

The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics

Oren M. Becker, and Martin Karplus

Citation: *The Journal of Chemical Physics* **106**, 1495 (1997);

View online: <https://doi.org/10.1063/1.473299>

View Table of Contents: <http://aip.scitation.org/toc/jcp/106/4>

Published by the *American Institute of Physics*

Articles you may be interested in

[Free energy disconnectivity graphs: Application to peptide models](#)

The Journal of Chemical Physics **117**, 10894 (2002); 10.1063/1.1517606

[A climbing image nudged elastic band method for finding saddle points and minimum energy paths](#)

The Journal of Chemical Physics **113**, 9901 (2000); 10.1063/1.1329672

[Equation of State Calculations by Fast Computing Machines](#)

The Journal of Chemical Physics **21**, 1087 (1953); 10.1063/1.1699114

[Markov models of molecular kinetics: Generation and validation](#)

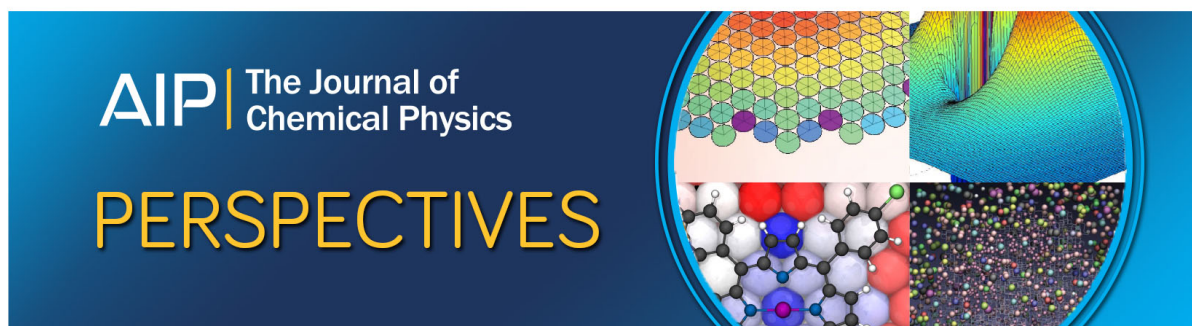
The Journal of Chemical Physics **134**, 174105 (2011); 10.1063/1.3565032

[Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points](#)

The Journal of Chemical Physics **113**, 9978 (2000); 10.1063/1.1323224

[Comparison of simple potential functions for simulating liquid water](#)

The Journal of Chemical Physics **79**, 926 (1983); 10.1063/1.445869



The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics

Oren M. Becker

School of Chemistry, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel and Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138 and Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, 67000 Strasbourg, France

(Received 15 April 1996; accepted 18 September 1996)

Topological characteristics of multidimensional potential energy surfaces are explored and the full conformation space is mapped on the set of local minima. This map partitions conformation space into energy-dependent or temperature-dependent “attraction basins” and generates a “disconnectivity” graph that reflects the basin connectivity and characterizes the shape of the multidimensional surface. The partitioning of the conformation space is used to express the temporal behavior of the system in terms of basin-to-basin kinetics instead of the usual state-to-state transitions. For this purpose the transition matrix of the system is expressed in terms of basin-to-basin transitions and the corresponding master equation is solved. As an example, the approach is applied to the tetrapeptide, isobutyryl-(ala)₃-NH-methyl (IAN), which is the shortest peptide that can form a full helical turn. A nearly complete list of minima and barriers is available for this system from the work of Czerminski and Elber. The multidimensional potential energy surface of the peptide is shown to exhibit an overall “funnel” shape. The relation between connectivity and spatial proximity in dihedral angle space is examined. It is found that, although the two are similar, closeness in one does not always imply closeness in the other. The basin to basin kinetics is examined using a master equation and the results are interpreted in terms of kinetic connectivity. The conformation space of the peptide is divided up in terms of the surface topography to model its “folding” behavior. Even in this very simple system, the kinetics exhibit a “trapping” state which appears as a “kinetic intermediate,” as in the folding of proteins. The approach described here can be used more generally to classify multidimensional potential energy surfaces and the time development of complex systems. © 1997 American Institute of Physics. [S0021-9606(97)50948-7]

I. INTRODUCTION

Mesoscopic systems with many degrees of freedom include liquids, glasses, solutions, and macromolecules. They are under intensive study in physics, chemistry, and biology. In all cases the thermodynamic and dynamic properties are determined by the nature of the potential energy surfaces.¹ Because of the complexity of the potential surfaces of such mesoscopic systems, a method for analyzing their multiple minima can serve a useful role. It can aid, for example, in the interpretation of conformational transitions in peptides and the kinetics of protein folding.^{2,3} Also, an understanding of the special properties of glasses and the nature of the glass transition³⁻⁵ depends on a detailed characterization of the potential surface.

Potential energy surfaces (PES) can be characterized by their minima, which correspond to locally stable configurations, and by transition regions connecting the minima. In small systems, which have only few minima, it is possible to use a direct approach and describe the entire potential energy surface. Such is the case for small reactive systems⁶ and for the alanine dipeptide, which has only two significant degrees of freedom.^{7,8} For systems with many degrees of freedom and a very large number of minima, a direct approach to the PES becomes very difficult.

Stillinger and Weber introduced a useful procedure for characterizing the multimimum PES of large systems.^{9,10} They investigated the local minima by taking configurations from a molecular dynamics trajectory and quenching the points along the trajectory to the nearest local minimum by following the steepest descent path. In this way, a sampling of the PES and local minima accessible at a given temperature were obtained. The minima provide a “mapping” of the PES, referred to as “hidden structures,” which they used to separate the partition function into thermal and configurational contributions.⁹ This procedure has been applied to a variety of systems, including water,^{10,11} rare gas clusters^{12,13} and the proteins myoglobin¹⁴ and bovine pancreatic trypsin inhibitor (BPTI).¹⁵ The protein studies showed that there are a very large number of minima in the vicinity of the native state and that the local minima are kinetically “clustered” into subsets; i.e., within a cluster or “basin” the minima tend to be connected by low barriers.¹⁵

A map of the minima and regions in their neighborhood gives only a partial description of the energy surface since it does not contain information about the barriers. While the set of minima and the surrounding portions of the potential surface are often sufficient to determine the thermodynamics,⁹ it is the barrier distribution that is required for the kinetics. A

knowledge of the transition states connecting the minima makes possible an exploration of the systems kinetics by use of the master equation approach.¹⁶ Recently, a rather complete map of the minima and barriers of a tetrapeptide in vacuum was obtained by Czerminski and Elber^{17,18} and a more limited study of transitions states in TIPS2 water was performed by Tanaka and Ohmine.¹¹

Experimental evidence concerning the complexity of the PES for a system with many degrees of freedom can be obtained by the observation of a multiplicity of relaxation times.¹⁹ A particularly interesting study has been performed by Frauenfelder and collaborators^{20–22} of the complex kinetics of CO to myoglobin. The results have been explained in terms of the existence of a hierarchy of minima, referred to as “conformational substates” that are thought to be arranged in “tiers” corresponding to different energies and barriers. The question of the role of solvent in the observed relaxation behavior has been raised recently.²³ The substates were introduced to explain the different time scales on which certain kinetic phenomena take place in this system. It was assumed that, in general, slower phenomena are associated with larger scale motions. These studies, which represent the most detailed analysis of the dynamics of folded proteins, have inherent in them the limitations of an experimental approach. Because only a few probes can be used to obtain kinetic data, the interpretation is based on simplified models and the specific connection between the dynamics and structure is not clear; e.g., in a system as inhomogeneous as a protein, very localized motions can have very high barriers.²⁴ Peptides and proteins differ in an essential way from the glasses usually investigated in studies of complex potential surfaces and multiple time scales.^{5,19,25} The latter are inherently homogeneous; i.e., on an infinite time scale, each region of the usual glassy system is equivalent to every other region. By contrast, a protein is inherently inhomogeneous due to the nature of the native structure and the amino acid sequence of which it is composed.

Many terms have been used in describing various attributes of complex potential energy surfaces. They include “funnels,”^{26,27} “clusters of minima,”¹⁵ and “tendrils.”⁹ All these represent attempts to describe the general *shape* and overall *connectivity* that characterize multidimensional potential surfaces. Since the concepts of shape and connectivity are within the realm of topology the present paper attempts to describe potential energy surfaces using a topological approach.

In theoretical studies of the gas-phase chemistry of small molecules, it is customary to focus on state-to-state transitions as the property of fundamental interest, although rate constants which represent averages over initial and final states are often used.⁶ For systems with many degrees of freedom there are alternative approaches for analyzing the kinetics of the transitions among the multiple minima. In rare gas clusters^{12,13} and proteins, for example, where individual states are usually clustered in “basins,” the interesting kinetics involves basin to basin transitions. The internal distribution within a basin is expected to approach equilibrium on a relatively short time scale, while the slower basin-to-basin

kinetics, which involves the crossing of higher barriers, governs the intermediate and long time behavior of the system. Experimental studies of protein kinetics, such as those of the Frauenfelder group^{20–22} measure only certain basin-to-basin transitions, since they cannot resolve the individual localized minima.

In the present work we use topology to define a map of the potential energy surface that emphasizes its *shape* and *connectivity*, and thus, offers a tool for a better understanding of the physical system. This map, which partitions configuration space into energy or temperature-dependent basins, generates a formal “disconnectivity” graph that is used to characterize the multidimensional potential surface. Combined with an appropriate metric, this method can be applied to qualitative and quantitative studies of a variety of systems. In terms of kinetics, the topological connectivity is manifested as basin-to-basin kinetics, which is used to reformulate the transition matrix and master equation. To illustrate the topological concepts, an application to the analysis of a tetrapeptide potential energy surface and its kinetic consequences is presented. The results for the simple peptide are used to draw conclusions concerning the behavior of more complex systems.

In Sec. II several procedures are defined for mapping the overall potential energy surface on the set of local minima, and in Sec. III the resulting disconnectivity graph is defined. In Sec. IV this graph is related to the topography of the potential energy surface while Sec. V uses it to interpret various physical phenomena. A detailed discussion of basin-to-basin kinetics is given in Sec. VI. Section VII presents the tetrapeptide example, which is followed by a discussion in Sec. VIII.

II. MAPPING AND BASINS

A molecule composed of N atoms has a $3N$ -dimensional conformation space $\mathbf{r}=(\mathbf{r}_1,\dots,\mathbf{r}_N)\in\mathfrak{R}^{3N}$,²⁸ where the \mathbf{r}_i are the vectors giving the position of atoms i in the three-dimensional space. The potential energy of the system, Φ , is a function defined over the $3N$ -dimensional conformation space, $\Phi(\mathbf{r}_1,\dots,\mathbf{r}_N)$. The Φ hypersurface has a multitude of local minima, which can be described as a discrete set indexed by α . The map $M^d(\mathbf{r})$ from the $3N$ -dimensional continuum \mathfrak{R}^{3N} to the discrete set of minima $\{\alpha\}$, was introduced by Stillinger and Weber;⁹ the notation we use is similar to that used by them. It is defined by direct minimization from any point in \mathfrak{R}^{3N} along a steepest descent path to the nearest local minimum,

$$M^d(\mathbf{r}):\mathfrak{R}^{3N}\rightarrow\{\alpha\}, \quad (1)$$

where the d superscripts stands for *direct* minimization or *direct* “quenching.”

Following Stillinger and Weber,⁹ let $R(\alpha)\subset\mathfrak{R}^{3N}$ denote the set of system configurations \mathbf{r} which map to a local minimum α . $R(\alpha)$ is a connected set, since all $\mathbf{r}\in R(\alpha)$ are connected by a path through α , and the different $R(\alpha)$ are disjoint. The significance of this mapping is that the $R(\alpha)$ partition the $3N-6$ dimensional configuration space into

“attraction basins” around the α minima. The boundaries of these sets lay on the $(3N-7)$ dimensional bifurcation hypersurface. The union of the boundaries constitutes a zero measure set in the configuration space \mathfrak{R}^{3N} for which the mapping M^d is undefined. Since the boundaries are not included in the $R(\alpha)$ basins, they form open sets.

When the potential energy surface (PES) represents an ensemble of small molecules (e.g., a water cluster), the number of distinct $R(\alpha)$ can be further reduced by forming equivalence classes from all $R(\alpha)$ sets which are identical except for particle permutations.⁹ In this case a symmetry number σ is assigned to each equivalence class to account for this grouping. If the potential energy surface represents a single macromolecule, such as a protein, the bonding connectivity prevents such permutations. This is an aspect of the essential inhomogeneous character of the system mentioned in the Introduction. For such systems each $R(\alpha)$ basin is in an equivalence class by itself and the symmetry number associated with it is unity ($\sigma=1$).

Although the $M^d(\mathbf{r})$ map partitions the PES in a physically meaningful way (i.e., it gives the relation between any point in \mathfrak{R}^{3N} and the nearest minimum α), the information from this partition is limited. In particular, it does not contain information about the barriers between the minima. For example, the map is unable to distinguish between local minima which account for small defects in an otherwise ordered crystal and the minima in the disordered matrix of a glass. This limitation arises from the fact that the $R(\alpha)$ give dynamical information only at the $T \rightarrow 0$ limit, where any barrier, however small, traps the system in a given minimum.

To account for the behavior of a system at finite temperature, it is useful to have a description that groups minima which are connected by low barriers. This can be achieved by introducing “super basins”

$$R^T(\alpha') = \cup R(\alpha), \quad (2)$$

which are a union of all the $R(\alpha)$ sets connected by barriers Φ^\ddagger lower than kT (or some small multiple of kT); i.e., a super basin implies that the system moves rapidly between its component minima. Consequently, $R^T(\alpha')$ is defined as the union of all $R(\alpha)$ sets for which $\max(\Phi_{ij}^\ddagger, \Phi_{ji}^\ddagger) \leq kT$, where Φ_{ij}^\ddagger is the barrier going from basin j to i , and Φ_{ji}^\ddagger is the barrier going from i to j . We introduce both barriers to implicitly take into account the energy difference between basins. The symbol α' refers to the lowest minimum in $R^T(\alpha')$; i.e.,

$$\alpha' = \min\{\alpha | \alpha \in R(\alpha) \subset R^T(\alpha')\}. \quad (3)$$

We can define a new mapping procedure $M^T(\mathbf{r})$ that will partition the $3N$ -dimensional configuration space \mathfrak{R}^{3N} into super basins at temperature T and map it onto the smaller set of minimum $\{\alpha'\}_T$ defined by Eq. (3)

$$M^T(\mathbf{r}): \mathfrak{R}^{3N} \rightarrow \{\alpha'\}_T. \quad (4)$$

The resulting super basins are also disjoint open sets where the set $\{\alpha'\}_T$ depends on the temperature T . The significance of the M^T mapping is that it reflects the system connectivity at different temperatures. Assuming a *time scale* of observa-

tion, defined as the time in which equilibrium among all of the α minima associated with an individual $R^T(\alpha')$ is reached, the dynamics of the system at that temperature T is sensitive only to transitions between different $R^T(\alpha')$ basins, and not to the underlying structure within the basins themselves; i.e., this time scale, which depends on T , is less than the mean time for a dynamical trajectory to cross from one $R^T(\alpha')$ basin to another.

From the above definitions it is clear that the direct minimization mapping $M^d(\mathbf{r})$ is the $T \rightarrow 0$ limit of this more general mapping $M^T(\mathbf{r})$; i.e.,

$$\begin{aligned} M^d(\mathbf{r}) &= \lim_{T \rightarrow 0} M^T(\mathbf{r}) = \alpha, \\ R^{T=0}(\alpha') &= R(\alpha), \\ \{\alpha'\}_{T=0} &= \{\alpha\}. \end{aligned} \quad (5)$$

At the $T \rightarrow \infty$ limit, assuming that the physical system is well behaved (i.e., it does not have infinite barriers), all of configuration space is incorporated in a single super basin, and the α' minima that corresponds to it is the global minima α^g , i.e.,

$$\begin{aligned} R^{T \rightarrow \infty}(\alpha') &= \mathfrak{R}^{3N}, \\ \{\alpha'\}_{T \rightarrow \infty} &= \alpha^g. \end{aligned} \quad (6)$$

At this limit the system is ergodic and a dynamical trajectory spans the full conformation space. In the present context, ergodicity is defined as being time scale dependent. Namely, it means that the system is not trapped in any subregion of conformation space and that given a long but finite time, e.g., a time longer than the length of a “long” molecular dynamics trajectory (~ 10 ns), any dynamical trajectory will span all of the available space. Exceptions to Eq. (6) exist only in systems which have infinite barriers, $\Phi^\ddagger = \infty$, and are thus never ergodic. At finite T , the system completely spans any given $R^T(\alpha')$ on the specified time scale but not the whole space.

