundamental scaling structure is independent of material particulars. A second example arises in the consideration of diffusion for those cases in which a particular microscopic mechanism dominates. In this case, the robust result of interest is the assertion that the temperature dependence of the diffusion rate scales as

$$\text{diffusion rate} \propto \exp(-E_a/k_B T), \qquad\qquad 1.$$

where E_a is the activation energy for the process of interest and k_B is Boltzmann's constant. A final example is the existence of elasticity, with the key understanding embodied in the existence of an equation of the form

$$E_{strain} = \frac{1}{2}\int_\Omega C_{ijkl}\epsilon_{ij}(\mathbf{x})\epsilon_{kl}(\mathbf{x})d^3\mathbf{x}, \qquad\qquad 2.$$

which describes the energy stored in a strained solid. We have introduced the tensor $\epsilon_{ij}(\mathbf{x})$, which describes the strain at point $\mathbf{x}$ and the elastic modulus tensor C_{ijkl}. Although we are using the notion of universality in a sense that is more general than the standard nomenclature, it suits our purposes to note that certain features of materials transcend the particulars of any one material.

In fact, each of the examples cited above has a complementary feature that strikes to the heart of what we mean by specificity. For example, in the case of diffusion, it is well known that from one case to the next the overall rate will depend in a detailed way on the particulars of the diffusion pathway, a fact that is embodied in the activation energy, E_a. Similarly, although materials as diverse as rubber and

steel both exhibit elasticity, the particular characteristics of rubber and steel are widely different and are reflected in their quite different values for the elastic moduli, as implied by the elastic modulus tensor, C_{ijkl}. This type of specificity has been elegantly expressed in graphical form in the work of Ashby & Jones (1) who depict the huge diversity in material parameters for quantities such as the elastic moduli, the thermal conductivity, the diffusion constant, the yield strength, and fracture toughness. From the standpoint of computational materials science, one of the most compelling challenges is to first understand such specificity, with the ultimate goal being the more ambitious task of tailoring such specificity for particular applications.

Universality and Specificity in Biology

As noted in our description of materials, there are two levels of understanding, both of which are indispensable. On the one hand, the physics approach is often built around searching for those features of systems that are universal. For example, in the context of macromolecules, the way in which the radius of gyration depends upon the number of monomers would constitute a universal result. Similarly, the existence of allosteric reactions is a kind of universal insight. On the other hand, biological specificity is a well known precept and often requires the use of atomic-level arguments in order to determine the precise relation between structure and function. Indeed, both in the materials setting and in the biological setting this insight has been elevated to the status of a central dogma. One of our main arguments is that much of the work of explicating the relationship between structure and function must be carried out at the atomic scale.

One area in which the relationship between structure and function described above is especially evident is in the context of the emerging field of single-molecule biomechanics. The basic point is that as a result of the tremendous experimental advances that have followed on the heels of such tools as the atomic-force microscope and laser tweezers, it is now possible to pull on macromolecules in a controlled way. For examples, see the papers of Bustamante et al. (5), Essevaz-Roulet et al. (6), Marszalek et al. (11) and Strick et al. (20). One of the conclusions to emerge from this analysis is the existence of what could be called mechanical spectroscopy, in which each molecule exhibits its own unique force-displacement curve, with the detailed features of this curve emerging in turn from the underlying macromolecular structure. A key challenge at the level of molecular specificity is to determine the way in which structure dictates the specific details of such force-displacement curves.

An example of the type of specific understanding that can be gleaned in the macromolecular setting is that associated with the force-induced unfolding of immunoglobulin domains in titin. With the emergence of experimental methods for exerting force on single macromolecules, a new era in mechanical manipulation has been ushered in in which it is possible to measure the force-displacement curves of large molecules, as shown, for example, in Marszalek et al. (11). The quantitative assessment of these measurements requires a precise atomic-level description of the structural specifics of the molecule of interest. The direct numerical simulation

of this process in the case of titin has been undertaken in Lu et al. (9) and Lu and Schulten (10), in which it was found that the key unfolding event corresponds to the collective breaking of hydrogen bonds between different strands.

Degree of Freedom Thinning and Effective Theory Construction

The central thesis of this review is now spelled out as follows. First, we argue that in many circumstances, it is detailed, material-specific understanding that is the objective. In these cases, we are inevitably led to mechanistic investigations aimed at ferreting out the precise atomic or molecular events that correspond to the phenomenon of interest. We argue that despite the clear importance of such atomic-level investigations, it is necessary to construct a computational formalism in which only those atomic-level degrees of freedom that are really of interest are kept. We imagine the creation of a new suite of atomic-level simulation tools in which the tandem challenges of multiple scales in both space and time have been successfully hurdled. However, for our purposes here, we concentrate on the question of spatial degree of freedom thinning and refer the reader to the works of Voter (22) for a taste of some of the innovative thinking that is being done in the context of the time scale problem.

