Lattice Representations of Globular Proteins: How Good Are They?

Adam Godzik, Andrzej Kolinski, and Jeffrey Skolnick*

Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Received 16 December 1992; accepted 30 April 1993

Using a number of different lattice models of proteins, the problems introduced by the discretization of a protein backbone are discussed and examples of the most typical errors arising in low coordination number lattices presented. The geometric properties of different lattices used in the literature are compiled, and for all of them the resulting α-carbon models of proteins are described in detail and compared to the original structures obtained from experiment. © 1993 by John Wiley & Sons, Inc.

INTRODUCTION

Recently, simplified models of protein structure have become increasingly popular.[1] It is now clear that methods employing a detailed, all-atom description of protein structure are not able to describe long-time processes such as the folding of even small single-domain proteins.[2] One natural way to reach longer time scales is to reduce the number of degrees of freedom that are explicitly treated. This can be achieved by ignoring the presence of all backbone atoms except the α-carbons, ignoring the side chains entirely, or replacing them by entities with a smaller number of degrees of freedom, such as the side-chain center of mass.[3-7] Further, interactions between pieces of the resulting reduced representation of the chain are modeled by potentials of mean force; thus, the solvent is implicitly rather than explicitly included.[7-11] The most drastic of these simplifications occurs in lattice models of proteins, where in addition to simplifying the description of the backbone and the amino acid side chains their positions are restricted to a discrete set of possibilities, i.e., the vertices of a given specified lattice.[12-22] Such a description offers many advantages, which together result in a substantially increased speed of calculation over more conventional techniques. Discretization reduces the available conformational space and, at the same time, can smooth the free energy surface, thereby eliminating a number of local hills and valleys present in an off-lattice representation. Most calculations can be carried out using integer arithmetics, and it is possible to precalculate many time-consuming operations. Overall, over a 100-fold speedup, as compared to the traditional molecular dynamics simulations, on equivalent reduced models can be obtained. Thus, routine simulations of long-time processes, such as protein folding, become feasible.[22]

On the other hand, protein structures in a lattice representation are obviously distorted in comparison to the real structures, and it is not clear what level of similarity must be achieved for the model to be useful. Therefore, in the present contribution we analyze and compare several different lattice models of proteins to understand the nature of the differences between discrete and continuous descriptions of protein structure and evaluate various levels of accuracy obtained in the different types of lattice models.

The first argument in favor of using a discrete model of protein structure comes from the regularities observed in the local structure of the protein backbone (Fig. 1). The distances between consecutive Cα—Cα atoms are sharply peaked around 3.8 Å (with a small peak around 2.95 Å for proline residues). The angle between three consecutive Cα atoms has a sharp maximum peaked at 90° and a more diffuse one at around 120°, and the tetrahedral angle between four Cα-atoms has one sharp peak around $-130°$ and one broad peak centered at 20° with a secondary peak around 125°. The peaks correspond roughly to the threefold rotational symmetry around the central Cα—Cα bond[23] (see also Fig. 1). These values result from the specific geometry of the peptide bond and strongly suggest that a discrete representation with the correct set of symmetries should be able to closely mimic the geometry of the protein backbone.

Thus far, we have focused on the intrinsic geometric properties of various lattices that have been used for the discretized representation of protein chains. To go further, we must develop criteria to

*Author to whom all correspondence should be addressed.