Similar to the $M^T(\mathbf{r})$ mapping, which is formulated in terms of the *canonical* ensemble using the relative barriers Φ_{ij}^\ddagger , a *microcanonical* mapping $M^E(\mathbf{r})$ can also be defined

$$M^E(\mathbf{r}): \mathfrak{R}^{3N} \rightarrow \{\alpha'\}_E. \quad (7)$$

This map uses the total energy E rather than the temperature T as the control parameter. The analogous microcanonical super basins $R^E(\alpha') = \cup R(\alpha)$ are defined by the mapping $M^E(\mathbf{r})$ as a union of elementary basins connected by barriers $\Phi^\ddagger \leq E$, where Φ^\ddagger is the energy value of the barrier. The significant difference between the two maps $M^E(\mathbf{r})$ and $M^T(\mathbf{r})$ is in the physical content of the result. The microcanonical map is concerned with the absolute energies and reflects the *topography* of the potential energy surface, while the canonical map highlights the relative energies and reflects the *kinetic connectivity* of the system. This difference in interpretation is further discussed in Secs. IV and V.

An important aspect of potential energy surface is that the dynamical system with a finite energy, $E < \infty$, is confined to a subspace Γ of the full configuration space, $\Gamma \subset \mathfrak{R}^{3N}$. This

subspace, which we name the “accessible configuration space,” is strictly defined as the union of all configurations which are accessible to the system at a given energy; i.e.,

$$\Gamma(E) = \{\mathbf{r} | \Phi(\mathbf{r}) \leq E\}, \quad (8)$$

where E is the total energy of the system. Γ is bound by the $\Phi=E$ isoenergy manifold, which is the classical “turning” manifold of the system; i.e., the manifold on which any trajectory (with total energy E) instantaneously has zero kinetic energy. This corresponds to the use of constant energy contours to characterize the PES.³⁰ The disjoint sets $R^E(\alpha')$ cover this accessible Γ space since their union is Γ ; i.e.,

$$\Gamma(E) = \cup R^E(\alpha'). \quad (9)$$

The accessible configuration space Γ is more loosely defined in the canonical representation, since the system has a nonzero (though often small) probability for reaching every point in configuration space. The canonical accessible configuration space is therefore the region in which the systems spends *most* of its time [where the time scale was defined following Eq. (4)]. We define this canonical accessible configuration space as

$$\Gamma(T) = \{\mathbf{r} | \Delta\Phi(\mathbf{r}) \leq nkT\}, \quad (10)$$

where $\Delta\Phi(\mathbf{r})$ is the potential energy at \mathbf{r} relative to the global minimum [$\Delta\Phi(\mathbf{r}) = \Phi(\mathbf{r}) - \Phi(\alpha^g)$], and n is a small number (which depends on the particular time scale used). As T increases the accessible configuration space Γ becomes larger.

Although the “ergodic limit” (as defined above) is formally reached only as $T \rightarrow \infty$, it is expected to be reached at finite temperatures in many physical systems. It is physically reasonable, therefore, to assume that there is an “ergodic” temperature $T_{\text{erg}} < \infty$ for which all of the accessible configuration space (defined in terms of long but finite times) is incorporated in a single basin; i.e.,

$$R^{T_{\text{erg}}}(\alpha') = \Gamma \subset \mathcal{R}^{3N}. \quad (11)$$

$M^{T_{\text{erg}}}(\mathbf{r})$ maps all configurations $\mathbf{r} \in R^{T_{\text{erg}}}(\alpha') = \Gamma$ to the deepest minimum, α'^g , in the accessible Γ space. Although there is no guaranty that $\alpha^g \in \Gamma$ (i.e., that the global minimum α^g is within the accessible configuration space Γ) it is likely to be true for most cases of physical interest. Equation (11) also assumes that the Γ space is connected. Again, this assumption is expected to be met for most physical cases, although in principle there can be systems in which Γ becomes connected only at $T \rightarrow \infty$, when it becomes identical to \mathcal{R}^{3N} .

III. THE TOPOLOGICAL DISCONNECTIVITY GRAPH

The map and its partitioning of configuration space into basins can be investigated further by use of topology.^{31–34} In particular, it allows us to use the set of local minima $\{\alpha\}$, instead of the continuous $R(\alpha)$ basins, in the analysis of the topology of the potential energy surface. Details of this analysis are given in the Appendix.

An interesting topological structure is obtained when considering the set-of-sets generated by applying the $M^T(\mathbf{r})$ map or the $M^E(\mathbf{r})$ map several times each at a different temperature or energy [the set A_3 in Eq. (A4)]. Each application

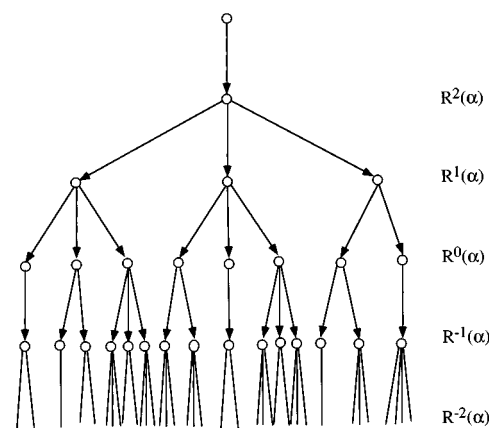


FIG. 1. A sample disconnectivity graph of a potential energy surface. Each vertex on this graph represents an attraction basin $R^T(\alpha')$ [or $R^E(\alpha')$] and the successive horizontal levels correspond to the different temperatures T (energies E) in which the mapping was performed. The arrows represent inclusion relations between vertices (basins).

of these mapping procedures yields a different partitioning of the potential energy surface, and the way these partitionings change as a function of temperature or energy reveals the underlying topography of the surface. Technically, this relationship can be described by a graph which is defined by the inclusion relationship of the different basins, i.e., the way smaller basins coalesce to form larger basins. An example of such a graph is given in Fig. 1. Each vertex on this graph represents an attraction basin $R^T(\alpha')$ [or $R^E(\alpha')$] and the successive horizontal levels correspond to the partitioning obtained from the map at different temperatures T (or energies E). Using common conventions, which represent the inclusion relation $x \supset y$ by assigning an arrow $x \rightarrow y$ to the edges (x, y) of the graph, we see that the graph in Fig. 1 has a very well defined directionality associated with it. All the arrows on this graph are directed downward, pointing in the direction of splitting super basins into smaller sub-basins. Therefore, we name this directed graph, $G(\Phi)$, the “disconnectivity graph” of the potential energy surface Φ (in future references to this graph we omit the arrows). Figure 1 is one possible example of $G(\Phi)$; other examples may have different branching characteristics.

The disconnectivity graph $G(\Phi)$ has a natural hierarchy it (see Fig. 1). Each vertex at a level j of the graph corresponds to a basin $R^j(\alpha')$ defined by the mapping $M^{T_j}(\mathbf{r})$ at a temperature T_j . All basins associated with an arbitrary “current” temperature (or energy) are indexed $j=0$, i.e., $R^0(\alpha')$. All “parent” basins, which belong to those levels of the graph associated with higher temperatures (energies), are indexed $R^1(\alpha'), R^2(\alpha')$ etc., where larger indices $j=1, 2, \dots$ correspond to higher temperatures (energies). All “descendent” basins, which belong to levels of the graph laying below the current level, are indexed $R^{-1}(\alpha'), R^{-2}(\alpha')$ etc., where smaller indices $j=-1, -2, \dots$ correspond to lower temperatures (energies). Thus, the graph $G(\Phi)$ preserves the temperatures (energies) related connectivity information as well as the minima information.

Another important characteristic of the graph $G(\Phi)$ is the number of edges originating at each vertex. This quantity is the main feature that distinguishes one graph from another. We call this branching number the *multiplicity* m of a vertex.

Although $G(\Phi)$ does not contain information about the size of a given attraction basin in the $3N-6$ dimensional conformation space, it does provide a *qualitative* measure of the *relative size* of the attraction basins associated with the graph's vertices; i.e., the number of local minima included in the basin. It is expected that often, though not always, basins which include many minima will have a larger volume in configuration space than basins which include only few minima. Thus, a qualitative measure is obtained from the number of branches below a given vertex; i.e., the configuration space volume of a basin $R^T(\alpha')$ is roughly proportional to the sum over the multiplicities of all the vertices which are under the vertex of interest

$$\int_{R^{3N}} R^T(\alpha') \propto 1 + \sum_j (m_j - 1), \quad (12)$$

where m_j is the multiplicity of a vertex j and the \propto sign stands for proportionality. The term on the left is the configuration space volume of basin $R^T(\alpha')$, and the summation on the right is over all the vertices of the graph which are under the vertex of interest. The "1" stands for the central branch and the $(m_j - 1)$ summation counts the number of edges that branch off the main branch. The relation given in Eq. (12) is not rigorous and there may be situations that do not conform to it; e.g., a small rough region, with many local minima, in an otherwise relatively smooth surface will be overweighted by Eq. (12).

The graph $G(\Phi)$, which highlights the basin structure of the PES, is related to the local connectivity tree, which represents all the local barriers connecting two local minima on the PES. In fact $G(\Phi)$ can be considered to be a coarse grained version of this local connectivity tree. However, while the latter is characterized by binary branching along a continuous energy scale (indicating the pairwise character of the barriers), in $G(\Phi)$ the vertices have a higher multiplicity and are arranged on a discrete energy level. At any level of the $G(\Phi)$ graph, all minima that are interconnected by barriers lower than the energy (or temperature) that defines that level are joint to a single vertex of the graph. Namely, any two minima that belong to the same vertex (basin) are connected by a path with barriers not higher than the defining energy (temperature) of that level.

IV. PES TOPOGRAPHY AND THE $G^E(\Phi)$ DISCONNECTIVITY GRAPH

The disconnectivity graph $G^E(\Phi)$ generated by the microcanonical map $M^E(\mathbf{r})$, which uses the absolute barrier heights Φ^\ddagger , reflects the underlying structure of the potential

energy hypersurface Φ . Consequently, there is a relation between the topography of the PES hypersurface and the structure of its energy-parametrized disconnectivity graph $G^E(\Phi)$. The feature of the graph that is sensitive to differences between particular topologies is the multiplicity m associated with each vertex; i.e., the number of edges going out of each vertex.

A. The metric

Before using the disconnectivity graph $G(\Phi)$ to classify PES topographies, it is necessary to define the *metric* that will be imposed on the abstract graph structure. The choice of metric is essential, since the graph $G(\Phi)$ does not represent a metric-independent topology but rather a specific reducible basis of such a topology. It is not unique and therefore depends on the metric used to generate it.

The metric used here has two features:

(1) It defines the discrete spacing between the levels of the graph, i.e., sets the intervals of the energy or temperature parameter with respect to which successive mappings are performed. The resulting set of parameters can be linear, logarithmic or obey other functional forms. This setting of level spacing is essential and may significantly affect the resulting graph $G^E(\Phi)$ or $G^T(\Phi)$. In the extreme case of an infinite level spacing [i.e., only $E=0$ and $E=\infty$ ($T=0$ and $T=\infty$) are used], the graph has the trivial structure of edges connecting a single root (the full configuration space) to the multitude of local minima. It has no more information than the direct minimization map $M^d(\mathbf{r})$. In the other extreme, where the parameter spacing is infinitesimally small, each individual barrier will be represented as a vertex and the notions of super basins and long range topographical features will be lost. The graph $G(\Phi)$ that is generated in this limit is identical to the local connectivity tree mentioned in Sec. III above. It is a graph where all local barriers are explicitly marked and every vertex has exactly two edges coming out of it (like an *irregular* $k=3$ Bethe lattice).

(2) The second element of the metric, which is employed only for $G^E(\Phi)$, is that the "dead branches" of the graph (i.e., edges that do not split any more) are drawn only down to the energy of their lowest minima. In principle, the graph itself could be drawn with all dead branches going down to the baseline. The elimination of the dead branches, although not essential, helps in visualizing the energy spectrum of the system.

B. Classification of PES

We now use the $G^E(\Phi)$ disconnectivity graph to classify some basic potential energy surface topographies. This classification is summarized in Table I. We assume that the metric used has a level spacing which reveals the internal structure of the surface. In real physical systems one expects the relevant energy spacing to be on the order of kT , i.e., on the order of 0.5–2 kcal/mol at room temperature. Naturally, in a large multidimensional system a combination of several of these elementary topographic features will be needed for a complete description of the surface.

TABLE I. Classification of basic PES structures in terms of their $G^E(\Phi)$ graphs (in terms of the microcanonical ensemble).

Φ	$G^E(\Phi)$	m
Potential energy surface	Disconnectivity graph	Vertex multiplicity
Rough	An ever branching graph.	$m > 1$ for all E
Single minimum (with weak fluctuations)	A single nonsplitting branch (some splitting at lower end)	$m = 1$ for $E > 0$ ($m > 1$ for $E \rightarrow 0$)
Funnel	A single splitting branch with many dead branches splitting from it.	Main branch: $\langle m \rangle > 1$ for all E Other branches: $\langle m \rangle \geq 1$ at high E $m = 1$ at low E

1. A rough PES

Many descriptions of the PES for a protein-forming polypeptide chain have assumed that it is "rough."² A model of such a rough PES is the fractal structure of a Koch curve;³⁵ i.e., each basin can be divided into k equal sub-basins, and so on, at all levels. For this simple case the multiplicity m of the vertices of the graph is fixed; i.e., $m = k = \text{constant}$ for all vertices. The corresponding graph with $k = 3$ is shown in Fig. 2 along with a one-dimensional schematic cut through the $3N - 6$ dimensional Φ hypersurface corresponding to it. In this case the graph is identical to a Bethe lattice of rank k .³⁵ In general, rough potential energy surfaces (though not necessarily fractal in form) are expected to characterize glassy systems.

2. A single minimum PES with small fluctuations

Another possible shape for a PES is a surface with a dominant single minimum that may have some small fluctuations in the neighborhood of the minimum. On such a surface a decrease in the energy E reduces the size of the "accessible" surface Γ but does not branch the $G^E(\Phi)$ graph; i.e., $m = 1$ for all vertices. Only at $E \rightarrow 0$ the weak variations that are expected to be present in complex systems like proteins,^{14,15} may become significant and the single descending line will split at its lowest end into many small branches. In this case the multiplicity is $m = 1$ down to

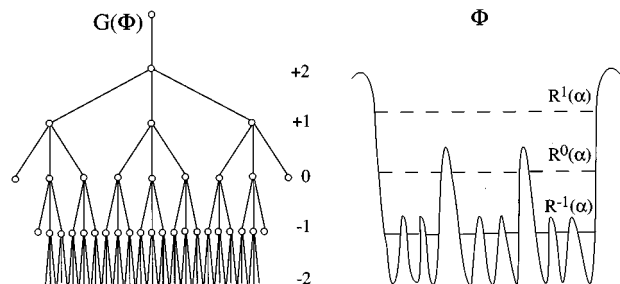


FIG. 2. A schematic one-dimensional partial cross section through a rough $3N$ -dimensional potential energy hypersurface Φ (triadic Koch curve). The horizontal lines on the PES Φ correspond to the energies used by the map $M^a(\mathbf{r}; E)$ to define the $R(\alpha)$ basins. These energies determine the corresponding levels on the disconnectivity graph $G^E(\Phi)$.

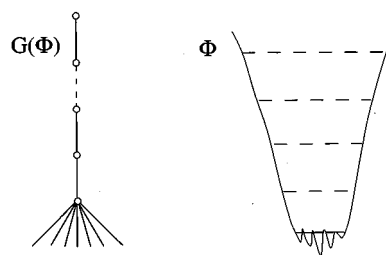


FIG. 3. A schematic representation of the PES Φ and corresponding $G^E(\Phi)$ graph for a single minima with weak noise.

$E \rightarrow 0 + \epsilon$ where $m > 1$ (in fact one expects $m \gg 1$). This PES topography and the corresponding $G^E(\Phi)$ graph are shown in Fig. 3.

3. A funnel shaped PES

The "funnel" structure, which has been postulated to be important in protein folding,^{26,27} is related to the single minimum potential; the role of such funnels in protein folding has yet to be confirmed.³⁶ In both the funnel and the simple minimum potential there is a deep minimum flanked by weaker local minima. The main difference between the two topographies is that in the funnel the flanking local minima exist over a wide range of energies, and therefore appear on all levels of the graph. As in the rough surface, each decrease in the mapping parameter can split the basins and branch the graph, but in a funnel most of these branches are expected to "die out" quickly; i.e., each branch reaches a local minimum and appear in the $G^E(\Phi)$ graph as an unbranched descending line. Ideally, only one branch, which corresponds to the core of the funnel, continues to branch down to $E = 0$. The graph of this topography, is therefore characterized by a single splitting branch which has an average multiplicity larger than 1 ($\langle m \rangle > 1$) over the whole temperature range; an average $\langle m \rangle$ is used because there may be vertices where no branching occurs due to the irregularity of the surface, although the overall feature is of a branching branch. Other branches may start with $m \geq 1$ but then collapse to a $m = 1$ branch. Figure 4 presents a schematic picture of the funnel potential energy surface and its corresponding $G^E(\Phi)$ graph.