THE PHYSICS OF CONSTRAINT

The Geometry of Constraint

It is clear that the computational challenges of the type mentioned above will continue to stretch molecular simulations to their very limits. As a result, the key point of this section and the culmination of this article is the argument that we must continue to actively seek alternatives to brute force atomic simulation. All of the methods that one might wish to bring to bear on problems of the type described above involve in one way or another the notion of constraint. For example, if we are to invoke a continuum description, there is an inherent assumption that nearby molecular sites suffer identical deformations. From the perspective of atomic-level simulation in thinking about the spatial domain, the key idea involves in some form or another trying to eliminate some subset of degrees of freedom, keeping only those that are relevant. Another example is those cases in which density functional calculations are used to examine macromolecules such as enzymes (15). In these cases, only atoms in the vicinity of the active site are resolved explicitly and the remaining atoms are either frozen or ignored altogether. Similarly, in the temporal domain, the search for constraint involves fixing certain bonds in the hope that the resulting dynamics will respect the processes of real interest. A particularly provocative recent idea has involved the use of artificially high temperatures with an associated extrapolation scheme for backing out the nature of the diffusive processes that will occur at lower temperatures (19).

Thus far, we have argued that the traditional tools for carrying out atomic-level simulation in computational materials science and computational biology suffer from a variety of limitations including the inability to handle sufficiently large systems or long times. Indeed, such simulations are usually restricted to spatial scales much smaller than microns and times shorter than nanoseconds. One of the thrusts of this article centers on the idea of exploiting certain geometric regularities found in both crystals and certain macromolecules that might allow uniting atomic-level simulation and the machinery of the finite element method. The basic idea is to resolve the following question: Given a potential energy $\Pi(\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_N)$, which depends upon all of the atomic coordinates, how can we find a surrogate energy function that features only a subset (or a linear combination) of the full set of original degrees of freedom?

One example of this type is the quasicontinuum method in which the tools of atomistic simulation are united with those of structural and fluid mechanics (12, 16). The key kinematic idea behind this method is the systematic use of constraint such that a large fraction of the atomic-level degrees of freedom is relegated to a form of kinematic slavery in which the motions of a small subset of master atoms dictate the positions of all the others. These kinematic constraints then allow for a replacement of the full, total energy function, $\Pi(\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_N)$, which depends on the coordinates $\{\mathbf{R}_i\}$ of all N atoms with an approximate surrogate, $\Pi(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_M)$, which depends only upon the coordinates $\{\mathbf{X}_i\}$ of the M master atoms. From a computational point of view the key advantage arises from the fact that $M < N$ and usually, $M \ll N$.

In order to effect the type of degree of freedom reduction advocated above, we have invoked the existence of an underlying template. For example, in the context of crystalline solids, the atomic positions of the unaccounted for degrees of freedom are determined (if and when they are needed) by displacing them from their reference (template) positions according to the prescription

$$\mathbf{u}(\mathbf{X}) = \sum_i \mathbf{u}_i N_i(\mathbf{X}), \qquad\qquad 3.$$

where $\mathbf{u}(\mathbf{X})$ is the displacement of the atom at position $\mathbf{X}$, the $\mathbf{u}_i$'s are the displacements of the nodes surrounding the atom of interest and $N_i(\mathbf{X})$ is the shape function (see Figure 1) associated with the ith such node. The basic idea is that the displacement at a position within a given element is computed as a weighted average of the displacements on the three nodes surrounding the point of interest. As is revealed below, we propose that the same sort of geometric thinking can be brought to bear on macromolecules, where the underlying template is provided by the various structural motifs (e.g., proteins are built up of a sequence of amino acids, DNA is built up of a sequence of nucleotides) that make up such molecules.

Constrained Energy Minimization and Constrained Dynamics

Our characterization of the physical foundations of degree of freedom thinning has been to divide such analyses into two parts, the first of which is kinematic and
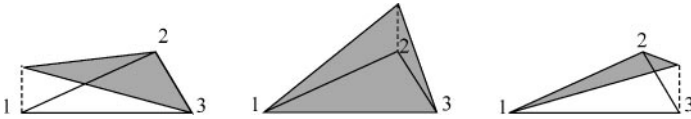
Figure 1   Illustration of the shape functions used in conjunction with two-dimensional finite elements. $N_1(\mathbf{X})$ is given in the leftmost schematic, $N_2(\mathbf{X})$ is shown in the middle figure and $N_3(\mathbf{X})$ is shown in the right figure.

was described above and the second of which considers the motions of the thinned degrees of freedom. This separation is to some extent artificial, although it suits the pedagogical purposes of the present discussion.

Given the set of reduced degrees of freedom $\{\mathbf{X}_i\}$ and the associated energy function $\Pi(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_M)$, our next question is how do we determine the energy minimizing configurations (statics) or the dynamical trajectories of the effective degrees of freedom. We begin by noting that a number of different variants of the question of how to find either the energy minimizing configurations or an effective dynamics for constrained kinematic representations have already been investigated. The most conceptually transparent scheme is that of zero-temperature energy minimization in which the effective energy function is relaxed with respect to the unknown nodal coordinates $\{\mathbf{X}_i\}$ using methods such as the conjugate gradient method or the Newton-Raphson method.