CCC 0192-8651/93/101194-09

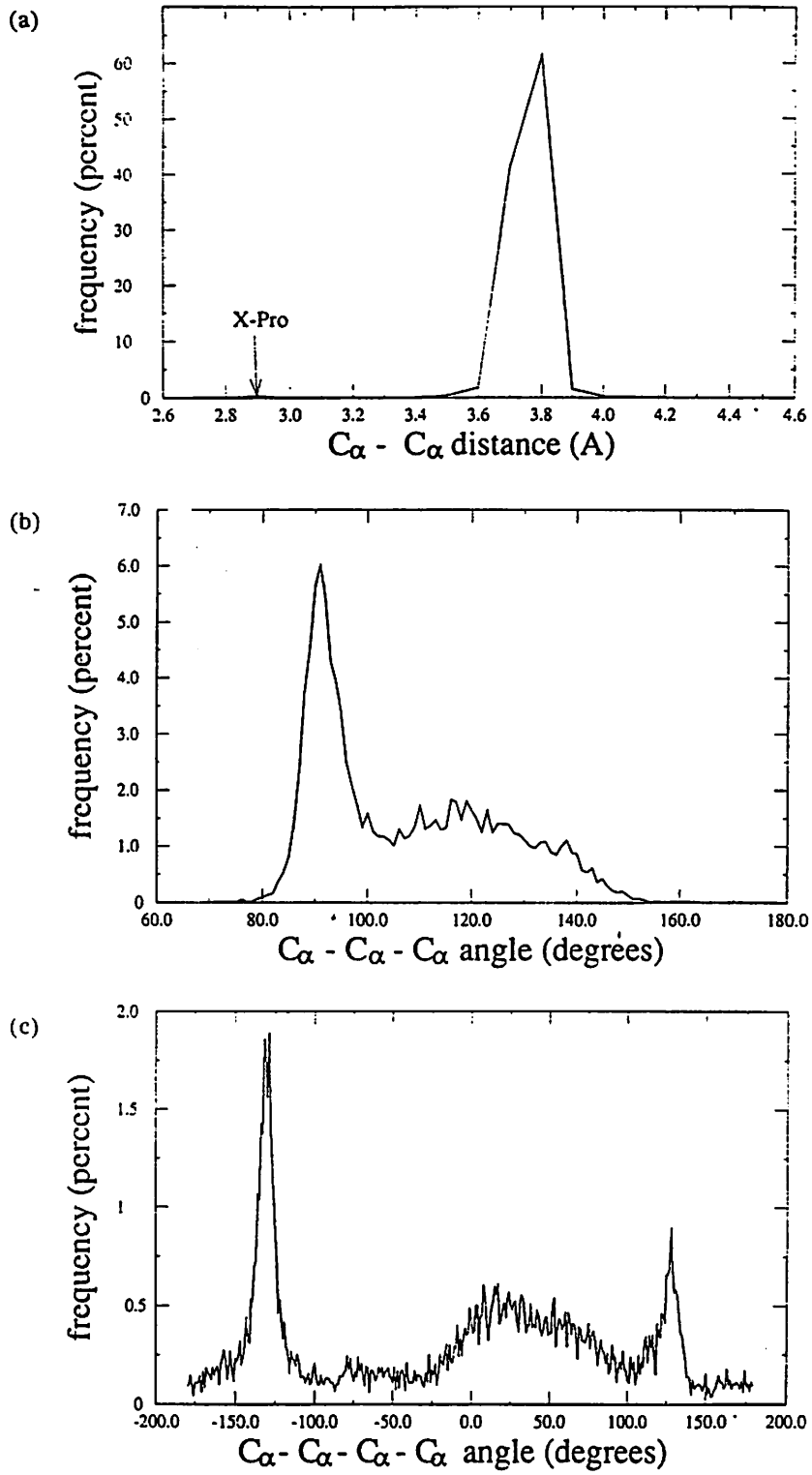

Figure 1. Local regularities in protein structures. (a). Distribution of Cα—Cα distances. (b). Distribution of Cα—Cα—Cα angles. (c). Distribution of Cα—Cα—Cα—Cα torsional angles.

compare and evaluate the similarity of the different lattice models of proteins relative to their respective crystallographic structures. The most commonly used measure of structural similarity is the rms distance between equivalent Cα positions after optimal superposition. Here, because the real structure and its model are compared, the identification of equivalent residues is trivial and the problem has a well-defined solution. Another criterion that will be used throughout this article is based on the local orientation of the chain, as measured by the direction of the bisector of the $C\alpha_{n-1}$-$C\alpha_n$-$C\alpha_{n-1}$ angle (Fig. 2).

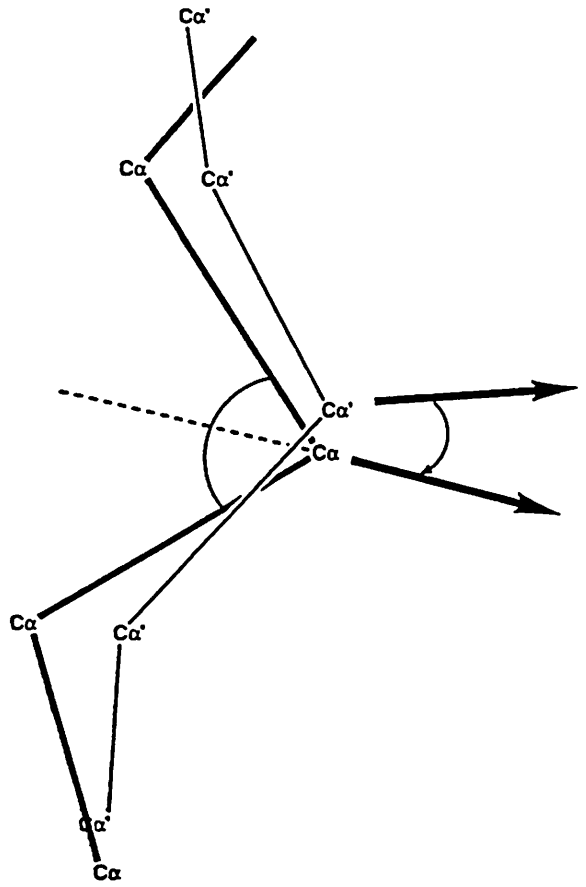

Figure 2. Angular similarity between two structures. Definition of the bisecting vector.

The local or angular correlation of the model and the real protein is crucial both in calculating the energy for simplified protein models as well as in rebuilding the full atom structure from the simplified description. Both problems will be discussed in detail in subsequent publications.