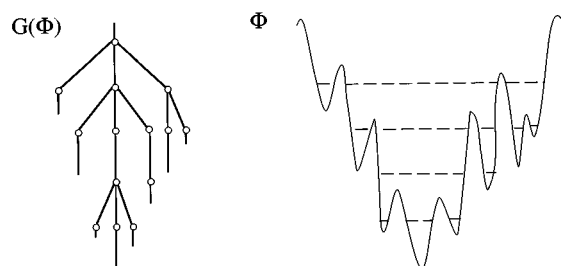


FIG. 4. A schematic representation of a funnel potential energy surface Φ and its corresponding $G^E(\Phi)$ graph.

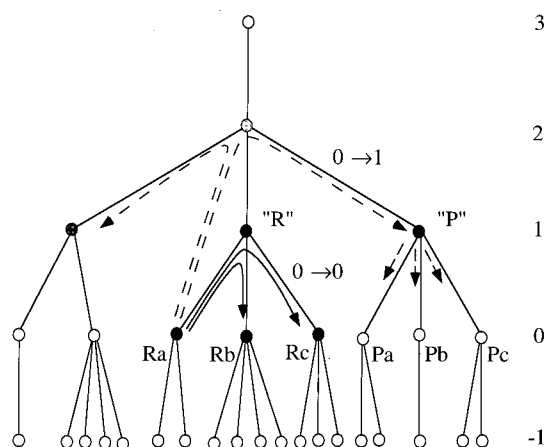


FIG. 5. A schematic representation of kinetic transitions on a canonical disconnectivity graph $G^T(\Phi)$. $0 \rightarrow 0$ transitions are denoted by solid arrows, $0 \rightarrow 1$ transitions are denoted by dashed arrows.

V. THE RELATION BETWEEN PHYSICAL PHENOMENA AND THE $G^T(\Phi)$ GRAPH

While the microcanonical (energy dependent) disconnectivity graph $G^E(\Phi)$ can be used to classify the shape and the topography of potential energy surfaces, the canonical (temperature dependent) disconnectivity graph $G^T(\Phi)$ is useful for describing the kinetics, both qualitatively and quantitatively. In this section we present a qualitative description of the kinetics of the system in terms of the canonical disconnectivity graph $G^T(\Phi)$ and the super basin mapping procedure $M^T(r)$. In Sec. VI we present a quantitative treatment of the kinetics based on the master equation. In general this graph can also be used to discuss and describe other physical phenomena such as glass transitions and freezing.

The disconnectivity graph $G^T(\Phi)$ at a temperature T partitions the PES into a set of $R^T(\alpha')$ basins. This offers a simple representation that can be used in a qualitative description of the kinetics of a complex system. We define a "transition" as motion along a path that crosses a barrier *higher* than the thermal energy. Since local states incorporated in a given basin are separated by barriers lower than kT , there are no formal "transitions" between these states, which are in rapid equilibrium. Therefore, the initial and final states in any kinetic transition are, in fact, not pure "states" (local minima) but collections of local minima that belong to the same basin. Namely, basins, and not minima, the elements that define the kinetics of the system. Basins that are separated by low barriers have a high probability for transition between them and basins that are separated by high barriers have low transition probabilities.

The disconnectivity graph $G^T(\Phi)$ reflects this type of kinetic connectivity (Fig. 5). In fact, the levels of the graph reflects a *kinetic* hierarchy. As described in Sec. III, we number the levels of the graph relative to the system temperature, which is assigned the index 0. All nodes on the 0-level of the graph represent actual super-basins at this temperature. These are denoted as Ra, Rb, Pa etc., and in many cases are connected to several sub-basins, which in turn are in rapid

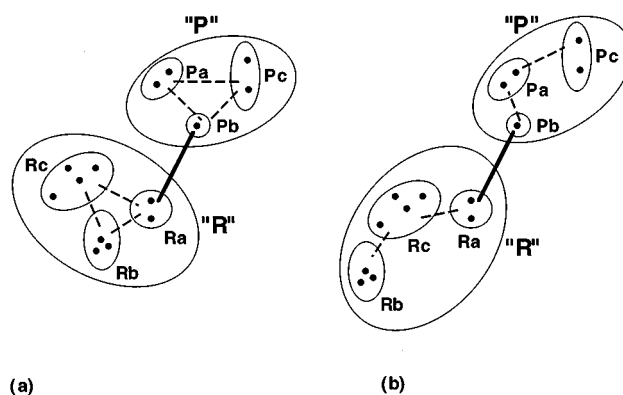


FIG. 6. A "basin connectivity graph" schematically representing the connectivity among the basins defined in Fig. 5: (a) maximal connectivity within the super basins, (b) minimal connectivity within the super basins. Dots indicate local minima, dashed lines are direct barriers between basins, and the solid line is the level-2 barrier connecting super basin R and super basin P .

equilibrium. The transition rate between the 0-level vertices of $G^T(\Phi)$ (i.e., super basins on the PES) is related to the location of their *least upper bound* (or *join*) on the graph, which in turn is proportional to the height of the lowest barrier separating the two basins. The higher the *join* (i.e., the higher the barrier), the slower the transition. Figure 5 illustrates kinetic transitions on the disconnectivity graph. Transitions between 0-level basins with a join at level 1 [denoted as $0 \rightarrow (1) \rightarrow 0$ transitions] are expected to be faster than those between basins with a join at level 2 [$0 \rightarrow (2) \rightarrow 1$ transitions].

Another feature that influences the rate of transitions, especially if the kinetic hierarchy is not strongly "separated" (i.e., the energy differences between levels are not large), is the connectivity among the different basins. A given transition probability depends not only on the barrier height but also on the probability of reaching the vicinity of that barrier from different parts of the surface. In case of a transition that crosses a level-2 barrier, it means that the internal connectivity within the level-1 super basins can effect the rate. We use the $G(\Phi)$ graph (Fig. 5) to illustrate this point. Two $R^0(\alpha')$ basins within a given $R^1(\alpha')$ super basin are *directly connected* if the path connecting their two respective α' minima does not go through any other $R^0(\alpha')$ basin. A limiting case is that of maximal connectivity within a super basin, in which all possible pairs of basins $R^0(\alpha')$ that belong to the same "parent" $R^1(\alpha')$ super basin are directly connected. For example, in the system shown in Fig. 5 the maximal connectivity limit means that there is a direct path connecting each of the three possible basin pairs: Ra–Rb, Ra–Rc, and Rb–Rc, that belong to the "R" super basin. This is schematically represented in the "basin connectivity graph" of Fig. 6(a), which shows the direct connections between basins. In this case, all the level-0 basins (Ra,Rb,Rc) are not more than one basin away from the level-2 barrier, which is located between Ra and Pb. If the internal connectivity within super-basin R is less than maximal [as illus-

trated in Fig. 6(b)] some basins are significantly further away from the level-2 barrier.

It is useful to distinguish between a strong kinetic hierarchy and a weak kinetic hierarchy:

(i) If there is a strong hierarchy, reflected in a large energy separation between the levels of the graph, there is a time separation between $0 \rightarrow (1) \rightarrow 0$ transitions, $0 \rightarrow (2) \rightarrow 1$ transitions etc. (Fig. 5), which is essentially independent of the multiplicity of the connectivity. Before each $0 \rightarrow (2) \rightarrow 1$ transition there is sufficient time for equilibration among the 0-level basins on the “reactant” side, and the rate of the $0 \rightarrow (2) \rightarrow 1$ transition will not depend on the 0-level connectivity (within the level-1 reactant basin). This type of time separation between $0 \rightarrow 0$ transitions and (at least) $0 \rightarrow 1$ transitions was observed in simulations of the protein bovine pancreatic trypsin inhibitor (BPTI),¹⁵ and corresponds to the tier picture suggested by Frauenfelder and co-workers for transitions in myoglobin.^{20–22}

(ii) If the kinetic hierarchy is weak, as reflected in a small energy separation between levels on the graph, there will not be a well-defined time separation between transitions relating to the different levels. This means that internal equilibration within the “reactant” super basin is not achieved before the higher order kinetic transitions take place, and the reaction rate will depend on specific connectivity among the reactant 0-level basins. A graph of basin connectivity (as in Fig. 6) shows the internal structure of the basin and helps in locating entropic contributions to the kinetics.

In a similar way, the connectivity within the “product” $R^1(\alpha')$ basin influences the *short-time nonequilibrium* distribution among final states (the final equilibrium distribution is insensitive to this). The $G^T(\Phi)$ graph shows that a $0 \rightarrow (2) \rightarrow 1$ transition starts at a level-0 basin (within the reactant super basin), crosses a level-2 barrier (the *join*) and ends up in the level-1 “product” super basin. Although the system finally ends up with an equilibrium distribution among the various level-0 product basins (Pa,Pb,Pc), the $G^T(\Phi)$ graph does not reveal the kinetics of this distribution, especially on short time-scales. An insight into this process requires the “basin connectivity graph” described in Fig. 6.

VI. “BASIN-TO-BASIN” KINETICS

The basin picture of the potential energy surface focuses on the basins, and not the energy minima, as the elementary entities that govern the behavior of the system. It highlights the role of groups of states that are in rapid equilibrium at a given temperature. The focus on *basin to basin* kinetics contrasts with the *state to state* kinetics often employed in gas phase and molecular beam reactions⁶ or the complete averaging over initial and final states in solution reactions. In this section we explore quantitatively the basin-to-basin kinetics that was discussed qualitatively in the previous section. We analyze the transition matrix for the system and introduce a new representation of the master equation that reveals the basin to basin kinetics by summing over the states that belong to any given basin (see Sec. VI C).

A. The master equation

The master equation is a loss–gain equation which describes the time evolution of the probability $p_i(t)$ for finding the system in a state i .¹⁶ The basic form of this equation is

$$\frac{dp_i(t)}{dt} = \sum_j [W_{ij}p_j(t) - W_{ji}p_i(t)], \quad (13)$$

where W_{ij} is the transition probability from state j to state i . Equation (13) can be rewritten in matrix form by defining the transition matrix elements as

$$W_{ij} = W_{ij} - \delta_{ij} \left(\sum_k W_{ki} \right). \quad (14)$$

The matrix \mathbf{W} has the properties that $W_{ij} \geq 0$ for $i \neq j$ and that the sum over each column is zero; i.e., $\sum_i W_{ij} = 0$ for all j . This last property is required for a closed system so that the flux out of any given state remains within the system (i.e., goes into the other states of the system). In matrix form Eq. (13) becomes

$$\dot{\mathbf{p}}(t) = \mathbf{W}\mathbf{p}(t), \quad (15)$$

which has the formal solution $\mathbf{p}(t) = e^{t\mathbf{W}}\mathbf{p}(0)$; where $\mathbf{p}(t)$ is the probability vector at time t .

B. The transition matrix

We now consider the structure of the transition matrix \mathbf{W} in the basin description given in Sec. II. For the peptide and protein systems of primary interest here, each state i corresponds to one of the potential energy minima α . In such systems the W_{ij} transition probabilities are the state to state rate constants, k_{ij} , for a transition $j \rightarrow i$. Given a knowledge of the minima and transition states it would be possible to use transition state theory, for example, to evaluate the rate constants.³⁷ In this case, we have

$$k_{ij} = W_{ij} = \frac{kT}{h} \frac{Q_{ij}^\ddagger}{Q_j} \exp(-E_{ij}/kT), \quad (16)$$

where k is the Boltzmann constant, h is the Planck constant, Q_j is the partition functions of the “reactant” state, Q_{ij}^\ddagger is the partition function of the transition state and E_{ij} is the barrier height measured relative to state j . Depending on the nature of the environment and the interest in more quantitative kinetics, a stochastic description (Kramers theory) or more accurate rate theory should be used instead of Eq. (16).³⁸

In the usual case of nondegenerate states

$$E_{ij} \neq E_{ji} \quad \text{for } E_i \neq E_j, \quad (17)$$

so that the transition matrix \mathbf{W} is not symmetric, with $W_{ij} > W_{ji}$ in most cases if $E_i < E_j$. The diagonal matrix elements are assigned the value $W_{ii} = -\sum_{i \neq j} W_{ij}$ to fulfill the requirement that $\sum_i W_{ij} = 0$ for all j .

From the formulation of the potential energy surface in terms of temperature dependent basins, $R^T(\alpha)$, (Sec. II) all barriers that connect minima that belong to the same basin are lower than kT . Therefore, transitions within such a basin have a high probability (relative to inter-basin transitions) and there is rapid equilibration. As a result the physically interesting kinetics of such systems is best described in terms of *basin to basin* transitions and not *state to state* transitions. To introduce this behavior into the master equation, we rewrite the transition matrix \mathbf{W} in block form

$$\mathbf{W} = \begin{pmatrix} W_{AA} & W_{AB} & W_{AC} & \cdots & W_{AT} \\ W_{BA} & W_{BB} & W_{BC} & \cdots & W_{BT} \\ W_{CA} & W_{CB} & W_{CC} & \cdots & W_{CT} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_{TA} & W_{TB} & W_{TC} & \cdots & W_{TT} \end{pmatrix}, \quad (18)$$

where the W_{AA}, W_{BB}, \dots blocks along the diagonal include the transitions within a given $R^T(\alpha')$ basin while the off-diagonal blocks (e.g., $\mathbf{W}_{AB}, \mathbf{W}_{BA}$), include the transition elements between two such basins. The W_{AA}, W_{BB}, \dots blocks may be rearranged, without loss of generality, so that the deepest minimum α' of the basin corresponding to block W_{AA} is deeper than the deepest minimum of the basin corresponding to block W_{BB} , etc. In the last block along the diagonal, denoted here by W_{TT} (for “transient”), we include all the high energy states; the W_{TT} block is somewhat loosely defined and is introduced for completeness. It includes all states with a very small input flux and a very large output flux; the fact that it could be further partitioned is unimportant for the present discussion.

The definition of W_{ij} [Eq. (16)] implies that *all* off-diagonal terms are nonzero. This means that, in general, \mathbf{W} cannot be decomposed into block diagonal form and the equilibrium distribution $\mathbf{p}(\infty)$ is expected to be spread over a large portion of state space. However, since states may not be connected directly (as was discussed in Sec. V), many of the off-diagonal terms in Eq. (18) are expected to be very small. At short and intermediate time scales these terms can often be neglected leading to a simpler description of the kinetics of the system.

C. Basin population probabilities and master equation

To focus on the basin populations and inter-basin transitions, it is useful to transform the master equation [Eq. (15)]. Instead of expressing the probability vector $\mathbf{p}(t)$ on a state-by-state basis, Eq. (15) can be rewritten by introducing basin probabilities in the form

$$\mathbf{p}(t) = \begin{pmatrix} \vdots \\ p_i(t) \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{P}_A(t) \\ \mathbf{P}_B(t) \\ \mathbf{P}_C(t) \\ \vdots \\ \mathbf{P}_T(t) \end{pmatrix}, \quad (19)$$

where $\mathbf{P}_A(t), \mathbf{P}_B(t)$ etc. are vectors which incorporate all the states that belong to a given basin **A**, **B**, etc.; i.e., each of these vectors are defined

$$\mathbf{P}_I(t) = \begin{pmatrix} \vdots \\ p_i(t) \\ \vdots \end{pmatrix}, \quad i \in I. \quad (20)$$

With this definition of basin vectors, the master equation [Eq. (13)] can be rewritten in terms of these $\mathbf{P}_I(t)$ as

$$\frac{d\mathbf{P}_I(t)}{dt} = \sum_J [\tilde{\mathbf{W}}_{IJ}\mathbf{P}_J(t) - \tilde{\mathbf{W}}_{JI}\mathbf{P}_I(t)] \quad (21)$$

or

$$\dot{\tilde{\mathbf{P}}} = \tilde{\mathbf{W}}\tilde{\mathbf{P}}, \quad (22)$$

where the $\tilde{\mathbf{W}}$ matrix is in the basin-block form introduced in Sec. VI B. The new equations, Eqs. (21) and (22), which are cast in terms of conformation basins, are a compact and focused form of the master equation [Eq. (13)]. In addition, we can define a new probability vector $\mathbf{P}'(t)$ which gives the accumulated probability of finding the system in a given basin, rather than in a given state. The scalar elements $P'_I(t)$ of $\mathbf{P}'(t)$ are defined as the sum over the probabilities of the individual states in basin I ,

$$P'_I(t) = \sum_{i \in I} p_i(t). \quad (23)$$

Since $\mathbf{p}(t)$ is normalized, $\mathbf{P}'(t)$ is also normalized; i.e.,

$$\sum_I P'_I(t) = \sum_I \sum_{i \in I} p_i(t) = 1. \quad (24)$$

The time evolution of this probability vector is given by the transition matrix in the block form of Eq. (18). In Sec. VII D, we examine the transition matrix \mathbf{W} and the time evolution of the basin probability distribution $\mathbf{P}'(t)$ for the specific case of the tetrapeptide IAN.