As the aim of this review is to highlight both successes and challenges, we note that in addition to the ability to carry out static energy minimization, quasicontinuum calculations have also been carried out dynamically (but in the absence of thermal effects) and using free energy minimization in which an effective free energy function associated with the master degrees of freedom is minimized (17). However, critical challenges remain surrounding the use of these methods as a full-fledged alternative to molecular dynamics because the dynamics of the master degrees of freedom at finite temperatures continues to pose challenges.

CASE STUDIES IN DEGREE OF FREEDOM THINNING

Case Studies from Conventional Materials Science

The development of systematic techniques for degree of freedom thinning is both a conceptual and computational necessity. Recent success in the conventional materials setting has been especially revealing in the context of defects in crystalline solids. Indeed, significant progress has been made in elucidating the structure, energetics, and interactions of dislocations, grain boundaries, and cracks. As a concrete incarnation of these methods, we describe the application of the quasicontinuum method to the nucleation dislocations.

NUCLEATION OF DISLOCATIONS   The general program of constructing first-principles models of plasticity in which plastic deformation is transmitted directly on the basis of dislocation motions remains one of the holy grails of computational materials science. Indeed, much progress has been made in this quest with the construction of a host of dislocation dynamics codes in which the dislocation population within a material responds to externally applied stresses. Nonetheless, one of the many difficulties with such analyses is the fact that they rely on a variety of mechanisms that are not within the direct purview of the linear elastic theory of dislocations. In particular, the treatment of dislocation nucleation, cross slip and dislocation interactions within dislocation dynamics codes is usually put in by hand in the form of various rules. The determination of the appropriate rules for these processes is one of the insights that can come from atomic-level analysis.

The conceptual underpinning for the development of dislocation dynamics has been the idea that rather than appealing to an uncertain phenomenology concerning plasticity, it would be more appealing to build up plasticity directly on the basis of the nucleation, motion, and interaction of dislocations. Unfortunately, just as there are uncertainties that attend the use of constitutive models of single-crystal plasticity, so too do models of dislocation dynamics call for insights external to the theory itself. One of the most damning uncertainties is that of the ways in which dislocations are nucleated at free surfaces, crack tips, and grain boundaries. An attractive experimental system within which such questions can be addressed is that of nanoindentation in which the surface of a material is deformed by a tip with a small ($\approx$60 nm) radius of curvature. As the force associated with this tip is increased, the deformation of the underlying material passes from innocent elastic deformation to permanent deformation in the form of dislocation nucleation.

An intriguing set of dislocation dynamics simulations of this phenomenon has been undertaken by Robertson & Fivel (14). An example of the type of experimental results it is the aim of these simulations to explore is shown in Figure 2.
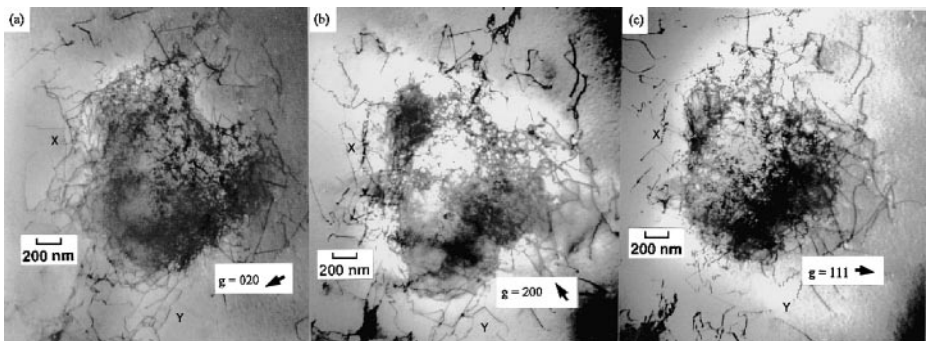

Figure 2   Experimental results revealing the nucleation of dislocations beneath an indenter (courtesy of M. Fivel & C. Robertson).
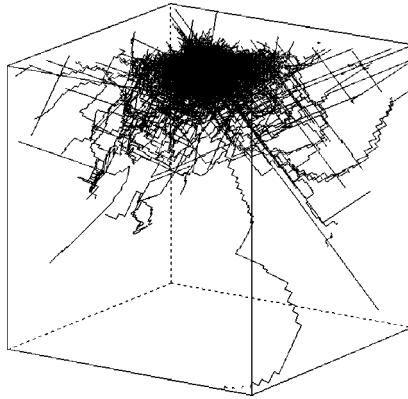
Figure 3 Simulation results revealing the nucleation of dislocations beneath an indenter (courtesy of M. Fivel & C. Robertson).

These results can be complemented using simulations of dislocation dynamics, as shown in Figure 3. These methods are used to examine the size and shape of the plastic zone beneath an indenter directly on the basis of dislocation dynamics. However, the dynamics is founded upon a series of ad hoc rules that determine when and where new dislocation loops are to be formed beneath the indenter.