Note that there are several features of lattice models of protein structure that make them different from continuous representations of proteins. These are outlined below.

1. The lattice representation depends very much on the orientation of the principal axis of the lattice. For instance, an element of the secondary structure might be described well if its axis coincides with one of the symmetry axes of the lattice; otherwise, it might be poorly described. As a result, certain supersecondary structure elements, such as supertwisted helical hairpins, and most of the larger proteins, where different secondary structure elements can appear with several different mutual orientations, cannot be fitted to some lattices with results as good as certain small molecules that are often presented as examples.[15,18]
2. Despite the apparent similarities between the local regularities of the protein backbone and the lattice chain, small differences can accumulate quickly and cannot be corrected due to the inflexibility of the lattice. This may result in a lattice structure that follows the backbone trace, but its local structure might be essentially random.
3. Regularities of the protein backbone, such as illustrated in Figure 1, are dominated by the statistics coming from the secondary structure elements. Other fragments of the protein structure, such as some types of turns, may have different bond lengths and angles that cannot be reproduced on the lattice. Due to this problem, the lattice structures may have uneven quality, with overall good rms, but at the same time it is grossly incorrect in some regions.

These peculiarities of the discretized description of proteins will be studied in detail for a number of lattices.

METHODS

We begin by defining a consistent way of describing any lattice model. In a cubic lattice, each point is described by a set of three integer indices, giving its distance in lattice units along the three principal Cartesian axes (x,y,z) from an arbitrarily chosen origin. The smallest distance between two points along any principal axis is taken to be unity. Now, every ordered set of points (which would usually be called a chain) can be described either by listing all its points or the starting point and then the sequence of vectors between the consecutive points of the chain. We can restrict ourselves to sets of points that are joined by vectors fulfilling certain criteria. Among the models presented here, sets of vectors (called basis vectors) used to build the chain could contain as few as 6 vectors or as many as 90. Different lattice models differ by the sets of basis vectors, and when we say "different lattices" this really means "lattice models built using different sets of basis vectors." For instance, we could study chains joined only by the vectors of length $\sqrt{2}$ in lattice units, i.e., the basis vectors of the type [1,1,0] plus all sign combinations and all possible permutations (12 in total: [1,1,0], [−1,1,0], [1,−1,0], [−1,−1,0], [0,1,1], [0,−1,1], [0,1,−1], [0,−1,−1], [1,0,1], [−1,0,1], [1,0,−1], and [−1,0,−1]). This particular choice limits the possible location of chain points to a subset of the original cubic lattice that is called the face-centered cubic (fcc) lattice. Some sets of basis vectors have specific names (cubic, diamond, fcc, bcc, etc.) while others do not. In each case, listing all basis vectors constitutes a unique and sufficient definition of the lattice representation of the chain.

The actual physical length of the lattice unit is calculated by comparing the length of the basis vector in lattice units to the mean Cα—Cα distance (about 3.8 Å). For instance, the length of the [1,1,0]

Table I. Compilation of properties of various lattices.

Basis vectors	Lattice unit (Å)	No. vectors	Name	Ref.
100	3.80	6	Cubic	31
110	2.69	12	fcc	15
111	2.19	8	bcc	15
111[a]	2.19	8(4)	Diamond	12, 20
200.110	1.9	18		
220.111	1.9	14		18
210	1.7	24	Knight walk	32
200.211.110	1.9	42	Ext. FCC	18
210.211	1.7	48		
210.211.111	1.7	56	Hybrid	25
310.311.300.221.220	1.2	90	Ultra	26
111[b]	0.81	8(36)	Diamond	33

[a]Vectors are divided into two separate groups: The odd vectors are taken from one group and the even ones from the second. There are eight vectors in total (as in the 111 lattice), but only four are available for any given point.

[b]In this particular representation, Cα, C, and N backbone atoms are represented on the lattice; therefore, there are 4 * 3 * 3 possibilities to reach from one Cα atom to the next.

vector equals $\sqrt{2}$ lattice units. To make it equal to the Cα—Cα distance, the lattice unit must equal 2.69 Å ($3.8/\sqrt{2}$; see Table I for the length of the lattice units in other lattices). In this article, we discuss several commonly used sets of basis vectors that closely match the regularities of the protein backbone. Most have already been used or suggested by different authors to build models of protein structure.