VII. AN EXAMPLE: THE IAN TETRAPEPTIDE

The topological analysis in Sec. II C and subsequent discussions are based on the M^T and M^E maps of the multidimensional potential energy surface. Constructing a full M^T map, is a time-consuming task for a complex system. At each of the temperatures it is necessary to generate a long

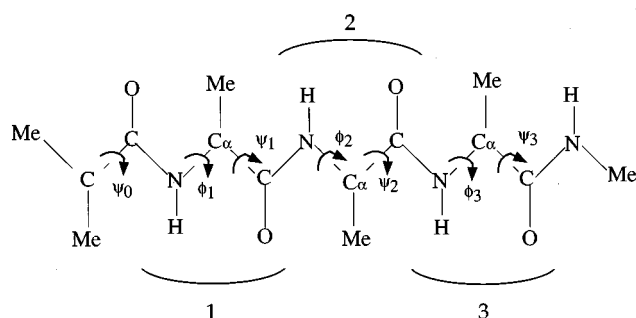


FIG. 7. The tetrapeptide IAN [isobutyryl-(ala)₃-NH-methyl]. The soft torsions (ϕ, ψ) are on each side of the C(*i*) carbons.

dynamics trajectory that samples the conformation space which is accessible at that temperature. It is then necessary to perform “finite time” simulated annealing for many conformations along the trajectory. Because of the difficulty of obtaining complete results by this approach even for relatively small systems (e.g., rare gas clusters of 15 or more atoms, peptides of 10 or more residues) approaches based on statistical sampling have been introduced.¹³ However, if the system is simple enough, it is possible to determine the minima and local barriers directly and to construct the disconnectivity graphs $G(\Phi)$ from them. A molecule for which the necessary information about the minima and barriers is available is the tetrapeptide isobutyryl-(ala)₃-NH-methyl (IAN) in vacuum (Fig. 7), which was studied by Czerninski and Elber.^{17,18} This peptide, which is a derivative of tetra-alanine, is the simplest model system that can form a full α -helical turn, including the stabilizing hydrogen bond between the first and the fourth amide plane.

In the polar hydrogen representation,³⁹ with methyl groups treated as single extended carbon atoms, IAN has a 72-dimensional conformation space, excluding overall translation and rotation. However, the conformational transitions of the peptide are well described in terms of a much smaller coordinate manifold. It consists of the seven soft torsions, corresponding to the dihedral angles ϕ and ψ (Fig. 7). Czerninski and Elber^{17,18} located 139 local minima and 502 barriers on the potential energy surface of this tetrapeptide. With this information they looked at the barrier spectrum and found a possible pathway from an “ α -helix” conformation to an extended “ β -sheet” conformation. They also calculated the relaxation times for the system using the master equation formulation.¹⁸ We use the data for IAN, which they kindly made available to us,⁴⁰ to construct the disconnectivity graph and investigate the topography of this potential energy surface. We then examine its basin-to-basin kinetics. The present results provide a viewpoint that complements the work of Czerninski and Elber and present a clear illustration of the utility of the formulation.

A. Potential energy surface topography and disconnectivity graphs

The search algorithm used by Czerninski and Elber started with the minimum energy path between two configura-

tions resembling an α -helix and a β -sheet, respectively, and then recursively determined alternative paths between any two minima encountered during the search.^{17,18} Each path calculation involved several refinement steps and the search was ended when no new minima were found. The resulting 139 minima are a good representation of the important portions of the vacuum potential energy surface, although some high energy minima are probably missing. Since the search was biased towards the α -helix and β -sheet like conformation, it is possible¹⁸ that some structures were missed. The 139 minima were connected by 502 direct transition states, i.e., barriers that connect two minima without intermediates. Of the 502 barriers, only 393 were barriers that define connectivity between different minima. The other 109 barriers were either self-connecting paths [i.e., paths that starts at a given minimum, crosses a saddle point and returns on the other side to the same minimum (20 barriers)] or correspond to higher-lying saddle points between already connected minima (89 barriers). When two minima were connected by more than one path, we used the lowest transition state to describe the connectivity and the basin kinetics.

Figures 8 shows the energy disconnectivity graph $G^E(\Phi)$ for the tetrapeptide IAN on different scales. The energy is relative to the global minima ($E=0$), and the point where the branches terminate are the exact energies of the corresponding minima (Sec. IV A) (except for some high lying minima). In Fig. 8(a) the levels of the graph are separated by $\Delta E=1$ kcal/mol. Since each vertex of the graph corresponds to a basin on the PES (Sec. III), the vertices at level $E=6$ kcal/mol represent basins which include minima connected by barriers no higher than 6 kcal/mol, while vertices at level $E=5$ kcal/mol represent basins with barriers not higher than 5 kcal/mol. Figures 8(b) and 8(c) are the corresponding $G^E(\Phi)$ graph with different level spacing; Fig. 8(b) is on the scale of $\Delta E=0.5$ kcal/mol and Fig. 8(c) is on the scale of $\Delta E=2$ kcal/mol. While all three graphs correspond to the same system, they show that the impression obtained from the graph depends on the choice of energy scale. In particular, Fig. 8(b) shows that, although the funnel structure can still be recognized even at higher resolution, the more detailed branching picture makes it harder to identify global features. The more coarse graph in Fig. 8(c) shows the funnel, but misses some of the interesting features in Fig. 8(a), as described below. Thus, it is important to use several different scales in constructing disconnectivity graphs to obtain a full understanding of a given system.

The graph in Fig. 8(a) has one main central branch (vertices A–B–C–D); i.e., the overall structure of the potential energy surface is that of a multidimensional “funnel” (see Sec. IV B and Fig. 4). There is only one relatively large side branch indicated in Fig. 8(a) as vertex E. It is significantly smaller than the main branch and ends at an energy above the global minimum. All 10 lowest minima and 14 out of the 15 lowest minima are inside the main funnel. The 20 lowest minima are numbered in Fig. 8(a) and their energies are given in Table II.

As the energy E of the $M^E(\mathbf{r})$ map decreases, the “size”

of the basin, which is related to the configuration space volume, of the main funnel also decreases. The size of a basin can be described by the number of minima incorporated within it [see Eq. (12) and Sec. III above]. As we move down from higher energies towards vertex A [using the notation in Fig. 8(a)], we find that initially only few high lying minima become disconnected from the main basin, and that no significant splitting is observed. Vertex A, defined at the $E=7$ kcal/mol level of the graph, corresponds to a basin of size 103; i.e., 103 out of the 139 minima are incorporated within it (and connected by barriers lower than 7 kcal/mol). At one level down ($E=6$ kcal/mol), the A basin has split into three sub-basins B, E, and F plus many smaller basin "fragments" which include only a single minimum. Of the three main sub-basins, vertex B incorporates 55 minima, vertex E incorporates 17 minima and vertex F has only 3 minima (a ratio of 1:0.31:0.05). The structure under vertex E is not that of a funnel, but rather a region on the PES where a group of minima of similar energy are separated by relatively high barriers. As the energy is reduced to $E=5$ kcal/mol, the main A-B-C branch shrinks further in volume and incorporates 39 minima at vertex C, with only very minor side branches.

Significant splitting occurs at the next level ($E=4$ kcal/mol), where three identifiable sub-basins D, G, and H, are formed. The central basin D, which leads down to the bottom of the funnel, has 14 minima, including 5 of the lowest 10 minima. Basin G incorporates 8 minima (including 4 of the lowest 10 minima) and basin H has only 2 minima including the extended " β strand" structure. From this level down it is hard to identify a central funnel as multiple branching occurs, with two different branches containing the lowest and next to lowest minimum.

In Fig. 8(b) with an energy spacing of 0.5 kcal/mol, all the branches are shown down to their lowest minima [as in Fig. 8(a)]. At this "finer" resolution the splitting of vertex A to three sub-basing is preserved; at $E=6.5$ kcal/mol, there are 63 minima under vertex B, 21 under vertex E and 7 under vertex F [a ration of 1:0.33:0.11, similar to that obtained from Fig. 8(a)]. Note, that the additional minima, now incorporated in basin F, were classified in Fig. 8(a) under basin A. Of interest in this more detailed graph is that the basin which corresponds to vertex H is larger than shown in Fig. 8(a); i.e., it includes more local minima than appear in Fig. 8(a). This basin splits from the main branch of the funnel at

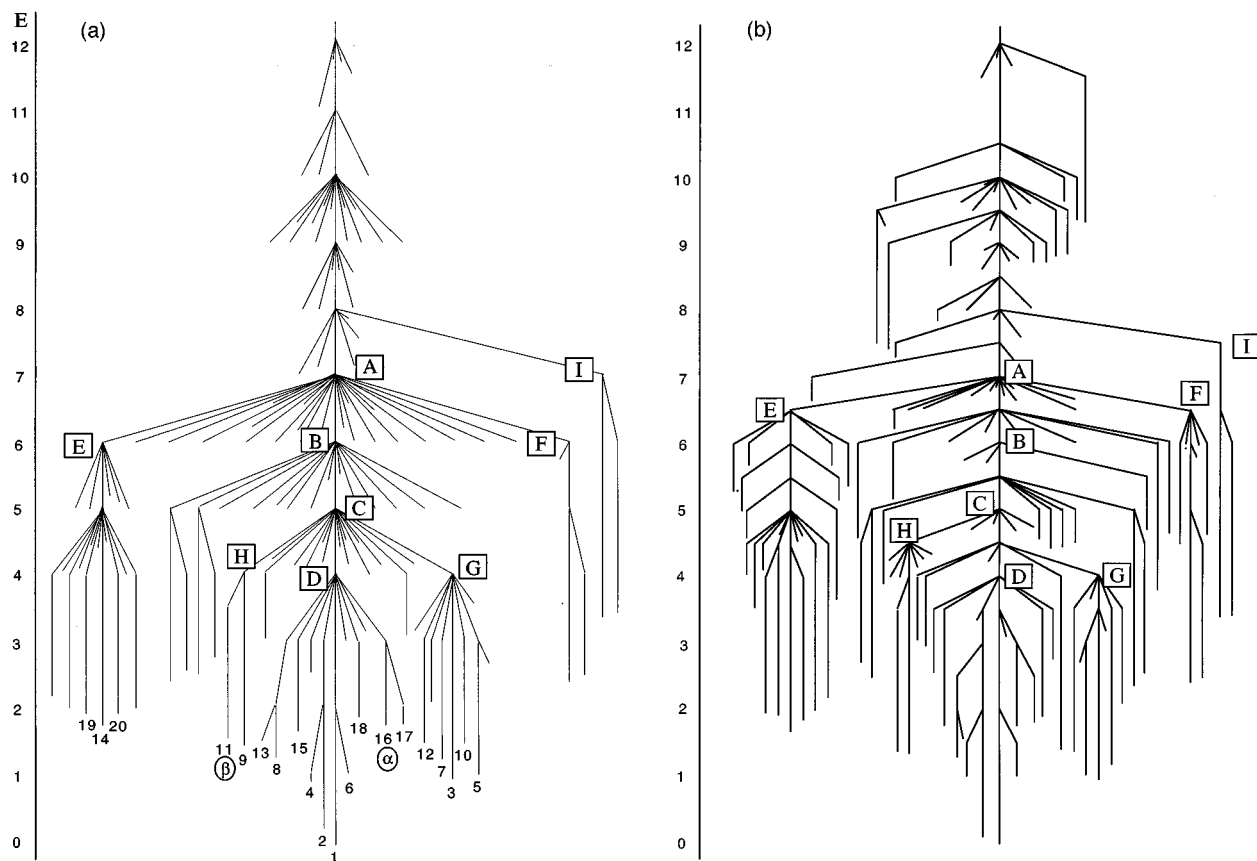


FIG. 8. Disconnectivity graphs $G^E(\Phi)$ of IAN: (a) mapped using a level spacing of $\Delta E=1.0$ kcal/mol. The overall structure approximates that of a funnel, though there are additional features outside the funnel. The numbers indicate the ordering of the 20 lowest minima on the potential energy surface. The energy scale is set relative to the lowest minima, i.e., the energy of minima #1 is 0 kcal/mol and the energy of minima #20 is 1.902 kcal/mol. For clarity, branches leading to high lying minima were drawn only down to the next level of the graph. Other branches were drawn down to their respective minima. Low lying minima were drawn down to their respective minimum energy. The letters A-I indicate vertices of interest. The number of minima below each of these vertices are: A-109; B-55; C-39; D-14; E-17; F-3; G-8; H-2; I-2. The symbols α and β indicate minima closest to the α -helix and β -sheet conformations, respectively. (b) The same with $\Delta E=0.5$ kcal/mol, (c) the same with $\Delta E=2.0$ kcal/mol.

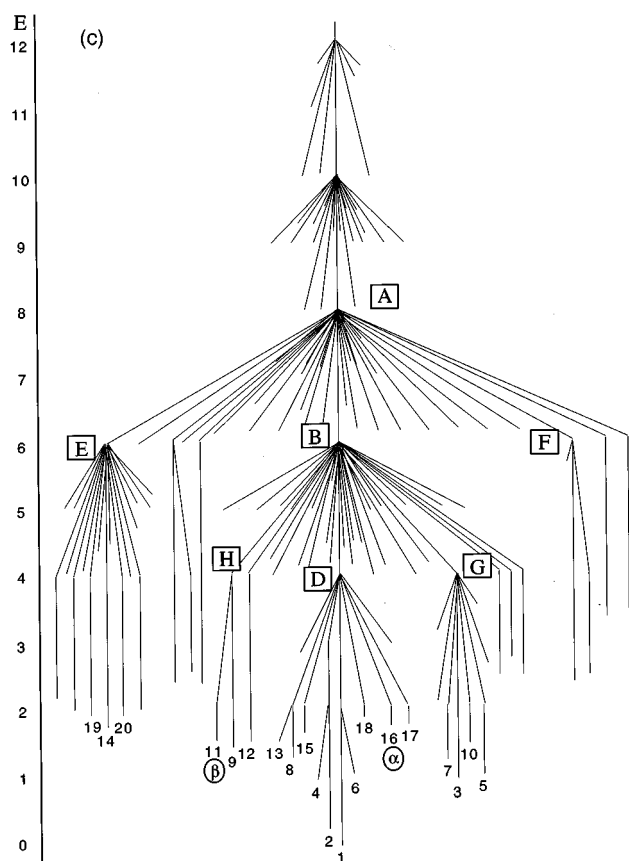


FIG. 8. (Continued.)

$E=4.5$ kcal/mol and carries with it a cluster of 8 minima. The “additional” 6 minima, with respect to Fig. 8(a), are relatively high in energy; they rank between number 36 and 52 and were included in vertex C in Fig. 8(a).

As discussed above (Sec. III), a continuous mapping with infinitesimal level spacing ($\Delta E \rightarrow 0$) will result in a picture of the local connectivity. Such a continuous mapping of IAN for a small portion of the PES can be found in Ref. 18.

B. Conformations and basins

An important question is the extent to which connected configurations are near each other in conformation space, and whether the similarity decreases as the location of the *join* increases in energy. In other words, how similar are conformations that according to the $G(\Phi)$ graph belong to the same basin. We use the $G(\Phi)$ graph in Fig. 8(a) and analyze the structures of the 45 lowest minima, plus several of the high energy minima. To do the analysis we make use of the seven mainchain dihedral angles of the tetrapeptide shown in Fig. 7. They are the Φ and Ψ angles of the three alanines and Ψ_0 , the dihedral angle of the *N* terminal $C(\text{CH}_3)$ group. Table II lists the energies and dihedral angles of the 45 lowest minima of IAN grouped according to the basins depicted in Fig. 8(a). The three high energy minima, plus minimum 44, are not associated with any of these basins. In particular, minimum 44 branches off above basin B

(in basin A) and the other three branch off at the very top of the graph. Table III summarizes the data in terms of characteristic angles for each basin. From the data one sees that, while some dihedral angles have essentially unique values in a given basin (e.g., the central angle Φ_2 , which has the range $-70^\circ \pm 20^\circ$ in basin D and $58^\circ \pm 7^\circ$ in G), others exhibit a number of alternative values.

The central funnel beginning with basin B and extending downward in energy, includes a wide range of dihedral angle values. The angle Φ_3 is negative for all structures and mainly in the neighborhood of $-70^\circ \pm 20^\circ$, although basin D shows a wider range of values. For Φ_3 , the basins (E,F,I) outside the funnel all have Φ_3 values of $60^\circ \pm 2^\circ$.