What atomic-level analyses of the quasicontinuum type have to offer is the possibility of converting such ad hoc rules into well-formulated nucleation criteria. To that end, a series of calculations were performed by Tadmor et al. (21) and Shenoy et al. (18) in which a pseudo-two-dimensional indenter geometry was examined from the standpoint of the quasicontinuum method. In addition to examining the limits and validity of various analytic approaches to dislocation nucleation, these quasicontinuum results permitted the formulation of a nucleation criterion in which new dislocations are nucleated when the resolved shear stress on the slip plane admitting dislocation nucleation reaches a critical value. Two examples of the indented sample are given in Figure 4, which shows the way in which the plastic deformation which takes place beneath the indenter depends upon the relevant crystal orientation. The key insight to emerge from these simulations is a prescription for incorporating dislocation nucleation into higher level dislocation dynamics models directly on the basis of atomic-level understanding of the nucleation process.

Figure 4 Application of the quasicontinuum method to the nucleation of dislocations beneath an indenter. This figure reveals the type of dislocation geometries induced by the presence of the indenter, which is represented by the white rectangle on the crystal surface (from Reference 21).
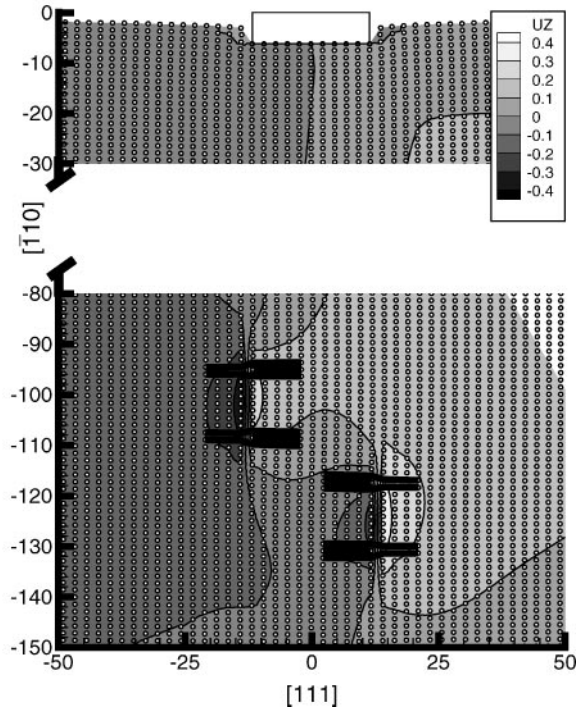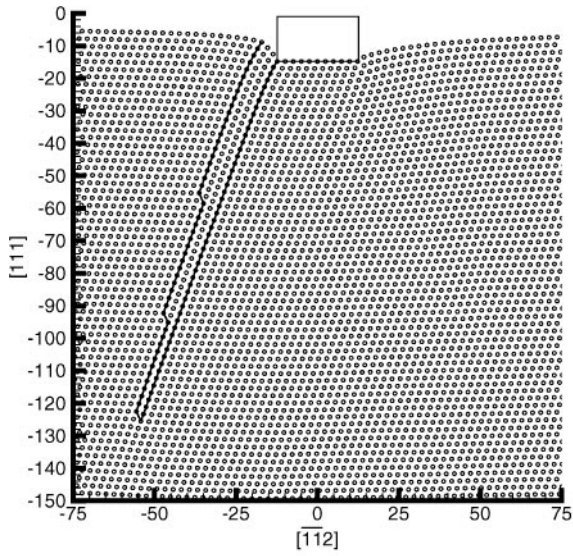
Case Studies from Computational Biology

Although models like those described above have already revealed their usefulness in the context of traditional materials, the most intriguing possibilities for such methods lie in the biological setting. Several examples of the application of multiscale analysis to problems of biological significance are discussed below.

MECHANICS OF DNA-PROTEIN INTERACTIONS   Coarse-graining methods can clearly have an impact in the context of DNA-protein interactions. The importance of such interactions as a central element in genetic control is spelled out eloquently by Ptashne (13). For example, the regulation of gene expression associated with the enzyme used to digest lactose is carried out by a protein known as the lac repressor. The lac repressor binds to two spatially distinct sites on the DNA molecule thus inducing a loop in the region between these two sites and excluding this part of the DNA from further transcriptional action. An elegant set of experiments aimed at examining the kinetics of loop formation under the action of lac repressor was carried out by Finzi & Gelles (7) who attached one end of a DNA molecule to a glass slide and had the other end tagged by an optical bead. The fluctuations of the tethered DNA as a result of Brownian motion were monitored with the insight that when a loop had formed, the length of the tethered molecule would be effectively shortened resulting in a different dynamics than in the absence of looping.

Despite the impressive progress in deducing both the existence and action of the lac repressor, as well as the resulting DNA dynamics, a wide variety of mechanistic questions remain concerning this system that range from the precise structure of the DNA-protein binding complex, to the energetics of this binding, to the kinetic processes that govern the life history of the lac repressor. All of these problems call for atomic-level understanding. On the other hand, the computational demands that attend attacking these problems are exorbitant and have led to the development of mixed atomistic-continuum formulations much like those discussed above in the context of defects in solids. One way in which degree of freedom thinning can be brought to bear on this problem is to examine the DNA-protein complex without having to pay the full atomistic price for the treatment of the looped DNA region. To that end, Balaeff et al. (2) have constructed an elastic rod model of the looped DNA that is solved for boundary conditions of the DNA-protein complex. The representation of the looped DNA is shown in Figure 5, with Figure 5a illustrating the kinematic description of the DNA in terms of a local set of vectors $\mathbf{d}_1(s)$, $\mathbf{d}_2(s)$ and $\mathbf{d}_3(s)$, where s parametrizes the position along the DNA molecule; Figure 5b is a schematic of the way in which the elastic solution can be fed back to the full atomistic simulation, adding loop-induced forces to the latter. In particular, the idea is that the loop above the actual simulation box is present only in terms of the forces it imposes on the parts of the molecules that are subject to direct atomistic investigation. Although our discussion does not spell out the details of how these calculations were done, our main intent is to show that the results of such calculations can be used as a boundary condition for the purposes of traditional atomic-scale analysis.