A summary of some selected properties of different sets of basis vectors used for building lattice models of proteins is presented in Table I. In each case, the type of the basis vectors is given, and then all possible permutations and sign combinations must be added, as explained above for the case of the ([110]) lattice. The diamond lattice is built from eight [1,1,1]-type vectors, but only the combinations of pairs of basis vectors that produce a tetrahedral bond angle (109.5°) are allowed (so, for instance, the [1,1,1] vector might be followed by [1,1,−1] but not [1,−1,−1]). This restriction limits the number of possible paths leading from any single point to four (including the one leading to this point). On the other hand, in the full (all heavy atom) backbone diamond lattice only every third point is equivalent to a Cα position. So, when we start from any particular Cα atom there are 4 possible N positions (three if we exclude the path leading to the Cα under consideration), for each N atom there are 3 C positions (reversal is not allowed), and again there are 3 Cα positions for each C position, bringing the total of Cα neighbors of the first Cα atom to 36 (27 for the Cα atom in the middle of the chain. In some cases, all the basis vectors have the same length {([100]), ([110]), ([111]), diamond and ([210]) lattices}, while in others combinations of vectors of different lengths are allowed {([200], [111]), ([200], [110]), ([200], [110], [211]), ([210], [211], [111]) and ([310], [311], [300], [221], [220]) lattices}. In such cases, the first vector that is listed is used to define the actual length of the lattice unit, as discussed above. The different lengths of the basis vectors do not reflect the variation of the Cα—Cα distances in real proteins: instead, they are introduced to increase the flexibility of the lattice so as to reduce the accumulation of errors in the construction of the model protein backbone (see the discussion below).

Three different proteins of a moderate size were chosen for the analysis. They are (the proteins are identified by their PDB[24] abbreviations): myoglobin (1mba), triose phosphate isomerase (2ypi), and plastocyanin (1pcy). They represent three different structural classes (all α, mixed α/β and all β, respectively) and are sufficiently large that for any axis orientation some secondary structural element is in an unfavorable orientation. Rop protein (1rop), fragments of bacteriochlorophyll-A protein (3bcl, residues 39–69 and 39–50), and myohemerythrin (2mhr, residues 40–65) serve as examples of supersecondary and secondary structure fragments (α and β hairpins, a β strand, and an α helix, respectively) and crambin (1crn) is used to compare the present results to those obtained in previous studies.

For a given lattice, the lattice representation of a particular protein structure is built by one of the following three procedures:

1. The lattice chain is built step by step, starting from the lattice position closest to the position of the first Cα atom position. Subsequent lattice Cα positions are added by choosing the lattice position closest to the next protein Cα, subject to the constraint that the new position can be connected to the existing chain by one of the basis vectors allowed in the particular lattice model. The chain is not allowed to intersect or overlap. This procedure is strictly deterministic and depends solely on the orientation of the protein structure being fitted. It does not produce the best possible fit, as a decision that is optimal at the beginning of

Table II. Quality of protein models in different lattices.

Lattice	α	β	$\alpha\alpha$	$\beta\beta$	1crn	1mba	2ypi	1pcy
100	1.7	2.5	1.9	2.8	2.4	2.7	3.3	4.3
110	1.4	0.7	1.5	1.6	1.6	1.9	2.0	3.3
111	1.6	1.2	2.0	1.9	2.0	2.4	2.8	2.9
Diamond	3.4	1.2	3.8	2.3	3.2	4.0	3.8	3.8
200,110	1.1	1.4	1.3	1.5	1.4	1.5	2.0	1.9
200,111	1.2	1.0	1.4	1.5	1.4	1.6	1.9	1.8
210	1.1	0.7	1.2	0.9	1.2	1.5	1.5	1.2
Ext. FCC	1.0	0.8	1.1	1.0	1.0	1.2	1.1	1.2
Hybrid 211	0.7	0.6	0.9	0.8	0.8	1.0	1.0	1.0
Ultra 311	0.5	0.3	0.6	0.6	0.6	0.7	0.7	0.7
Backbone	1.0	0.7	1.2	1.0	1.2	1.2	1.5	1.5

rms difference in Å between the lattice model and crystallographic Cα coordinates.

chain construction may result in a poorer structure later in the building process. In fact, for low-coordination lattices this procedure sometimes reaches a dead end (the chain cannot propagate without violating excluded volume) and the fitting procedure must retrace its steps to find another solution.

2. A lattice chain of appropriate length and random conformation is built. The penalty function is calculated as the sum of the squared distances between the lattice Cα positions and the equivalent real Cα positions, which are not allowed to move. Sets of one, two, and three atom movements are automatically constructed from the set of basis vectors by the program, and simulated annealing is performed to minimize the penalty function. The result depends both on the parameters used in the minimization and the orientation of the protein molecule.
3. This approach is similar to the second, but here the rms distance between equivalent Cα positions after optimal superposition is used as a penalty function. As in 2, simulated annealing is performed to minimize the penalty function. The results depend on the parameters used in the minimization, but do not depend on the orientation of the protein molecule.

The third approach is used to obtain the best possible models in a given lattice, as reported in Table II. The second algorithm is used to obtain best models for a given orientation of the protein structure, such as illustrated in Figures 3 and 4.

In both cases, the best fits are obtained by using the first procedure to generate the initial conformation for use by the second or third procedure. As the lattice fits are obtained by simulated annealing, there is no guarantee that these are the lowest possible values. However, the same values were obtained in a number of independent minimizations, so there is a reasonable assurance that the results presented here are correct.