The split of the funnel (basins B and C) into its two main components, basins D and G, is defined mainly by the two “central” dihedrals Ψ_1 and Φ_2 . Their values lead to a bend that is common to basins D and G [see Fig. 9(a) and 9(b)]. In D this bend is similar to a type I turn ($\Phi_2 \approx -60^\circ$, $\Psi_2 \approx -30^\circ$ and $\Phi_3 \approx -90^\circ$, with Ψ_3 variable for the lowest energy structures). The *N*-terminal dihedrals (Φ_1, Ψ_1) seem less significant. Structure 16, which is described as a standard α -helix by Czerminski and Elber deviates significantly from an α -helix, in terms of its dihedral angles and its hydrogen bond (1–3 rather than 1–4).

Minima #34 and #36 (grouped under “C (not in D, G, H)”, which are the conformations that form a bridge between basin D and basin G (Fig. 10(c)), correspond to a mixture of the two basin structures. They are similar to the D basin in two dihedrals (Φ_2 and Ψ_2) and to basin G in two other dihedrals (Φ_1 and Ψ_1); Φ_3 is common to both basins and the terminal dihedrals are less specific. Similar behavior is found in the minima grouped under “B [not in C]”. Here dihedrals Φ_1 and Ψ_1 are similar to basin D while dihedrals Φ_2 and Ψ_2 are like basin G.

Basin H, which also is part of the central funnel, exhibits an extended conformation (β -strand type) very different from the “bent” conformations of basins D and C [see Figs. 9(a) and 9(c)]. It has been noted recently⁴¹ that the minima of such extended strands are near $\Phi \approx 70^\circ$, $\Psi \approx 75^\circ$, instead of the “ideal” values in the neighborhood of $\Phi = -140^\circ$, $\Psi = 140^\circ$. The conformations of basin H differ from those of basin G by only two dihedrals; they are $\Phi_2 = -70^\circ$ and $\Psi_2 = 70^\circ$ in basin H, vs $\Phi_2 = 55^\circ - 65^\circ$ and $\Psi_2 = -50^\circ - -70^\circ$ in basin G. From Fig. 8(b) one sees that basin H splits from the central funnel at a higher level than basins D and G, indicating that a higher barrier is associated with the change of dihedral angles.

Although the nonfunnel basins share the same C-terminal dihedral angles (Φ_3 and Ψ_3), they differ in the other angles. An interesting characteristic of basin E is that while all its minima share angles Φ_2 , Φ_3 , and Ψ_3 ($-70^\circ - -90^\circ$; 60° ; -60° , respectively), they can have one of two values for Ψ_2 (70° and -20°). This results in conformations that look very different from each other [see, for example, Figs. 9(d) and 9(e)].

The energetic effect of the *N*-terminal dihedral Ψ_0 is very small. Its variation consistently leads to changes of about 0.05–0.10 kcal/mol for pairs of minima such as those

TABLE II. The energies and dihedral angles of the lowest 45 minima of IAN grouped together according to their basin assignment. A few high lying minima are added for comparison. The energies, in kcal/mol, are given relative to the lowest minima on the surface.

No.	Energy	Ψ_0	Φ_1	Ψ_1	Φ_2	Ψ_2	Φ_3	Ψ_3
Basin D [including α -helix (#16)]								
1	0.000	-60.54	50.97	-71.95	-88.03	-31.32	-76.49	73.28
6	1.021	-58.18	44.91	-83.68	-67.18	-25.82	-94.16	12.62
2	0.109	103.73	48.93	-72.11	-93.29	-28.75	-77.33	71.62
4	0.959	109.94	41.02	-84.71	-66.82	-25.73	-99.05	14.45
8	1.291	125.10	-52.35	-32.95	-70.74	-22.19	-101.86	-32.52
13	1.546	124.70	-48.23	-34.14	-58.25	-20.63	-84.38	-6.84
15	1.748	118.00	57.84	-62.81	-75.16	70.10	-72.85	72.12
18	1.828	-59.73	58.10	-62.88	-75.21	69.92	-72.92	72.09
16	1.750	-62.53	-53.37	-31.78	-69.52	-22.31	-102.81	-32.63
17	1.776	-62.14	-49.16	-32.84	-58.16	-20.18	-83.54	-7.50
29	2.526	125.06	-51.69	-22.03	-72.90	-20.39	-77.06	67.36
33	2.773	-62.36	-52.56	-20.87	-72.29	-20.19	-76.83	67.58
35	2.958	128.58	-56.15	-48.83	-79.15	45.11	-167.12	-48.84
37	3.211	-62.83	-59.36	-50.19	-79.28	48.31	-167.12	-49.09
Basin G								
3	0.897	-63.61	-72.79	80.81	64.56	-56.64	-72.91	74.95
5	0.969	126.71	-71.21	81.63	64.92	-55.89	-72.69	74.94
7	1.161	-62.00	-73.40	78.46	55.03	-69.37	-93.85	-23.99
10	1.360	126.60	-71.42	81.35	56.59	-68.68	-92.31	-24.64
12	1.427	124.40	-56.55	124.51	63.21	-47.31	-69.99	-25.06
23	2.103	-64.62	-58.34	124.23	63.17	-47.65	-69.69	-24.76
32	2.720	120.23	-86.48	66.18	62.32	-54.11	-82.86	146.25
38	3.225	-83.00	-90.72	60.56	60.75	-55.00	-83.97	151.61
42	3.710	-56.68	-71.83	145.28	51.06	-69.00	-88.47	-19.79
Basin H [including β -strand (#11)]								
9	1.344	-61.34	-73.90	72.91	-72.14	71.31	-72.43	72.46
11	1.386	122.05	-73.50	72.56	-72.28	71.27	-72.44	72.46
Basin C (outside of sub-basins D, G, H)								
34	2.957	-61.33	-74.95	70.27	-53.39	-22.61	-77.43	-11.99
36	3.012	122.17	-74.57	69.88	-53.59	-22.50	-77.46	-11.99
45	3.863	-61.33	-74.44	72.67	-68.85	-28.09	-75.21	69.90
Basin B (outside of sub-basin C)								
26	2.414	118.58	58.64	-62.13	57.97	-60.69	-74.56	71.68
27	2.482	-59.98	58.82	-62.23	57.97	-60.69	-74.56	71.68
30	2.618	118.50	58.67	-61.25	49.73	-70.30	94.43	-21.56
31	2.683	-59.95	58.87	-61.33	49.77	-70.30	-94.48	-21.56
Basin E								
14	1.703	115.31	56.07	-67.34	-79.15	69.12	60.40	-61.14
19	1.873	-57.55	57.25	-66.47	-78.49	70.30	60.97	-60.69
20	1.902	-61.33	-73.74	73.20	-72.41	74.64	61.79	-59.42
21	1.943	122.04	-73.34	72.84	-72.56	74.59	61.78	-59.43
22	1.966	116.76	48.80	-71.88	-90.42	-23.41	59.74	-59.93
24	2.274	-57.85	50.45	-71.87	-89.03	-24.86	59.67	-59.82
41	3.544	124.84	-52.59	-22.87	-75.29	-15.19	58.86	-61.14
43	3.711	-62.41	-53.49	-21.95	-74.72	-15.55	58.87	-61.14
Basin F								
25	2.410	-61.77	-73.74	77.04	61.91	-60.21	59.26	-59.99
28	2.492	123.43	-72.97	77.29	62.06	-60.12	59.28	-59.99
Basin I								
39	3.359	118.57	58.71	-62.02	58.19	-61.27	58.96	-60.21
40	3.428	-59.97	58.89	-62.11	58.21	-61.27	58.96	-60.21
Some minima outside all of the above basins								
44	3.792	123.53	-69.67	70.38	176.43	-40.67	-104.01	-34.30
127	9.330	125.98	-67.91	-20.55	-143.31	144.46	-143.67	147.80
128	11.475	116.45	43.44	34.61	48.28	29.09	58.76	17.03
137	11.675	-72.68	-143.39	102.60	47.99	53.34	54.53	-74.75

TABLE III. Characteristic dihedral angles of the different basins (based on the lowest 41 minima specified in Table II). Angles are rounded to the closest multiple of 5. The numbers in parentheses indicate the percentage that a specific angle range appears within the basin.

Basin	Ψ_0	Φ_1	Ψ_1	Φ_2	Ψ_2	Φ_3	Ψ_3
D (14 minima)	100 to 125 (50%) -60 (50%)	40 to 60 (43%) -50 to -60 (57%)	-20 to -85 (-20 to -35) (43%) (-50 to -85) (57%)	-60 to -90	-20 to -30 (71%) 70 (14%) 45 (14%)	-70 to -100 (86%) -165 (14%)	70 (43%) 15 (14%) -10 (14%) -30 (14%) -50 (14%)
G (9 minima)	120 to 125 (44%) -55 to -65 (44%) -85 (11%)	-55 to -90	60 to 80 (67%) 125 (22%) 145 (11%)	50 to 65	-50 to -70	-70 to -90	150 (22%) 70 (22%) -20 to -25 (56%)
H (2 minima)	120 (50%) -60 (50%)	-75	70	-70	70	-70	70
C (not D,G,H) (3 minima)	120 (33%) -60 (66%)	-75	70	-50 to -70	-20 to -30	-75	-10 (66%) 70 (33%)
B (not in C) (4 minima)	120 (50%) -60 (50%)	60	-60	50 to 60	-60 to -70	-75 to -95	70 (50%) -20 (50%)
Summary of basin B (32 minima)	120 (47%) -60 (53%)	-50 to -95 (66%) 40 to 60 (33%)	-20 to -85 (60%) 60 to 80 (33%)	-55 to -90 (60%) 50 to 65 (40%)	-20 to -70 (80%) 45 to 70 (20%)	-70 to -100 (94%)	-10 to -30 (40%) 150/70/15/-50
E (8 minima)	115 to 125 (50%) -60 (50%)	50 to 55 (50%) -50 to -70 (50%)	-70 (75%) -20 (25%)	-70 to -90	70 to 75 (50%) -15 to -25 (50%)	60	-60
F (2 minima)	120 (50%) -60 (50%)	-70	80	60	-60	60	-60
I (2 minima)	118 (50%) -60 (50%)	60	-60	60	-60	60	-60

of basins H, F, and I. On the other hand, the energy barrier associated with this simple 180° dihedral angle flip (between two values: -60° and $+120^\circ$) is rather high, typically 2 to 3 kcal/mol.

The high energy minima listed in Table II are quite different from any of the low energy structures; e.g., minimum 127 has most of the dihedral angles Φ_2 , Ψ_2 , Φ_3 in the ideal β -strand region, which is higher in energy than the actual strand minimum, as described above.

The basin designated by vertex H is particularly interesting. As already mentioned, the conformations incorporated in basin H are extended. However, Fig. 8(a) indicates that this basin is part of the central funnel, which otherwise incorporates nonextended conformations [Figs. 9(a) and 9(b)]. This difference can be understood from the finer $G^E(\Phi)$ graph generated with 0.5 kcal/mol resolution [Fig. 8(b)]. At this resolution one sees that the H basin branches off from the central core of the funnel at $E=4.5$ kcal/mol, 0.5 kcal/mol above the point where the D and G branches separate. The difference between the folded D-vertex configurations and the extended H-vertex configurations shows that connectivity is not always simply related to conformational similarity.

C. Details of connectivity

The $G(\Phi)$ graph (Fig. 8) shows the splitting of conformation space into smaller basins as the total energy E decreases. It does not indicate, however, whether the minima

inside a basin are highly connected to each other due to the presence of multiple paths or whether they have only the minimal internal connectivity needed to define a basin. This difference, which is of importance for understanding the kinetics can be addressed by using the basin connectivity graph introduced in Sec. V (Fig. 6). This graph shows the direct connectivity within a given basin at a given total energy E .

Figure 10 shows graphs of the internal connectivity at different energy levels of the 39 minima of IAN that are included in basin C, which corresponds to the “center” of the funnel [Fig. 8(a)]. The graph is a projection of the minima from the $3N$ -dimensional conformation space onto a plane. Each point in this graph is one of the 39 minima belonging to vertex C (numbered according to their energy), and each edge represents a direct pathway between two minima that is lower than the energy value that defines that graph; the different symbols for the points and lines are described in the caption. The projection was arranged so that there would be as few line crossings as possible. As a result, connectivity is related to proximity. However, due to the reduction in dimensionality, line crossings cannot be avoided altogether. Further, as no special optimization was performed, the relative position of the minima give only a qualitative indication of the connectivity shown explicitly by the pathways.

Figures 10(a)–10(d) can be used to follow the creation of basins of increasing size in this region of the PES. At $E=3.5$ kcal/mol three small basins are observed [Fig. 10(a)],

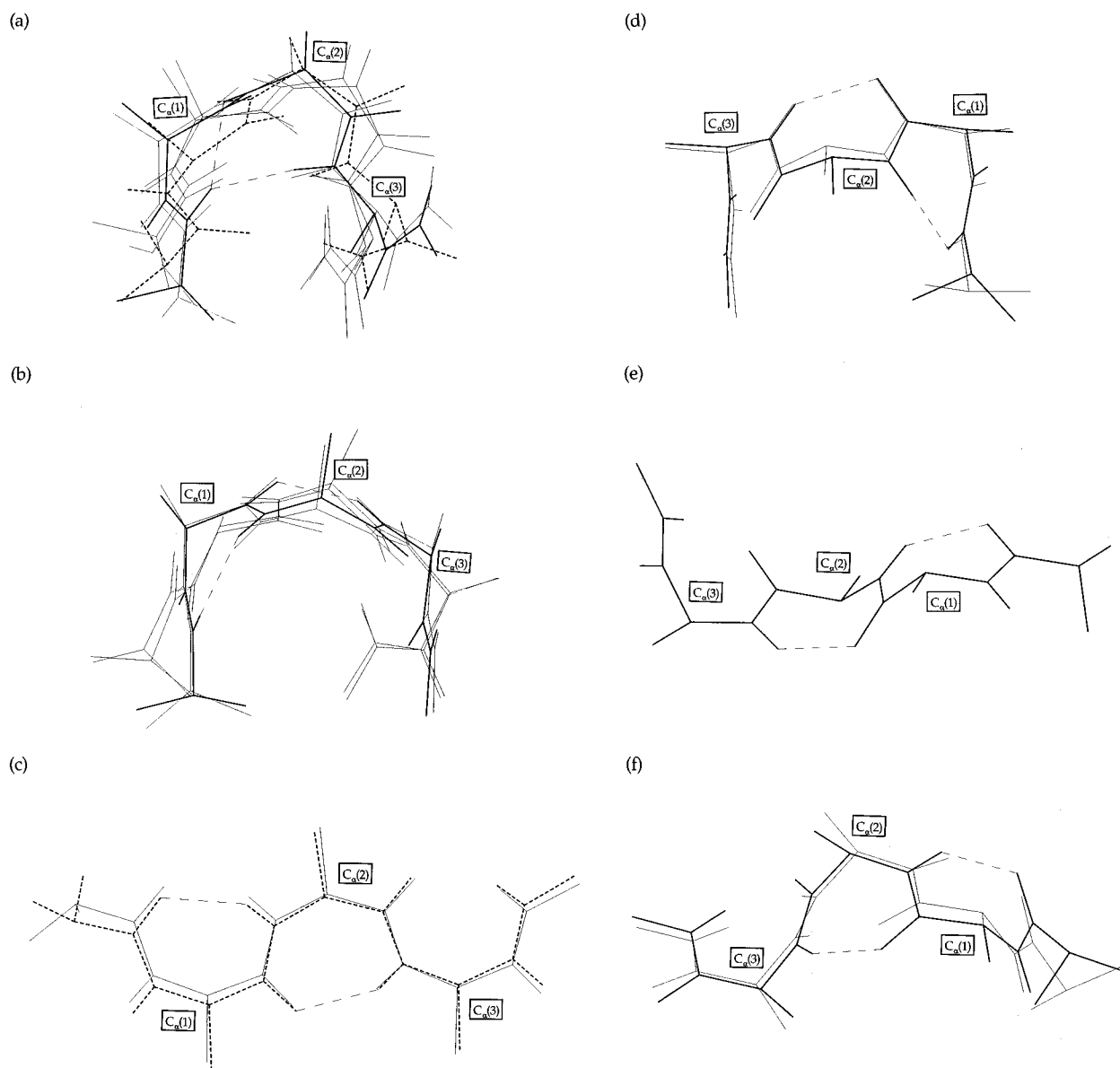


FIG. 9. A sample of IAN conformations from different regions of the potential energy surface, classified according to the vertex that defines the basin to which they belong. Dashed lines indicate hydrogen bonds. (a) Vertex D: minima number 1 (bold), 2, 4, 6, and $16=\alpha$ (bold dashed); (b) Vertex G: minima 3 (bold) 5, 7, and 10; (c) Vertex H: minima 9 and $11=\beta$ (bold dashed); (d) Vertex E: minima 14 (bold) and 19; (e) Vertex E: minimum 20; (f) Vertex F: minima 25 (bold) and 28.

each centered around one of the three lowest minima and including 3 to 4 other minima. The basins around minima number 1 and 2 are associated with vertex D and the basin around minimum number 3 belongs to vertex G. The internal connectivity of these basins is nearly minimal; i.e., with one exception (4, 8, 13) each minimum is connected only to two others. As the total energy increases to $E=4$ kcal/mol [Fig. 10(b)] the basins around minima 1 and 2 increase in size and become connected to form the core of the funnel (vertex D). The G basin around minimum number 3 also increases in size, and a new basin is beginning to form around minima 9 and 11 (vertex H). The interconnectivity within basins D and G has increased considerably; there is a square at the center of D connecting its four lowest minima (numbers 1, 2, 4, and 6), with each connected to several other minima.