A CALCULUS OF STRUCTURAL MOTIFS   The central theme of this article is the concept of a systematic elimination of degrees of freedom on the basis of constraint. It provides a powerful tool not only for the reduction of the computational burden but also as a measure to help judge which features are particularly important in determining a macromolecule's biological function. One venue within which we believe that such techniques will be especially useful is in the construction of mesoscopic models of biophysical significance. Herein we provide an outline of a strategy to devise coarse-grained models of biopolymers and biopolymeric complexes in their natural environments. We also present preliminary progress in our recent efforts to build a coarse-grained description for proteins.

Almost all approaches to simulate protein structure and protein dynamics, using e.g., molecular dynamics (MD) simulation techniques, tackle the problem from an all-atom point of view. However, consideration of a generic enzyme reveals that often there is an underlying separation of spatial scales. To make this idea more concrete consider an enzyme embedded in a membrane and performing hydrolysis of some compound in a living cell. Clearly, the active site where the chemical transformation of the substrate takes place may require a fully atomistic description. Upon receding from this site, however, one often encounters regions in the protein that solely serve a structural role, constituting a polypeptide framework to hold the active site in its correct shape. The same kind of arguments can be applied to the lipid bilayer membrane surrounding the protein. In order to achieve realistic simulations of membrane proteins, it is necessary to include membrane patches of considerable size as part of the MD simulations, often more than doubling the number of atoms to be considered. The membrane's purpose, on the other hand, is mostly that of a scaffold for the protein core, providing the necessary hydrophobic environment for intra-membrane parts of the enzyme. Treating the membrane in an all-atom fashion is probably overkill, and an approach similar to the one used for dislocations in solids seems far more appropriate; namely a more refined treatment of lipid molecules close to the protein core and increased coarse graining using a finite element interpolation scheme in regions farther away. The underlying theme for both the protein and its surroundings is the use of adaption, meaning that the description should be as coarse-grained as necessary to retain and capture the functionally relevant pieces, with superfluous degrees of freedom discarded.

The remainder of the discussion focuses on the description of proteins. Adopting a bottom-up point of view when considering the protein structural hierarchy, one first encounters the proteins' building blocks, namely, the 20 naturally occuring amino acids arranged into polypeptide chains of different length. The structural hallmark of amino acids is their identical backbone, their sole distinctive feature being the 20 different amino acid side chains. Advancing one step up from the linear arrangement of amino acids on the primary structure level, is the realm of secondary structure. Here a wealth of structural motifs is encountered, and it is a well-established fact that most proteins can be pictured as a collection of such structural units (4); the most prominent ones are the α-helix and the β-sheet. It is possible to judge a protein's function or even its location in the cellular

environment by its composition in terms of secondary structural elements, for example, the seven α-helix bundle typical for membrane embedded G protein–coupled receptors. The importance of structural motifs is also apparent in the fact that whole protein databanks such as CATH or SCOP are organized around the composition of proteins in terms of secondary structure elements. Based upon this, we aim to establish a coarse-grained description on both the single amino acid as well as the secondary structure level, thereby reducing the number of degrees of freedom by a factor of at least ten.

In the remainder of this section, we describe our ideas for a constrained representation of amino acids. To illustrate such a representation consider an amino acid at position i in a polypeptide chain. We focus our attention on the backbone consisting of an NH-group, a carbon C_α and a CO-group. The C_α atom is connected to a side chain of one of the 20 amino acids in proteins. The carbon atoms involved are labeled C_α, C_β, etc. The scheme of constraint that we propose rests on the notion of master atoms, which are being kept track of explicitly, and slave atoms, which are constrained to follow the master atoms in their motion. To keep track of the spatial location of residue i, we define by $\mathbf{R}^i_{C_\alpha}$ the position of the C_α atom. The orientation of the peptide bond and the side chain are described by a fixed set of vectors that connect C_α to the hydrogen H_N ($\mathbf{R}^i_H$), to the oxygen ($\mathbf{R}^i_O$) and to C_β ($\mathbf{R}^i_{C_\beta}$). The vectors $\mathbf{R}^i_H$, $\mathbf{R}^i_O$ and $\mathbf{R}^i_{C_\beta}$ are used to define the axes of an affine coordinate system centered on the C_α; N, C and H_α are slave atoms the positions of which are tied to $\mathbf{R}^i_{C_\alpha}$ as are the side chain atoms. The position vector $\mathbf{r}^i_S$ of a slave atoms S can then be obtained through the relation

$$\mathbf{r}^i_S = \mathbf{R}^i_{C_\alpha} + \alpha^i_S \Delta \mathbf{R}^i_H + \beta^i_S \Delta \mathbf{R}^i_O + \gamma^i_S \Delta \mathbf{R}^i_{C_\beta}, \qquad\qquad 4.$$

where $\Delta \mathbf{R}^i_\gamma = (\mathbf{R}^i_\gamma - \mathbf{R}^i_{C_\alpha})$ are constant vectors and the quantities α^i_S, β^i_S and γ^i_S are defined below.