The excluded volume requirement is satisfied only for the Cα atoms, and no provision is made for the

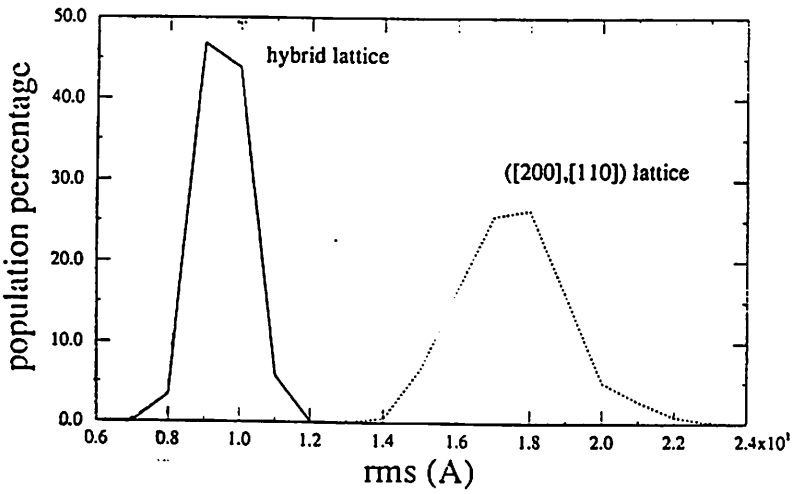

Figure 3. Distribution of rms of the best lattice fits to the native crambin structure in the ([200], [110]) lattice and the hybrid ([210], [211], [111]) lattice. The crambin structure was rotated around Euler angles Φ, Ψ, and Θ (six intervals in cos Θ were used) in 30° increments, and lattice fits to the rotated structure were grouped together according to their rms value.

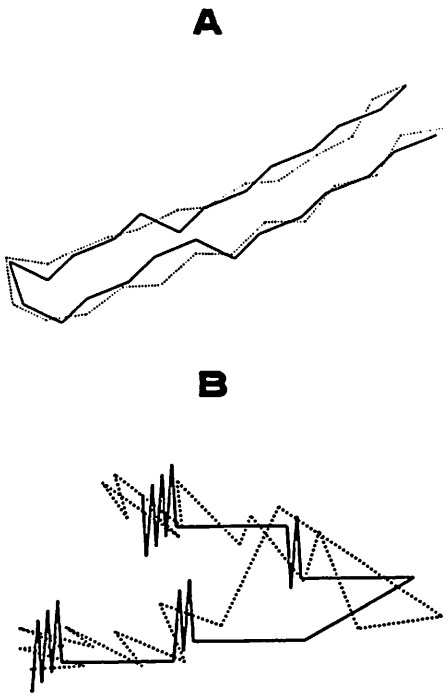

Figure 4. Example of the ([110]) lattice fit (solid line) to a β hairpin structure, (dotted line) as seen in: (a) from the side; (b) from the top.

introduction of side chains. It is possible to introduce side chains into lattice models and full side-chain libraries for the ([210]), the hybrid ([210], [211], [111]), and the ultra ([310], [311], [300], [221], [220]) lattice have been built and used in folding simulations.[22,25-27] However, the detailed analysis of the lattice structure with side chains, including the comparison of interaction patterns, goes beyond the scope of the present article, and will be presented in a separate publication.

RESULTS

The combination of procedures 1 and 3, as described above, was used to find the best possible lattice fit to the 8 test structures using 12 different lattices. The results, summarized in Table II, give an indication of how close a given lattice type can describe the protein structure. There is an obvious correlation between the number of available basis vectors and the quality of the fit, with the 90-vector "ultra" ([310], [311], [300], [221], [220]) lattice giving structures with the lowest rms. The cubic and diamond lattice (with six and four basis vectors, respectively) fits have the worst overall quality. Still, all of the lattices do a good job of describing the overall protein structure, with clearly recognizable topology, domains, and some secondary structure elements.

An important indicator of the lattice quality in describing protein structure can be obtained by comparing results for small protein fragments to that of larger proteins. For all lattices, it is more difficult to fit larger structures, but a few of them lose so much detail as to become almost unacceptable for larger proteins. This problem results from two effects discussed previously:

1. The dependence of the lattice structure on the relative orientation between the lattice principal axis and the protein being fitted, as discussed in the introduction. This effect is better illustrated in Figure 3, where the distribution of the rms of the best fit is plotted for the ([200], [110]) and the hybrid ([210], [211], [111]) lattices for all possible rotations of the original crambin crystal structure around Euler angles with the discrete steps of 10°. Here, procedure 2 (see above), where the target protein structure is not allowed to rotate, was used to generate lattice models. It is immediately obvious that these two lattices behave differently. Both lattices seem to be of acceptable quality on the basis of the data shown in Tables II and III. But, as seen in Figure 3, structures of such quality are obtained only for the few percent of the total number of orientations of the ([2,0,0], [1,1,0]) lattice, and for some orientations the lattice fit is bad.
2. In addition, some lattices can describe some types of secondary structures better than others. For instance, the ([100]) (cubic) lattice strongly favors α helical structures, and the diamond lattice does

Table III. Quality of protein models in different lattices.