Increasing the energy to $E=4.5$ kcal/mol results in an increase in size of all three basins with basins D and G becoming connected to form basin C [Fig. 10(c)]. Although the connectivity within D and G increases, the connection between them to form basin C is rather weak. In fact, there is only one connecting path through a relatively high lying minimum (number 34). This type of connectivity is indicative of an "entropic" barrier, which may disappear at higher energies as more pathways become available. There is a simple "star-shape" connectivity within basin H. At $E=5$ kcal/mol [Fig. 10(d)] basin H is integrated into basin C. While the internal connectivity within the D and G sub-basins continues to increase, the connectivity between basins D, G, and H is still weak and involves few pathways through high lying minima. Thus, the "entropic" barrier encountered

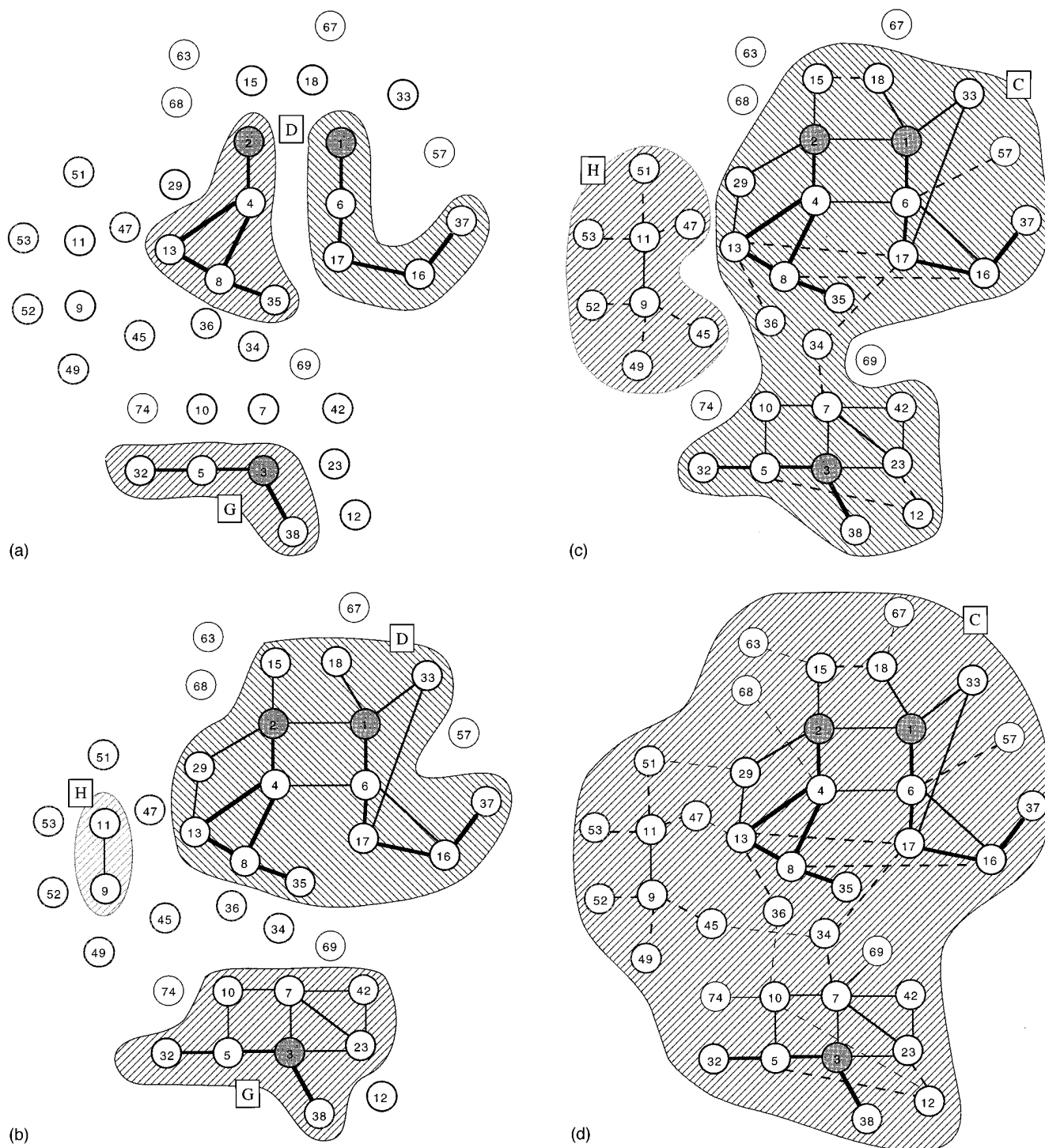


FIG. 10. Basin connectivity graphs showing the internal connectivity between the 39 minima that are included within basin C [Fig. 8(a)] at different energy levels. (a) At total energy $E = 3.5$ kcal/mol; (b) $E = 4$ kcal/mol; (c) $E = 4.5$ kcal/mol, and (d) $E = 5$ kcal/mol. The minima are numbered according to their energy (the global minimum is number 1) and the two-dimensional arrangement was chosen to minimize the complexity of the graph (no special optimization was performed). The dashed regions indicate connected sub-basins at the given energy, and the sub-basins are labeled C, D, G, and H according to the notation in Fig. 8(a). The lines connecting two minima indicate that there is a pathway with a barrier below a given energy. Thick bold lines: $\Phi^\ddagger < 3.5$ kcal/mol; bold lines: $3.5 \text{ kcal/mol} < \Phi^\ddagger < 4$ kcal/mol; dashed bold lines: $4 \text{ kcal/mol} < \Phi^\ddagger < 4.5$ kcal/mol; dashed lines: $4.5 \text{ kcal/mol} < \Phi^\ddagger < 5$ kcal/mol. The filled-in circles (1, 2, and 3) correspond to the minimum within a given sub-basin on Fig. 10(a).

in Fig. 10(c) is somewhat weakened but does not completely disappear. The weak connectivity between these three sub-basins reflects the fact that their conformations are distinctly different [Figs. 9(a)–9(c)].

The finding that the connectivity within a basin is significantly greater than that between basins supports the concept of basin-to-basin kinetics (Sec. VI above). When combined with the lower energy barriers within a basin, it should

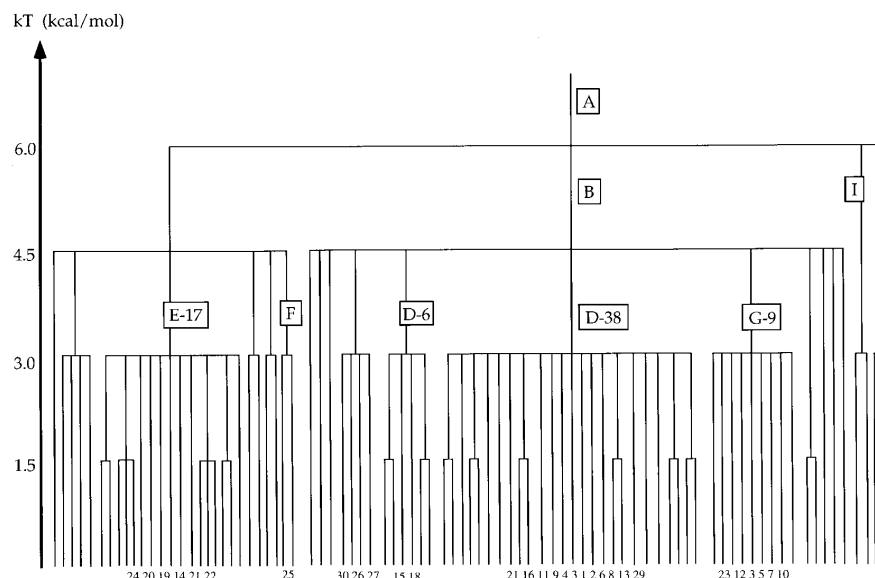


FIG. 11. The $G^T(\Phi)$ map for IAN. Shown are minima with absolute energy lower than -4 kcal/mol (lowest 91 minima). The level spacing in this map correspond to a relative barrier height of 1.5 kcal/mol. The lowest 30 minima are numbered. The letters correspond to the basin assignments defined in Fig. 8(a) and the numbers next to them to the numbers of the minima in the basin (see text).

lead to faster transitions within a basin and slower transitions between basins. Thus, the slower, larger-scale transformations correspond to basin-to-basin kinetics.

From the connectivity data it is possible to locate the shortest path between any two minima. It is defined in terms of the number of transitions required to go from one minimum to another. There are often several possible paths between two minima, even without taking into account that there can be more than one path between directly connected minima; the latter were not included in the graphs for simplicity (see Sec. VII A). For example the shortest path between the α -helix configurations (minimum 16) and the β -sheet configuration (minimum 11), involves four local transitions. It can be written as 16-8-13-47-11 (i.e., a move from minimum #16 to minimum #8, then through minima #13 and #47 to reach minimum #11); this path was discussed in Ref. 18. Three of the barriers crossed along this path are higher than 4 kcal/mol. An alternative path of equal length is 16-17-13-47-11, which has a corresponding set of barrier heights; this was not considered in Ref. 18. It is possible that the slightly longer path (16-6-4-13-47-11) is preferential (or at least competitive) since only two of its five barriers are higher than 4 kcal/mol (see Sec. VII D). Similar arguments can be made within the framework of the canonical ensemble, in which absolute barrier heights are replaced by relative values.

The INA example has shown that the disconnectivity graph $G^E(\phi)$ can be used to reveal the underlying structure of the PES. The overall shape of the potential energy surface is that of a funnel, although at least one relatively large side branch exists. Close to its bottom, the funnel splits to several sub-basins somewhat obscuring the funnel's core. The funnel is centered around the global minimum of this surface. Almost all the deep minima are in or close to the core of the funnel. There exists a high degree of conformational similar-

ity within the sub-basins at the bottom of the funnel, but as one goes up in energy and includes more branches the structural diversity increases and becomes significant even within the main funnel. There is high degree of internal connectivity within sub-basins at the bottom of the funnel, but the interconnectivity between these sub-basins is low. This finding is in accord with the observed conformational diversity.

Although the expected relation between connectivity and structural similarity (especially within local basins) is often present, it does not always hold. In some cases connected conformations exhibit structural similarity, but very dissimilar conformations can show a high degree of connectivity. Thus, the graph provides information that is not evident from simply looking at the structures.

D. Basin-to-basin kinetics

We now focus on the basin-to-basin kinetics of the IAN peptide and make use of the appropriate master equation for a semiquantitative analysis. Qualitative aspects of the kinetics of the system are revealed by the $G^T(\Phi)$ graph, which indicates which minima are kinetically connected; i.e., it shows that the "relative" barriers separating them are lower than a certain kT value. Figure 11 shows the canonical $G^T(\Phi)$ graph of IAN using a barrier height separation of 1.4 kcal/mol. As this is a "relative" energy scale, all the local minima are placed at zero and the nodes of the tree represent the height of the barriers connecting them. This construction is different than that used with the $G^E(\Phi)$ graph that is defined with respect to an absolute energy scale. We take the larger of the two relative barriers, Φ_{AB}^\ddagger for $A \rightarrow B$ and Φ_{BA}^\ddagger for $B \rightarrow A$, to define the kinetic connectivity. Although the $G^E(\Phi)$ graph and the kinetic connectivity graph $G^T(\Phi)$ do not necessarily have to resemble each other, Fig. 11 shows

that the basin structure, which is evident in the $G^E(\Phi)$ graphs in Fig. 8, is also reflected in the kinetic $G^T(\Phi)$ graph. A significant difference between the two graphs is that the relationships between some of the minima in the B basin are slightly altered in the $G^T(\Phi)$ graph.

The information provided by the canonical $G^T(\Phi)$ graph can be used to transform the system's transition matrix \mathbf{W} into the basin block form of Eq. (18). To analyze the resulting kinetic behavior, we use an *approximate* form for the transition rates. It is assumed that all transitions have the same characteristic frequency factor, which for this vacuum system, we set equal to the standard value kT/h . With this assumption Eq. (16) reduces to

$$k_{ij} = W_{ij} = \nu \exp(E_{ij}/kT), \quad (25)$$

where at 300 K, the frequency factor $\nu = kT/h = 6 \times 10^{12} \text{ s}^{-1}$, which corresponds to approximately 200 cm^{-1} . Use of Eq. (25) instead of Eq. (16) for evaluating the transition rates corresponds to a one-dimensional approximation for the transition and assumes that the basins all have the same shape. It is used here to simplify the calculation of the kinetics but has not been employed in constructing the potential surface. Thus, the differences in rates from one minimum to another arise only from the connectivity and from the barriers along each of the paths. Czerminski and Elber in Ref. 18 used complete transition state theory rate expressions; in most cases, the effect was small (Czerminski, private communication) and is not important for the present analyses.

1. Qualitative features of kinetics

In Sec. VI B we discussed how the full transition matrix \mathbf{W} can be rewritten in a block form, based on the basin assignment of the individual states. Any transition matrix can be rearranged so that all the states (local minima) that belong to a given basin appear next to one another in the matrix. The nature of the blocks depends on the temperature at which the mapping was done. Figure 12 shows the transition matrix \mathbf{W} of IAN, at two different temperatures, after rearranging it into a block form similar to Eq. (18). As described in Sec. VI B, the blocks themselves are arranged so that the first block corresponds to the basin which includes the deepest minimum on the surface, the second block corresponds to the basin with the lowest minimum among all the remaining basins, etc. The size of the block is determined by the number of minima included in the associated basin at the given temperature. Thus, in Fig. 12(a), the minima between 1 and 38 all belong to the same basin, to which the global minimum also belongs. The next block, between 39 and 47, includes states connected with the "second deepest" basin, etc. In Fig. 12(b) the first (and largest) block is associated with the B basin and the second block with the E basin (the "side-branch"). The gray level of each \mathbf{W}_{IJ} block (using the terminology of Sec. VI B) is proportional to the *average* of the individual transition probabilities included in that block (excluding the terms along the diagonal) and the dark points indicate the location of the actual nonzero terms.

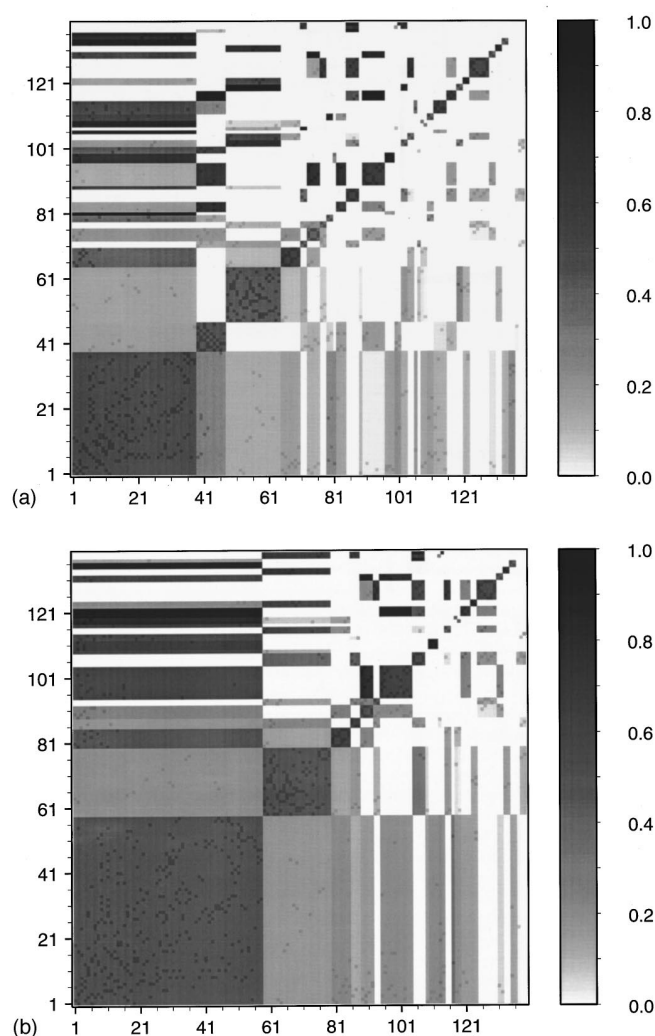


FIG. 12. The transition matrix \mathbf{W} of the tetrapeptide IAN in a basin-blocked form [similar to Eq. (28)] at two temperatures: (a) $kT=3.0$ kcal/mol, (b) $kT=3.5$ kcal/mol. Each block contains all the states (local minima) that belong to the same basin (as defined by the canonic mapping at that specific temperature). The blocks themselves are arranged in the matrix in order of their deepest minima. The gray scale is proportional to the average transition probability in that block [as defined in Eq. (27)] and the darker points indicate the actual transition elements.