Three of the 20 amino acids require a slightly modified description: In the case of glycine we invoke H_β instead of C_β; in the case of cysteine we keep track of sulfur S_γ to model disulfide bridges as closely as possible, and in the case of proline, we keep track of the ring hydrogen H_γ rather than C_β. For the case of alanine the representation outlined is shown in Figure 6. We also note that our eventual objective is to decide on the level of refinement used in our constrained calculations in an adaptive fashion, with the level of resolution changing during the course of a simulation.

According to Equation 4, the positions of the slave atoms are defined through their coordinates α^i_S, β^i_S and γ^i_S. These coordinates are chosen according to the initial protein structure but subsequently are independent of protein conformational changes. Naturally one has to address the question of how to best choose the coordinates. One can either select the coordinates to reproduce the initial PDB structure (3) or choose them to reflect optimally average protein conformations. The description as outlined in terms of master and slave atoms provides considerable motional freedom to the backbone itself but limits the movement of side

chains. Several remedies to provide more flexibility to side chains may be employed, e.g., separate affine coordinate systems for both backbone and side chains, both centered on the C_α. One might also envision a description of the backbone as explained but represent side chains through spherical or ellipsoidal shapes that are tracked through their centers and orientation.

The next crucial conceptual issue is the calculation of effective forces on the master atoms. The potential energy of the protein as described in the constrained scheme can be expressed as $\Pi(\{\mathbf{r}_S^i(\{\mathbf{R}_\alpha^i\})\}, \{\mathbf{R}_\beta^j\})$ where $\mathbf{r}_S^i(\{\mathbf{R}_\alpha^i\})$ denotes the position of the slave atoms in residue i in terms of the coordinates of the master atoms. The forces on the master atoms can be calculated via the chain rule

$$\mathbf{F}_\alpha^i = -\frac{\partial \Pi}{\partial \mathbf{R}_\alpha^i} - \sum_j \frac{\partial \Pi}{\partial \mathbf{r}_S^j} \frac{\partial \mathbf{r}_S^j}{\partial \mathbf{R}_\alpha^i}. \qquad 5.$$

To compute these forces requires knowledge of the forces on all atoms as evidenced by the second term on the right-hand side of the equation and, as a result, the present scheme is crippled if one cannot find a way to find effective forces on the master atoms without visiting every atom. Devising a scheme that requires only force calculation for master atoms is of utmost importance. We have already implemented the constrained kinematic representation outlined above and used Equation 5 to determine the energy-minimized protein structures as a diagnostic relative to the goodness of approximating protein structure on the basis of subsets of the full set of atomic degrees of freedom. The current implementation of this method is built around the suite of molecular modeling tools, known as TINKER, from the Ponder group at Washington University. Calculations were performed on oligopeptide chains as well as on bovine pancreatic trypsin inhibitor (BPTI). The structures agree well with with all-atom calculations in the interior of the protein with somewhat larger deviations for longer side chains at the periphery. The calculated root mean square deviation is 1.17 Å. This deviation results mainly from to a total of 16 residues having a root mean square deviation larger than 1.0 Å. With the exception of PHE4 and LEU29, all are charged or polar, and most are oriented toward the outside of the protein. This is illustrated in Figure 7. Further analysis shows that the root mean square deviation of 4 out of the 16 residues (ASP3, ARG39, ARG42, and GLU49) is due to a conformational change of the side chains, namely folding back of their charged groups onto the main chain. The root mean square deviation for the remaining 11 residues is caused by a rigid shift or a rigid rotation of the whole amino acid. The potential energies obtained from our calculations were generally found to be too high, the major deviation being due to the 4 charged residues that change their side chain conformation. Presently we seek to implement a dynamical implementation of the constrained representation to evaluate its performance in the dynamic setting. We will also devise and test routes toward calculating effective forces on nodes without the need to know the forces on every atom explicitly.

A VISION FOR COMPUTATIONAL SCIENCE: COMPUTATIONAL EFFECTIVE THEORIES

We have argued that universality and specificity pose complementary demands in the development of understanding in both the materials science and biological settings. One of the key tools in effecting the analysis of specificity is atomic-level simulation. Unfortunately, such simulations carry with them a huge computational burden that makes the simulation of large systems and long times prohibitive. As a result, one of the central challenges of computational science is the need to find computational surrogates in which the choice of degrees of freedom is made such that full atomic-scale resolution is used only where it is needed. Such ideas have already shown their worth in the setting of traditional materials, and we contend that a central challenge for the consideration of biological materials is the successful development of tools for systematic coarse graining.