Lattice	α	β	$\alpha\alpha$	$\beta\beta$	1crn	1mba	2ypi	1pcy
100	37	14	51	36	66	69	78	85
110	23	14	36	68	35	42	55	76
111	42	15	42	49	76	52	68	80
Diamond	69	14	73	107	85	76	89	97
200,110	21	15	43	35	41	34	48	47
200,111	24	14	33	66	42	34	55	44
210	17	16	21	40	30	40	40	35
Ext. FCC	16	19	20	25	26	20	29	32
Hybrid 211	16	21	15	25	20	20	27	23
Ultra 311	8	12	12	17	16	14	17	14
Backbone	26	29	37	58	51	36	55	43

Local similarity of lattice models of proteins as measured by $\langle\Theta\rangle$. See the text for additional details.

a better job of fitting β structures (see Table II). Thus, these lattices can achieve relatively good fits for short proteins of the appropriate type. For larger, mixed-motif proteins, this advantage disappears.

The crucial question that arises at this point is how to interpret the values displayed in Table II. When comparing real proteins, an rms on the level of 3.0 Å between 70% of the side chains would be regarded as close similarity.[28] However, here we are comparing real proteins to artificial structures, devoid of many features that are taken for granted in real proteins and possessing other features that may not be typical of proteins. Therefore, we introduced another measure of similarity between the real and the model structure. First, for each Cα atom the vector that bisects the angle formed by the three consecutive Cα atoms is constructed. It gives an indication of the direction of the side chain and for real structures is close to the Cα—Cβ vector (the Cβ positions are not included in the Cα lattice structures). Later, the angle Θ between such vectors for the real and model structures is calculated, and the mean value of this angle $\langle\Theta\rangle$ is presented in Table III. The average value for two uncorrelated chains is 90°, and values larger than 90° indicate anticorrelation and values smaller than 90° increasing correlation. If the directions of bisector vectors of the model structure are not correlated with the appropriate directions in the real structure, this means that the structural elements in the model have lost their hydrophobic/hydrophilic "faces," and despite the model backbone having the same topology the side chains have random orientations.

Table III presents the mean angular correlation $\langle\Theta\rangle$ for structures with the best rms, obtained with the procedure described previously. The results in Table III were calculated for the same structures as presented in Table II, and the angular correlation was not used in the fitting procedure. For the high-correlation lattices, it is still possible to improve the angular correlation with little price paid at the rms level. It is not possible to do the same for most low-coordination lattices, such as the cubic and diamond lattices, which therefore must be completely dismissed for building higher-quality models.

As expected, the higher coordination number lattices perform much better, and the difference between the native and the lattice protein can reach 10–20°. For the secondary structure fragments, particularly for the β strand, almost all lattices can produce a good model, but already at the level of two adjacent elements of secondary structure the quality of the fit gets much worse. This effect is much more pronounced for low-quality lattices. At the rms level, it may seem that it is much easier to reproduce the β structures, but as seen in Table III these seemingly good fits have severe orientation problems, with β strands in a β sheet facing each other instead of being parallel to each other (see Fig. 4).

For the α helices, the main problem is the strong difference between the best fit obtained for the optimal orientation of the lattice axis relative to the helix direction and fits obtained for other orientations. Again, this problem is much more pronounced for the low-coordination lattices. This is illustrated in Figure 5a for the ([110]) lattice. Here, the best fit (left-hand side) gives a reasonable looking helix (rms of 1.7 Å and angular correlation of 37°). For the worst orientation (right-hand side), the lattice representation deteriorates to the point that the model struc-