From Fig. 12 one sees that the block partitioning of the transition matrix reflects the kinetic character of the system. It also highlights how this character changes with temperature (also seen in Fig. 11). As expected, blocks along the main diagonal, which represent intrabasin transitions, have a dark shade reflecting the high probability of intrabasin transitions. There are many off-diagonal terms which are white, indicating that the two corresponding basins are kinetically completely disconnected; i.e., a transition between two such basins can occur only via a third basin. A closer look at the figure reveals that, while almost all basins have transition elements connecting them to the main funnel (the large square at the bottom left), the connectivity among the rest of the system is much less pronounced. It is particularly small for the high-laying *transient* basins (the lower right corner of the matrices in Fig. 12). There are a few high-laying basins

that are not directly connected to the main funnel (indicated by white “stripes” in the leftmost column). We also see that the transitions in the “lower” triangle of the matrix, from basins with lower minima to basins with higher minima, are smaller on the average than transitions in the upper triangle (indicated by their lighter shades of gray). This last observation indicates that, for *intermediate* times, when small probability transitions can be neglected, only the upper triangle of the matrix is kinetically significant. At these times the transitions from the “higher” basins to the “lowest” basin (the central funnel) dominate the kinetics. This results in an overall flow of probability from higher areas of the PES to the central funnel (and the native conformation). The reverse transitions become significant at longer times when equilibrium is established.

The kT values used in Fig. 12 were chosen because they make a difference in the observed structure. Due to the nature of the specific system, which is reflected in Fig. 11, the interesting kinetic structuring occurs around 3.0 kcal/mol. At lower kT values (e.g., 1.0 kcal/mol) the system is almost entirely kinetically disconnected, while at higher values (e.g., 4.5 kcal/mol) the system is almost entirely connected into one or two basins.

2. Quantitative features of kinetics

To follow the time development of the basin probability vector $\mathbf{P}'(t)$ it is necessary to solve the master equation [Eqs. (15) or (22)]. A convenient way of solving Eq. (15) is by expanding the probability vector $\mathbf{p}(t)$ in terms of the eigenvectors and eigenvalues of the transition matrix \mathbf{W} . If \mathbf{s}_i are the eigenvectors and λ_i are the corresponding eigenvalues, the time evolution of the probability vector $\mathbf{p}(t)$ can be written as¹⁶

$$\mathbf{p}(t) = \mathbf{p}^{\text{eq}} + \sum_{\lambda_i < 0} C_i \mathbf{s}_i e^{\lambda_i t}, \quad (26)$$

where the coefficient $C_i = [\mathbf{s}_i | \mathbf{p}(0)]$ is determined by the initial distribution, $\mathbf{p}(0)$. Since all $\lambda_i \leq 0$, the equilibrium distribution \mathbf{p}^{eq} is the eigenvector that corresponds to $\lambda_i = 0$; i.e., as t approaches infinity, $\mathbf{p}(t)$ approaches \mathbf{p}^{eq} and all the other terms decay to zero.

Because \mathbf{W} is a nonsymmetric matrix, it cannot be guaranteed that it has a complete set of eigenvectors spanning its space. However, with the condition of detailed balance ($W_{ij}P_j^{\text{eq}} = W_{ji}P_i^{\text{eq}}$), there exists a symmetry transformation that transforms \mathbf{W} into a real symmetric matrix \mathbf{V} ,¹⁶

$$\mathbf{V} = [\mathbf{P}^{\text{eq}}]^{-1/2} \mathbf{W} [\mathbf{P}^{\text{eq}}]^{1/2}, \quad (27)$$

with $[\mathbf{P}^{\text{eq}}]$ a diagonal matrix with the equilibrium vector \mathbf{p}^{eq} along its diagonal. The symmetric \mathbf{V} matrix can be easily solved with standard methods to obtain its eigenvectors \mathbf{u}_i and eigenvalues κ_i . Due to the nature of the transformation, the \mathbf{V} and the \mathbf{W} matrices have the same eigenvalues, $\kappa_i = \lambda_i$, and their eigenvectors are related by the transformation operator used in Eq. (27); i.e.

$$\mathbf{u}_i = [\mathbf{P}^{\text{eq}}]^{-1/2} \mathbf{s}_i. \quad (28)$$

Applying the transformation, Eq. (28), to the probability vector $\mathbf{p}(t)$ results in a transformed probability vector $\mathbf{q}(t) = [\mathbf{P}^{\text{eq}}]^{-1/2} \mathbf{p}(t)$, which obeys the master equation [Eq. (15)] with the symmetric transition operator \mathbf{V} ; that is

$$\dot{\mathbf{q}}(t) = \mathbf{V} \mathbf{q}(t). \quad (29)$$

As in Eq. (26), the solution of Eq. (28) can be cast in terms of the eigenvectors \mathbf{u}_i and eigenvalues κ_i of the \mathbf{V} matrix; that is

$$\mathbf{q}(t) = \mathbf{q}^{\text{eq}} + \sum_{\kappa_i < 0} C'_i \mathbf{u}_i e^{\kappa_i t}, \quad (30)$$

with $C'_i = [\mathbf{u}_i | \mathbf{q}(0)]$. Solving for Eq. (30) and transforming back to Eq. (26), we get the time evolution of the probability vector $\mathbf{p}(t)$,

$$\mathbf{p}(t) = \mathbf{p}^{\text{eq}} + \sum_{\kappa_i < 0} C'_i [\mathbf{P}^{\text{eq}}]^{1/2} \mathbf{u}_i e^{\kappa_i t}, \quad (31)$$

with $C'_i = [\mathbf{u}_i | [\mathbf{P}^{\text{eq}}]^{-1/2} \mathbf{p}(0)]$. In the final form of Eq. (26) the coefficients are given by $C_i = [\mathbf{s}_i | [\mathbf{P}^{\text{eq}}]^{-1} \mathbf{p}(0)]$, where $[\mathbf{P}^{\text{eq}}]^{-1}$ is the normalization factor of the \mathbf{s}_i vectors.

In performing the calculation we proceeded as follows: First we found the eigenvalues of the nonsymmetric \mathbf{W} matrix using standard techniques (i.e., we transformed to Hessenberg form and then found the eigenvalues using the QR algorithm⁴²). The equilibrium vector \mathbf{p}^{eq} was obtained through LU decomposition and “inverse iteration” starting from the $\lambda_i = 0$ eigenvalue.⁴² Then the \mathbf{W} matrix was transformed to the symmetric form \mathbf{V} [Eq. (27)] and solved for eigenvalues κ_i and eigenvectors \mathbf{u}_i by reducing it to tridiagonal form and then applying the QL decomposition algorithm.⁴² Finally the time evolution of the population vector $\mathbf{p}(t)$ was obtained by solving Eq. (31).

Figure 13(a) shows the time evolution of the population probability $P'_j(t)$ [Eq. (20)] for several basins at $kT = 3.0$ kcal/mol, starting from uniform distribution. Figure 13(b) shows the same results but the basin probabilities are divided by the number of minima in the basin to correct for size effects. The time is given in terms of the characteristic vibrational period used in Eq. (25); i.e., $\tau = 1/\nu = h/kT = 1.6 \times 10^{-13}$ s. Equilibrium is reached after about 60 time units. The “kinetic basins” are indicated by letters, which correspond to the notation used in $G^E(\Phi)$ and $G^T(\Phi)$ [Figs. 8(a) and 11], and by numbers, which indicate size of basin. The “kinetic basin” D-38 (defined by the map M^T) includes most of basins D and H of Fig. 8(a) (the M^E map); basin D-6 is also a part of basin D of Fig. 8(a). The notation T-7 and T-6 indicates groups of high-lying transient states [not really a basin in the strict sense; T stands for transients, as in Eq. (18)].

The flow of probability in Figs. 13(a) and 13(b), which starts with a uniformly distributed probability over all states of the system, clearly shows the tendency towards the central funnel on the PES. One sees that the probability of finding the system in the central basins increases both in absolute value and on a per minimum basis (basins D-38, D-6, and G-9). The probability of finding the system in high-lying

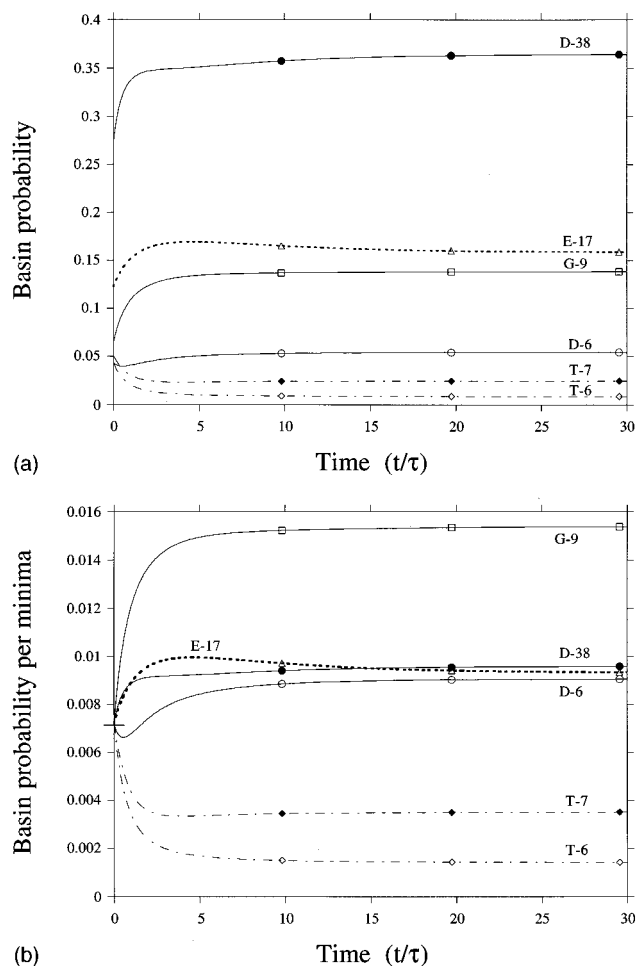


FIG. 13. The time evolution of several basin population probabilities $P'_j(t)$ at $kT=3.0$ kcal/mol, starting from uniform distribution. The time is given in terms of the characteristic vibrational period used in the approximation of Eq. (38), i.e., $\tau=1/\nu=h/kT=1.6\times 10^{-13}$ s. The basins are indicated by letters that relate to the notation of Fig. 8(a), and numbers, which indicate size of basin. Basins D-38, D-6, and G-9 are part of the central funnel, basin E-17 is a side branch [see Fig. 8(a)] and basins T-7 and T-6 are high laying transient basins (centered around minima that rank 66 and 117 on the scale of 139, respectively). (a) Actual basin population probabilities, (b) basin population probabilities divided by the number of minima in each basin, to account for size effects. The horizontal bar on the y axis represents the initial uniform probability of 1/139 per minima.

transient T basins decreases rapidly with time (T-7 is a seven membered basin whose lowest minimum ranks 66 on the scale of 139 minima; and T-6 is a six membered basin with a lowest minimum ranking 117). The behavior of the side branch E-17 [basin E in Fig. 8(a)] is interesting. The results indicate an initial flow of probability into this basin which is later depleted in favor of the central basins; the depletion is related to the approach to equilibrium, since the minima of basin E are higher in energy than some of the basins in of the main funnel. Other small basins (e.g., D-6) also exhibited nonmonotonic kinetics with either an initial decrease or an initial increase in probability followed by the opposite trend in approaching equilibrium.

The probability per minima of the G-9 basin in Fig. 13(b) is higher than that of the main D basin (both D-38 and

D-6). This is due to the fact that the larger D-38 basin includes alongside the lowest minima some local minima that are slightly higher in energy (on the slopes of the funnel). These reduce the equilibrium "per minimum" values. This phenomenon is less significant in the smaller G-9 basin.

Folding model. There is considerable interest in the kinetics of protein folding.³⁶ The present technique can be applied to this problem, in principle, by starting from the very large basin of "unfolded" states and following the system until it coalesces into the thermodynamically stable basin corresponding to the native state. Although the tetrapeptide is clearly too small to represent the folding behavior of a protein, it can be used as an interesting model system. We recast its behavior as a folding problem by making certain assumptions concerning the nature of the basins. We distinguish three groups of states (see Fig. 11):

- (i) *Folded states:* The 53 local minima that are included in the central funnel on the surface (basin B in Fig. 8). This group of states includes the global minimum and both the " α -helical" conformation and the so-called " β -sheet" conformation.
- (ii) *An "intermediate" or "trap" basin:* This is basin E in Fig. 8, which is outside of the main funnel; it includes 17 local minima.
- (iii) *"Unfolded" states:* The rest of the system, which consists of 69 local minima.

Figure 14(a) shows the time evolution of the population probability $P'_j(t)$ [Eq. (20)] of IAN at $kT=3.0$ kcal/mol, starting from a distribution in which *only* the unfolded states are populated with equal probability (the probability of the folded and intermediate states is initially zero). Figure 14(b) shows the same kinetics corrected for size effects (i.e., where the basin probabilities are divided by the number of minima in the basin). As before, the time is given in terms of the characteristic vibrational period used in the approximation of Eq. (25), i.e., $\tau=1/\nu=h/kT=1.6\times 10^{-13}$ s. Equilibrium is reached after about 60 vibrational periods; only the first 20 periods are shown. In general, both figures show the expected folding kinetics. There is an overall flow of probability from the unfolded states to the folded states as the system approaches equilibrium. Due to the small size of the system and the approximate character of the transition probabilities, the probability of occupying unfolded states does not vanish at equilibrium. It can be seen in Fig. 14(b) that the global minimum has the largest probability.

The most interesting aspect of the kinetics is the time evolution of the intermediate basin. As is seen in Fig. 14(b), the curve describing the probability per minimum of finding the system in these states first increases and then decreases somewhat. The behavior is commonly referred to, in protein folding as well as in general kinetics, as a "kinetic intermediate." However, the group of states are not kinetic intermediates which have a positive role in the folding process, but rather traps outside of the main folding pathway. These states interfere with the folding process rather than helping it, as has often been assumed.³⁶ This result is similar to the

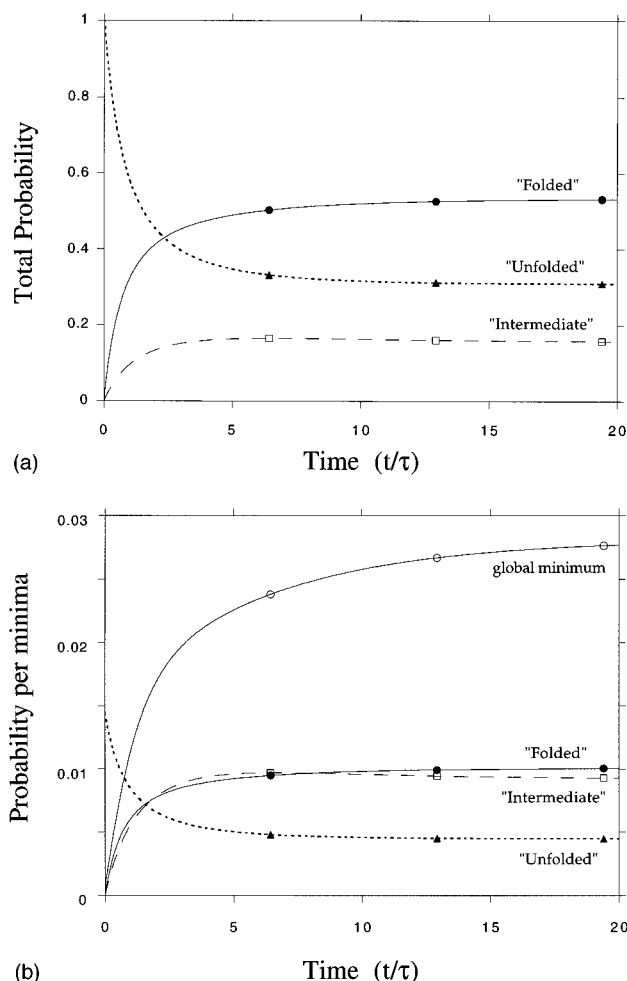


FIG. 14. The “folding kinetics” of alanine tetrapeptide. Defined are three groups of states: folded, intermediate, and unfolded (see text). Shown is the time evolution of the population probability $P_j^i(t)$ of these states at $kT=3.0$ kcal/mol, starting from an initial distribution in which *only* the unfolded states are populated with even probability (the probability of the folded and intermediate states is initially zero). The time is defined as in Fig. 13. (a) Actual basin population probabilities, (b) basin population probabilities divided by the number of minima in each basin, to account for size effects.

conclusion reached by Wolynes and co-workers from analyzing an analytic folding model.²⁷

VIII. CONCLUDING DISCUSSION

The multidimensional potential energy surfaces of complex systems are difficult to interpret and there is widespread interest in simplified representations that preserve the essential features. In this paper we have defined a procedure that maps the potential energy surface onto the set of local minima. Disconnectivity graphs that describe the results of this mapping procedure were used to analyze the nature of the potential surface. The map differs from the direct minimization maps used previously in that it preserves the temperature dependent picture of “attraction basins” and retains information about the connectivity of the surface, including the barriers between minima. From any given basin, a downward “look” along the branches of the graph reveals the

underlying structure of the surface, while an upward look towards its root reveals the structure of the transitions.