ACKNOWLEDGMENTS

**The *Annual Review of Materials Research* is online at
http://matsci.annualreviews.org**

LITERATURE CITED

1. Ashby MF, Jones DRH. 1996. *Engineering Materials*. Oxford, UK: Butterworth-Heinemann

2. Balaeff A, Mahadevan L, Schulten K. 1999. *Phys. Rev. Lett.* 83:4900

3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. *Nucleic Acids Res.* 28:235

3a. Bhandarkar M, Budescu G, Humphrey WF, Izaguirre JA, Izrailev S, et al. 1999. BioCoRE: a collaboratory for structural biology. Proc. SCS Int. Conf. Web-Based Modeling and Simulation, San Francisco. pp. 242–51

4. Branden C, Tooze J. 1999. *Introduction to Protein Structure*. New York: Garland. 2nd ed.

5. Bustamante C, Macosko JC, Wuite GJL. 2000. *Nat. Rev. Mol. Cell Biol.* 1:130

6. Essevaz-Roulet B, Bockelmann U, Heslot F. 1997. *Proc. Natl. Acad. Sci. USA* 94:11935

7. Finzi L, Gelles J. 1995. *Science* 267:378

8. Humphrey W, Dalke A, Schulten K. 1996. *J. Mol. Graph.* 14:33

9. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. 1998. *Biophys. J.* 75:662

10. Lu H, Schulten K. 2000. *Biophys. J.* 79:51

11. Marszalek PE, Lu H, Li H, Carrion-Vazquez M, Oberhauser AF, Schulten K, Fernandez JM. 1999. *Nature* 402:100

12. Ortiz M, Phillips R. 1998. *Adv. Appl. Mech.* 36:1–79

13. Ptashne M. 1992. *A Genetic Switch*. Cambridge, MA: Blackwell Sci. Cell

14. Robertson CF, Fivel MC. 1999. *J. Mater. Res.* 14:2251

15. Segall MD, Payne MC, Ellis SW, Tucker GT, Boyes RN. 1998. *Phys. Rev. E* 57: 4618.

16. Shenoy VB, Miller R, Tadmor EB, Rodney D, Phillips R, Ortiz M. 1999. *J. Mech. Phys. Solids* 47:611

17. Shenoy VB, Shenoy VB, Phillips R. 1999. In *Multiscale Modeling of Materials*. ed. VV Bulatov, TD de la Rubia, R Phillips, E Kaxiras, N Ghoniem. Warrendale, PA: Mater. Res. Soc.

18. Shenoy VB, Phillips R, Tadmor EB. 2000. *J. Mech. Phys. Solids* 48:649

19. Sorensen MR, Voter AF. 2000. *J. Chem. Phys.* 112:9599

20. Strick TR, Allemand JF, Bensimon D, Croquette V. 2000. *Annu. Rev. Biophys. Biomol. Struct.* 29:523

21. Tadmor EB, Miller R, Phillips R, Ortiz M. 1999. *J. Mater. Res.* 14:2233

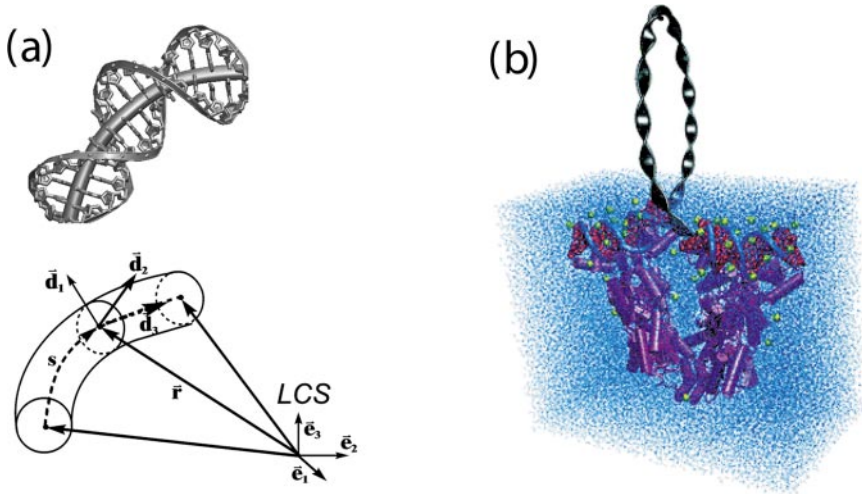22. Voter AF. 1997. *J. Chem. Phys.* 106:4665