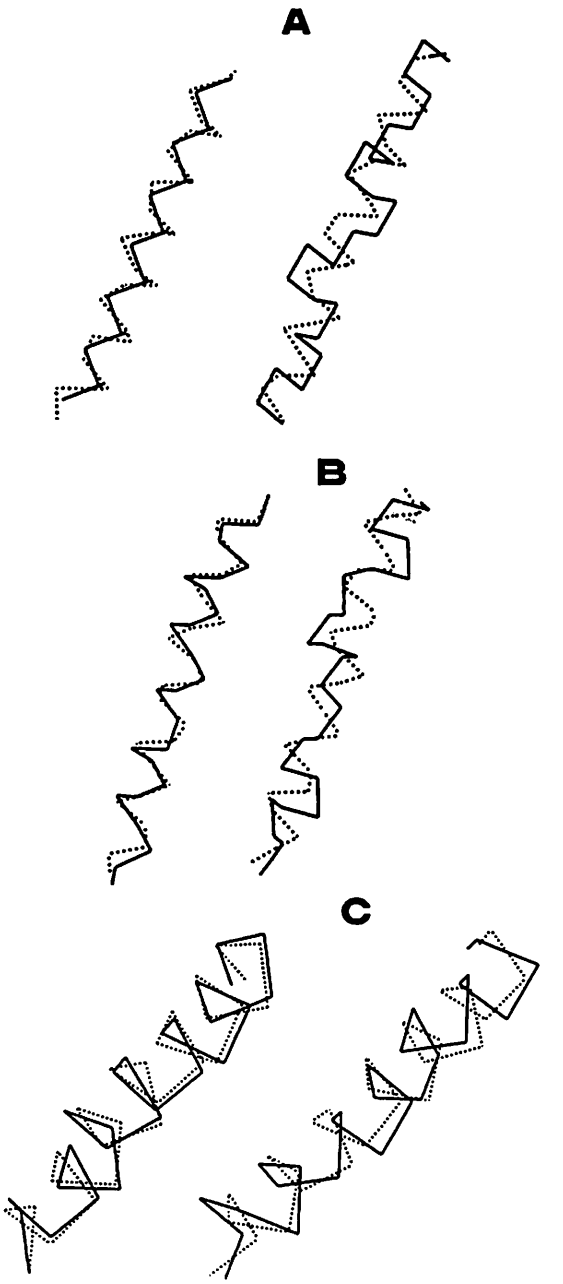

Figure 5. Examples of lattice fits (solid line) to a real α helix (dotted line) (a). Simple cubic lattice. (b). 210 lattice. (c). Hybrid lattice. For each lattice, the left-hand side shows the best lattice fit and the right-hand side the worst lattice fit.

ture (rms of 2.5 Å and angular correlation of 55°) is not even helical. As illustrated in Figure 4b, an intermediate-quality lattice, like the ([210]) lattice, on going from the best (1.1 Å, 17°) to the worst (2.0 Å, 45°) fit loses its regularity but retains the chiral character of the helix. At the other end of the spectrum, as seen in Figure 4c, the hybrid lattice with both the best (0.7 Å and 16°)- and worst (1.2 Å, 22°)-fit models are helical and regular.

CONCLUSIONS

In this article, we examined a number of different proteins and protein fragments in the context of a set of discretized models of protein structure. A series of lattices of increasing fidelity to protein structures have been compared. Not surprisingly, we demonstrated that by increasing the number of basis vectors available for backbone construction the lattice model of protein structure can be made as close to the real protein structure as required. By all measures of structural similarity, high-coordination lattices employed here can describe protein structure with a high level of accuracy.

The simplest lattices, such as the cubic or the diamond lattice, can describe the protein structure on the level of 3–4 Å rms. At this level, the lattice model has the correct overall shape and topology, but local regularities, such as the hydrophobic/hydrophilic faces of secondary structure elements, are almost completely lost. For most orientations, it is not possible to represent secondary structure elements other than the one favored by the particular lattice, for instance, a β sheet for the diamond lattice and an α helix for the cubic lattice.

On the next level, lattices such as the bcc and fcc can describe some proteins and protein fragments with high fidelity, yet their overall level of accuracy is closer to 2–3 Å, with serious problems for some orientations and some proteins.

The general level of description improves to 1–2 Å for the ([200], [111]), ([200], [110]), ([200], [211], [111]), and [210] lattices, in order of increasing quality. However, for some orientations the secondary structure still loses its regularity and the quality of the fit visibly deteriorates with increasing size of the molecules.

It is only at the level of the all-backbone atom, diamond lattice, the hybrid ([210], [211], [111]), or the ultra ([310], [311], [300], [221], [220]) lattice that one has a correct level of secondary structure description even for the most unfavorable orientation of the original protein structure. Protein models built using these lattices reproduce such details as the twist of β sheets, the preferred α-helix/α-helix packing angles, etc. The ultra lattice improves the quality of the model even further, with such details as the hydrogen bond pattern in secondary structure elements described correctly.

In general, it is number of the basis vectors that is most important for the quality of the lattice protein model. The main reason for this is that the differences between the lattice model and the actual protein accumulate quickly, and it is the ability to relax these differences that ultimately decides the quality of the lattice model.

It is, of course, possible to design even higher-order lattices that could describe protein structures even closer. The accuracy level presented by the hybrid ([210], [211], [111]) lattice is good enough for most demanding applications, including the *de novo* folding of simple motif proteins.[29] The ultra lattice ([310], [311], [300], [221], [220])[26,27] has been used in the refinement of the folded structures, with the possibility of predicting structures on the level of 2.1–3 Å.

Thus, we conclude that after deciding about the level of accuracy necessary for any particular purpose one can always find a lattice that can fulfill the necessary requirements. But, when using lattice models one always has to be aware of certain problems specific to the lattice representation: the dependence on the orientation in space, the error accumulation in some regions of the protein, and the intrinsic lattice preference for certain types of secondary structures. It is only in the limit of high-coordination lattices that these problems are effectively eliminated, and one can use a lattice representation of protein structure with confidence that the derogatory effects of discretization are negligible and their advantages can be safely exploited.