The topological analysis can be applied to the glass transition and the concept of freezing, which have been discussed in terms of the structure of the potential energy surface.^{5,9,25} It is generally assumed that above the fluid–solid transition temperature, T^{tr} , most physical system are ergodic; i.e., they are not trapped in any subregion of the configuration space and, given a reasonable time, any dynamical trajectory will span all of the available space. “Reasonable time” in this context may be considerably longer than the length of a “long” molecular dynamics trajectory (1–10 ns). This means that for $T_+ > T^{\text{tr}}$ all of the PES is incorporated in a single basin $R^{\text{erg}}(\alpha') = \Gamma$ [Eq. (11)]. Below the transition temperature, T^{tr} , a glass forming material is trapped in one of its many disordered conformation. This indicates that the corresponding PES is very rough, and that the roughness starts immediately below T^{tr} . The nature of the glass transition is reflected in Fig. 1, which shows the topological graph of such a potential energy surface. Above T^{tr} the system is ergodic (there is only one node, i.e., a single ergodic basin, at levels 2, 3, and 4 of the graph) but below the transition temperature, which is at level 2 of the graph, there is significant splitting which indicates that trapping occurs in subregions of the PES (levels 1, 0 etc. of the graph).

Upon cooling a glass forming material below the glass transition temperature, $T_- < T^{\text{tr}}$, the system that starts at a random configuration $\mathbf{r} \in \Gamma$ is trapped in one of the many branches of the graph and is likely to end up in a configuration that is significantly different from the global minimum; i.e., upon cooling, the system descends along the branches of the disconnectivity graph $G^T(\Phi)$ and, as the ergodic basin $R^{\text{erg}}(\alpha')$ splits into smaller $R^T(\alpha')$ basins, the system gets trapped in a localized region of the PES. Correspondingly, at any given T_- temperature, the $M^T(\mathbf{r})$ mapping will result in a large set of $\{\alpha'\}$ minima [Eq. (3)]. The difference between the basin structure of a crystal and that of a glass can be related to the way in which the $R^T(\alpha')$ sets become disconnected as the temperature drops. In contrast to the graph for a glass described above, the dominant features of the PES for a crystal forming substance are expected to be similar to those of a single minimum with small fluctuations, as in Sec. IV B.

As an illustration of the approach, we have applied it to the potential energy surface of the alanine tetrapeptide in vacuum, which is the smallest peptide that can form a single α -helical turn. The analysis was based on 139 local minima and 502 barriers determined by Czerminski and Elber.^{17,18} Although there are only seven significant dihedral angle degrees of freedom (Φ_1 through Φ_3 and Ψ_0 through Ψ_3), the “disconnectivity” graph exhibits some interesting features. It has a funnel-like structure that encompasses many minima, including the global minimum. Analysis of the detailed connectivity of the minima and the kinetics of transitions between them showed the complexity present in this very small system.

The topological mapping analysis presented in this paper can be compared with a recent statistical analysis by Kunz

and Berry of the potential surface for a 19-atom Lennard-Jones cluster.¹³ The two analyses share many concepts, in that they both characterize the surface by the basin structure, highlighting the connectivity between the basins and following the basin-to-basin kinetics. Nevertheless, there are several significant differences. The study of Kunz and Berry is closely related to the work of Stillinger and Weber^{9,10} and others.^{11–15} It differs from the earlier papers in that its emphasis is more on connectivity and kinetics than on thermodynamics. The paper determines minima sampled along high temperature dynamical trajectories with the connectivity, including saddle points determined for successive coordinate sets along a given trajectory. These results in a “linear” connectivity map, which is equivalent to a one-dimensional cut through the multidimensional surface. The topological mapping presented here analyzes the full multidimensional surface, without first reducing its dimensionality. It provides information regarding the overall structure of the entire surface and the multidimensional connectivity between the different basins as represented in the disconnectivity graph. In the specific examples, it is interesting to note that, although the number of particles (26 in the peptide; 19 in the cluster), the number of minima (139 for the peptide, 299 for the cluster), and the number of barriers (502 for the peptide and 461 for the cluster) are similar, the former correspond essentially to the entire set of minima and most of the barriers, while the latter is a trajectory-based sample out of an estimated 500 000 minima;¹³ no estimate of the number of barriers was given. The smaller number of minima in the peptide is a direct consequence of the bonded connectivity of the atoms.

Although the tetrapeptide is a very simple system, the properties of its potential surface can be used to comment on questions raised for proteins. One question in protein folding concerns the relative importance of pronounced local minima versus an overall funnel-like structure of the potential energy surface. In the tetrapeptide these two features are coupled and the largest funnel is centered around the global minimum. The deep minimum can be viewed as a “vortex” perturbing the surface around it to form a funnel; the deeper the minimum the larger the funnel is expected to be. Thus, it is possible that in proteins the thermodynamic requirement for folding (a deep global minimum) is coupled to a kinetic solution to the search problem (the funnel). This picture is the opposite extreme from the “golf-course” potential, in which the deep global minimum does not influence the rest of the surface.

Another question concerning proteins is the existence of metastable states. These correspond to minima that may have very different structures and energies. They are assumed to be separated by barriers that cannot be overcome on biological time scales^{3,43} so that the protein can be trapped in one of the minima during the folding process. This requires that the potential energy (or free energy) surface of a protein is ragged and glass like. As has been pointed out,³⁶ without a *unique* metastable state, folding would be ineffective because the same sequence could yield functional and nonfunctional folded polypeptide chains. In the tetrapeptide, almost all the deep minima are located close to each other in the main

funnel and are separated by relatively low barriers. The conformational similarity was generally correlated with connectivity and with spatial proximity in configuration space (although there are exceptions to this rule⁴⁴). Configurations that were far away on the surface were also generally dissimilar. Thus, a requirement of the metastability hypothesis (that there exist similar structures in remote regions of configuration space) is not supported by the tetrapeptide results. Further, similar energies are not necessarily related to similarity in structure.

The basin picture used in the analysis highlights the fact³⁸ that at physiological temperatures the tetrapeptide, like a protein, is not confined to a single minimum but rather is moving freely within the attraction basin, which corresponds to the accessible multimimum potential surface at that temperature. Although it includes the global minimum, a dynamical trajectory can glide over the low barriers and explore a large part of the basin, without being sensitive to the details of the underlying structure.

Although the tetrapeptide is a minimal system, it is sufficiently complex to show features that may be of interest for larger polypeptides and stable proteins. To test this possibility, a similar topological study of the potential surface of a protein is being made [O. M. Becker (in preparation)].

ACKNOWLEDGMENTS

We thank Ron Elber and Ryszard Czerminski for making their data on IAN available to us, and thank Eugene Shakhnovich and Andrej Šali for helpful discussions. This work has been supported in part by a grant from the National Science Foundation.

APPENDIX: TOPOLOGICAL CHARACTERISTICS OF THE MAP

Since all the $R(\alpha)$ and $R^T(\alpha')[R^E(\alpha')]$ basins are open sets they can, in principle, form a topological space. We begin by defining the *universal set* $X = \{\alpha_1, \dots, \alpha_i, \dots\}$, which contains all the local minima on the PES Φ . Each of these α_i minima correspond to a unique $R(\alpha)$ basin (or equivalence class if $\sigma > 1$). The mapping procedures M^d , M^T , and M^E partition the universal set X in different ways. The direct minimization map $M^d(\mathbf{r})$ partitions X into small sets, each consisting of a single local minimum, i.e., $\{\alpha_i\}$. The $M^T(\mathbf{r})$ map, on the other hand, partitions X into α_i^T sets [or α_i^E sets for $M^E(\mathbf{r})$] that may include more than a single local minimum. The members of a set α_i^T are all the local minima α_j included in $R^T(\alpha'_i)$, i.e.,

$$\alpha_i^T = \{\alpha_j | \alpha_j \in R(\alpha_j) \subset R^T(\alpha'_i)\}. \quad (\text{A1})$$

This partition depends parametrically on the mapping temperature T (or energy E).

Since a topological structure is the relationship between open sets, we now define the sets-of-sets to be investigated. We focus on the properties of three different sets-of-sets A_i , which are defined through the different mapping procedures:

(i) The set of all sets generated by the direct minimization map $M^d(\mathbf{r})$

$$A_1 = \{\{\alpha_{ij}\} \mid \text{all } \alpha_i \in X. \quad (\text{A2})$$

(ii) The set of all sets generated by the map $M^T(\mathbf{r})$ at a specific temperature as defined in Eq. (A1)

$$A_2 = \{\alpha_i^T\} = \{\alpha_1^T, \alpha_2^T, \dots\} \quad \text{all } \alpha_i \in X. \quad (\text{A3})$$

(iii) The sets-of-sets generated by applying the $M^T(\mathbf{r})$ map at several different temperatures

$$A_3 = \{\{\alpha_i^0\}, \dots, \{\alpha_i^T\}, \dots, \{\alpha_i^\infty\}\} \quad \text{all } \alpha_i \in X$$

$$\text{and } 0 \leq T \leq \infty. \quad (\text{A4})$$

A_1 is a special case of A_2 in which $T=0$. Also note that the $\{\alpha_i^0\}$ sets of A_3 are the elements $\{\alpha_i\}$ of A_1 and that $\{\alpha_i^\infty\} = X$. Thus, Eq. (A4) can be rewritten as

$$A_3 = \{A_1, \dots, \{\alpha_i^T\}, \dots, X\} \quad \text{all } \alpha_i \in X$$

$$\text{and } 0 < T < \infty. \quad (\text{A5})$$

To illustrate what is involved in these sets-of-sets, let's consider the simple example of a system with three local minima. The universal set in this case is $X = \{\alpha_1, \alpha_2, \alpha_3\}$. The M^d map generates from it the set $A_1 = \{\{\alpha_1\}, \{\alpha_2\}, \{\alpha_3\}\}$. Assuming that at an intermediate temperature T_1 minima α_1 and α_3 are connected, the set A_2 generated by $M^{T_1}(\mathbf{r})$ at this temperature is $A_2 = \{\{\alpha_1, \alpha_3\}, \{\alpha_2\}\}$. Finally, the A_3 set for this example is $A_3 = \{\{\alpha_1\}, \{\alpha_2\}, \{\alpha_3\}, \{\alpha_1, \alpha_3\}, \{\alpha_1, \alpha_2, \alpha_3\}\}$.

It is useful to describe some elementary properties of the A_i sets. First, the three sets *cover* the universal set X , i.e.,

$$X = \cup a \quad \text{for all } a \in A_i. \quad (\text{A6})$$

Also, having the relations $A_1 \subset A_3$ and $A_2 \subset A_3$ we see that A_3 is *finer* than both A_1 and A_2 . Note, that, although the elements of A_2 are unions of the elements of A_1 , $A_1 \not\subset A_2$, both A_1 and A_2 are *disjoint* sets; i.e., for every $x, y \in A_i$ ($i = 1, 2$) the intersection is empty, $x \cap y = \emptyset$. On the other hand the set A_3 is not disjoint; i.e., there exist $x, y \in A_3$ for which the intersection is not empty. For example, the intersections $\{\alpha_i^T\} \cap \{\alpha_i^0\} = \{\alpha_i^0\}$ or \emptyset , depending on the condition $\alpha_i^0 \in R(\alpha) \subset R^T(\alpha')$; i.e., whether or not α_i^0 belongs to the α_i^T super basin.

Since A_1 and A_2 are disjoint sets they are *trivial posets* (partially ordered sets), because in a disjoint set the partial order relation is trivially reflexive, transitive, and antisymmetric. The corresponding *diagram*,^{31,34} which displays the inclusion relations between the elements of the poset, is therefore a disconnected set of points. The structure of A_3 is more interesting, as it is a nontrivial poset. This structure is discussed in detail in Sec. III. The structure of the inclusion relation in this poset is given by a directed graph, an example of which is given in Fig. 1.

¹E. R. Davidson, Chem. Rev. (1993), 2337–2581.

²M. Karplus and E. Shakhnovich, in *Protein Folding* 127, edited by T. Creighton (Freeman, New York, 1992).

- ³P. W. Anderson, B. I. Halperin, and C. M. Varma, Philos. Mag. **25**, 1 (1972).
- ⁴C. Laermans and A. Vanelstraete, Phys. Rev. B **34**, 1405 (1986).
- ⁵J. L. Green, J. Fan, and C. A. Angell, J. Phys. Chem. **98**, 13780 (1994).
- ⁶R. D. Levine and R. B. Bernstein, *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford University Press, New York, 1987).
- ⁷B. M. Pettitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985).
- ⁸A. G. Anderson and J. Hermans, Proteins **3**, 262 (1988).
- ⁹F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984); Phys. Rev. A **25**, 978 (1982).
- ¹⁰F. H. Stillinger and T. A. Weber, J. Phys. Chem. **87**, 2833 (1983).
- ¹¹H. Tanaka and I. Ohmine, J. Chem. Phys. **91**, 6318 (1989).
- ¹²R. S. Berry, Chem. Rev. **93**, 2379 (1993).
- ¹³R. E. Kunz and R. S. Berry, J. Chem. Phys. **103**, 1904 (1995).
- ¹⁴R. Elber and M. Karplus, Science **235**, 318 (1987).
- ¹⁵T. Noguti and N. Gô, Proteins: Structure, Function, and Genetics **5**, 97, 104, 113, 125, 132 (1989).
- ¹⁶N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- ¹⁷R. Czerminski and R. Elber, Proc. Natl. Acad. Sci. USA **86**, 6963 (1989).
- ¹⁸R. Czerminski and R. Elber, J. Chem. Phys. **92**, 5580 (1990).
- ¹⁹Y. S. Bay and M. D. Fayer, Phys. Rev. B **39**, 11066 (1989).
- ²⁰A. Ansari *et al.* Proc. Natl. Acad. Sci. USA **82**, 5000 (1985).
- ²¹I. E. T. Iben *et al.* Phys. Rev. Lett. **62**, 1916 (1989).
- ²²H. Frauenfelder, S. G. Sliger, and P. G. Wolynes, Science **254**, 1598 (1991).
- ²³S. J. Hagen, J. Hofrichter, and W. A. Eaton, Science **269**, 959 (1995).
- ²⁴B. R. Gelin and M. Karplus, Proc. Natl. Acad. Sci. USA **72**, 2002 (1975).
- ²⁵M. Klein, Annu. Rev. Phys. Chem. **4**, 23 (1991).
- ²⁶P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
- ²⁷P. Wolynes *et al.* Science **267**, 1618 (1995).
- ²⁸The map is actually defined from the \mathfrak{R}^{3N-6} configuration space and not \mathfrak{R}^{3N} , since otherwise every local minima α will formally be in an infinite equivalence class, stemming from the fact that the energy function is invariant under rigid translation and rotation.
- ²⁹S. Kirkpatrick, C. D. Gelatt, Jr, and M. P. Vecchi, Science **220**, 671 (1992).
- ³⁰R. M. J. Cotterill and J. U. Madsen, in *Characterizing Complex Systems*, edited by H. Bohr (World Scientific, Singapore, 1990), pp. 177–191.
- ³¹R. E. Merrifield and H. E. Simmons, *Topological Methods in Chemistry* (Wiley, New York, 1989).
- ³²J. R. Munkres, *Topology* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
- ³³F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).
- ³⁴D. E. Rutherford, *Introduction to Lattice Theory* (Hafner, New York, 1965).
- ³⁵B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1983).
- ³⁶M. Karplus and A. Sali, Current Opinion in Structural Biology **5**, 58 (1995).
- ³⁷T. L. Hill, *Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, MA, 1960).
- ³⁸C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure & Thermodynamics*, Adv. Chem. Phys. LXXI (Wiley, New York, 1988).
- ³⁹B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
- ⁴⁰Reference 17 reported 513 barriers. We received 502 barriers which provide an adequate picture of the surface.
- ⁴¹P. H. Maccallum, R. Poel, and E. T. Milner-White, J. Mol. Biol. **248**, 374 (1995); T. Lazaridis (private communication).
- ⁴²W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1989), Chap. 11.
- ⁴³J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526 (1990); Biopolymers **32**, 695 (1992).
- ⁴⁴O. M. Becker, Proteins: Structure, Function, and Genetics **27** (1997).