Figure 5 DNA-protein interaction as described in a mixed atomistic-continuum framework. (*a*) Continuum elastic rod model of DNA. Shown is (*top*) a molecular model of DNA with its central axis depicted as a curved elastic rod and (*bottom*) the coordinate systems used to calculate the minimum energy conformation of the rod. (*b*) A twisted DNA loop (*top*) as described by elastic rod theory held by the lac repressor (*bottom*), the latter being simulated in a bath of water molecules and ions by a full atom molecular dynamics simulation that takes into account forces, computed by elastic rod theory, that resist the looping of the DNA.
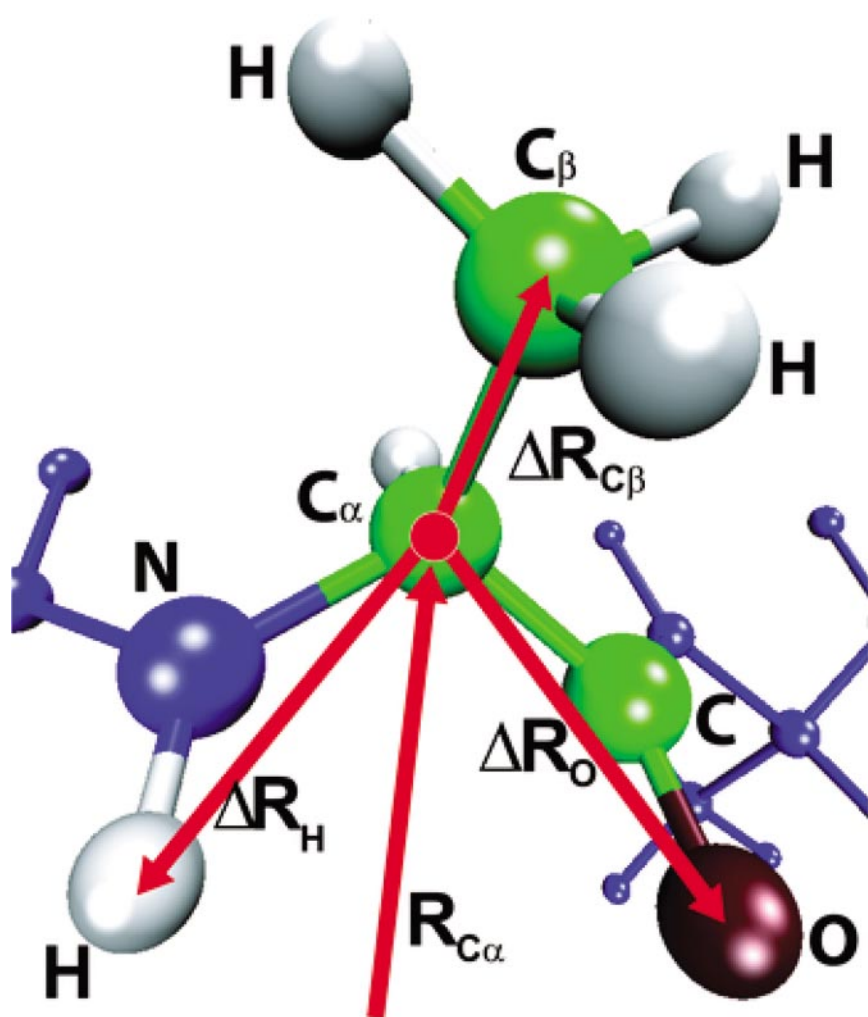
Figure 6 Illustration of the constrained representation for amino acids. This figure shows an alanine residue together with the affine basis vectors centered on the C_α atom. All slave atoms are represented as linear combinations of the basis vectors in terms of their coordinates α_S^i, β_S^i, and γ_S^i according to Equation 4.
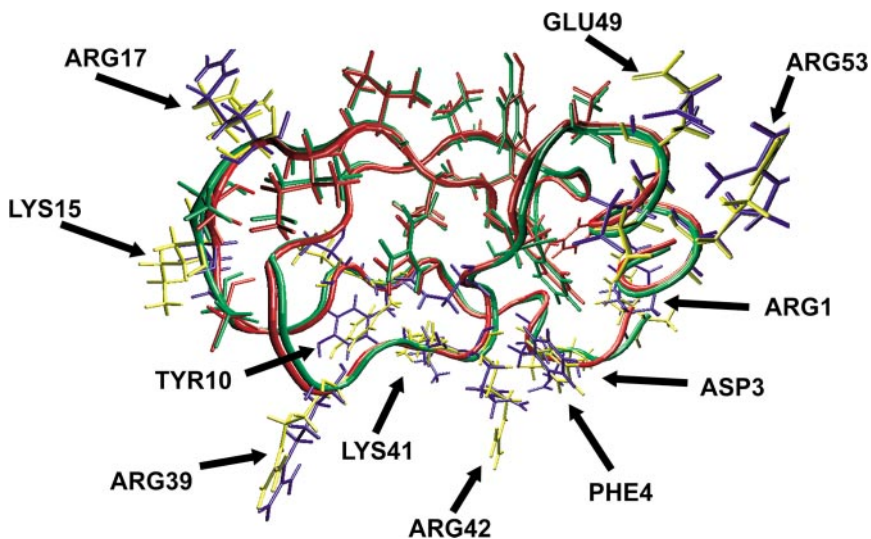
Figure 7 Comparison of the relaxed structures obtained using all atom calculations (*red*) and the constrained scheme (*green*). The backbone is shown in tube representation together with a subset of the side chains. Ten side chains with a root mean square deviation larger than 1.0 Å are shown in *blue* (all atom calculation) and *yellow* (constrained scheme). With the exception of PHE4, all are charged or polar and are located at the periphery of the protein.