This brings up the important question of how much the results of the lattice simulations reported in the literature are affected by the effects discussed here. For simulations in which the folding pathways are studied, the effects are negligible. In fact, the results even do not depend on whether the system is on lattice or not.[30] In these simulations, the target structure was known in advance and the type and mutual orientation of secondary structure elements was known to be compatible with the lattice. Also, the local secondary structure was specified in advance by defining the set of preferred local $C\alpha$—$C\alpha$ distances, and this helped enforce the regularity of the local secondary structure—where deficiencies of the lattice are most severe. If, however, no information about the correct secondary structure is provided and the final structure is not known in advance, some of the lattice deficiencies become crucial. When secondary structure formation depends on the details of local interactions, the quality of the representation becomes more important. If the existence of the given secondary structure element depends on lattice orientation, it might not be possible to describe the final structure at a low-coordination lattice. These problems motivated the de-

velopment of the high-coordination lattices, where, as illustrated in this article, most of them are eliminated.

This research was supported in part by Grant GM-37408 from the Division of General Medical Sciences, National Institutes of Health.

References

1. R.L. Jernigan, *Curr. Opin. Struct. Biol.*, **2**, 248 (1992).
2. M. Karplus and G.A. Petsko, *Nature*, **347**, 631 (1990).
3. M. Levitt and A. Warshel, *Nature*, **253**, 694 (1975).
4. I.D. Kuntz, G.M. Crippen, P.A. Kollman and D. Kimelman, *J. Mol. Biol.* **106**, 983 (1976).
5. A.T. Hagler and B. Honig, *Proc. Natl. Acad. Sci. USA*, **75**, 554 (1978).
6. B. Robson and D.J. Osguthorpe, *J. Mol. Biol.*, **132**, 19 (1979).
7. C. Wilson and S. Doniach, *Proteins*, **6**, 193 (1989).
8. S. Tanaka and H.A. Scheraga, *Macromolecules*, **9**, 945 (1976).
9. S.V. Narayama and P. Argos, *Int. J. Pept. Prot. Res.*, **24**, 25 (1984).
10. S. Miyazawa and R.L. Jernigan, *Macromolecules*, **18**, 534 (1985).
11. S.H. Bryant and C.E. Lawrence, *Proteins*, **16**, 92 (1993).
12. V.G. Dashevskii, *Mol. Biol.*, **14**, 105 (1980).
13. N. Go and H. Taketomi, *Proc. Natl. Acad. Sci. USA*, **75**, 559 (1978).
14. H. Taketomi, F. Kano, and N. Go, *Biopolymers*, **27**, 527 (1988).
15. W.R. Krigbaum and S.F. Lin, *Macromolecules*, **15**, 1135 (1982).
16. S. Miyazawa and R.L. Jernigan, *Biopolymers*, **21**, 1333 (1982).
17. K.F. Lau and K.A. Dill, *Macromolecules*, **22**, 3986 (1989).
18. D.G. Covell and R.L. Jernigan, *Biochemistry*, **29**, 3287 (1990).
19. E.I. Shakhnovich and A.M. Gutin, *Biophys. Chem.*, **34**, 187 (1989).
20. J. Skolnick and A. Kolinski, *Annu. Rev. Phys. Chem.*, **40**, 207 (1989).
21. D.A. Hinds and M Levitt, *Proc. Natl. Acad. Sci. USA*, **89**, 2536 (1992).
22. A. Godzik, J. Skolnick, and A. Kolinski, *Proc. Natl. Acad. Sci. USA*, **89**, 2629 (1992).
23. M. Levitt, *J. Mol. Biol.*, **104**, 59 (1976).
24. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Simanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
25. A. Kolinski and J. Skolnick, *Chem. Phys.*, **97**, 9412 (1992).
26. J. Skolnick, A. Kolinski, C. Brooks III, A. Godzik, and A. Rey, *Curr. Biol.*, **3**, 414 (1993).
27. A. Kolinski and J. Skolnick, *Proteins*, submitted.
28. A.M. Lesk and C. Chothia, *J. Mol. Biol.*, **136**, 225 (1980).
29. A. Kolinski, A. Godzik, and J. Skolnick, *J. Chem. Phys.*, **98**, 7420 (1993).
30. A. Rey and J. Skolnick, *Chem. Phys.*, **158**, 199 (1991).
31. Y. Ueda, H. Taketomi, and N. Go, *Biopolymers*, **17**, 1531 (1978).
32. A. Kolinski, M. Milik, and J. Skolnick, *J. Chem. Phys.*, **94**, 3978 (1991).
33. M. Milik and J. Skolnick, *Proc. Natl. Acad. Sci. USA*, **89**, 9391 (1992).