Theor Chim Acta (1987) 71: 75–88

Theoretica Chimica Acta

The energy well connection graph: a use of curvilinear coordinates to achieve dimension reduction for graph representation of molecular conformation energy data

T. M. Creese[1], **Gary L. Grunewald**[2] **and Mary W. Creese**[2]

[1]Department of Mathematics, University of Kansas, Lawrence, KS 66045, USA
[2]Department of Medicinal Chemistry, University of Kansas, Lawrence, KS 66045, USA

(Received August 5, revised and accepted November 21, 1986)

In the conformation space of a flexible molecule, curvilinear coordinate paths connecting conformations of segments of the given molecule are used to reduce the number of variables required for describing barriers between preferred conformations of the molecule as a whole. The technique is applied to a hypothetical example, then to *m*-trifluoromethyl-N-ethyl-amphetamine (fenfluramine), and to N-methyl-3-phenyl-3-(*o*-methoxyphenoxy)-propylamine (nisoxetine). In the latter example, the number of variables is thus reduced from six to three. In all three examples, a graph representation of low-energy well connection is achieved. The limit of easy comprehension has thus been moved back from about three torsion angles to three effective segments. Within this limit the procedure leads to a quantitative diagram, which is no harder to read than a contour map, showing the barriers to low-energy interconnection among the favorable conformations of a moderately complex molecule.

Key words: Non-linear dimension reduction — Curvilinear coordinates — Graphical presentation — low-energy well connection graph

1. Introduction

Many techniques of reducing the dimensionality of data are known. Some are familiar and are often used. We present an application of curvilinear coordinates whose uses in chemistry do not seem to have attracted attention.

Investigators have not neglected the possibility of analyzing the conformation energy map of a complex molecule by looking successively at important flexible segments [1-3]. However, the subsequent synthesis often requires so many variables that the collection, analysis, and presentation of data for the molecule as a whole may appear impractical. In this situation, simplification by introducing curvilinear coordinates is possible.

Suppose for now that a molecule may be divided into two segments, the first comprising n_1 rotation angles, and the second n_2. The rotation- or conformation-space, U_1, of the first segment has dimension n_1 while U_2, that of the second, has dimension n_2. The space of the whole molecule is the cartesian product $U_1 \times U_2$, which has dimension $n_1 + n_2$.

If arcs Γ_1 and Γ_2 are chosen in U_1 and U_2, then in $U_1 \times U_2$ those arcs will be orthogonal, no matter how they are chosen. If the arcs are somewhat smooth then the cartesian product of Γ_1 and Γ_2 may be thought of as embedding a plane into $U_1 \times U_2$ as a curved 2-dimensional submanifold. Because it is 2-dimensional, those features of the space which appear on (or even near) the embedded plane can be presented readily on paper for analysis or publication.

For a long time, linear techniques such as those based on regression have been used with great confidence to embed a plane into euclidian space, leaving its distance and angle properties calculable when needed. In contrast, the novelty of the present procedure is to allow non-linear and indeed virtually arbitrary embeddings of the plane, which necessarily appears flat in its own coordinate space. The concommitant cost is the immediate loss of all the metric properties, curvature, distance and angle. We illustrate that in some cases this cost is acceptable since those properties are virtually unused. As with linearly-embedded planes, points which lie close to a plane embedded non-linearly may often be considered as belonging in the resulting surface.

At times the technique may suggest realistic savings of calculational effort, but it is not presented as a new calculational device. It shows instead how to organize and reduce the amount of data necessary for consideration and thus minimize the dimensional obstacles to comprehension of the results. It is primarily a matter of analysis and presentation [4].

The representation which results is a "graph", an object of a well known class whose uses in chemistry are acquiring prominence [6-8]. In fact, the graphs presented in this study will have "capacity", and the final example will not be 2-dimensional but 3-dimensional.

2. A hypothetical example illustrating the technique only

Realistic examples of sufficient complexity to illustrate the technique tend to be hard to present. Therefore, the first example is artificial in order to be easily accessible.

Suppose the molecule M consists of a chain of three bonds with rotation angles τ_1, τ_2, τ_3 joining four groups Y_1, Y_2, Y_3, Y_4 whose rigidity is not in question

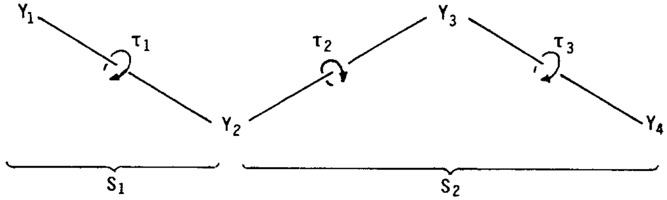

Fig. 1. Segment structure of the hypothetical molecule M

(see Fig. 1). Suppose further that M is thought of as divided into two segments, S_1 whose configuration is described by one rotation angle, τ_1, and S_2, described by τ_2 and τ_3. The rotation space of S_1, called U_1, is of dimension $n_1 = 1$; U_2, that of S_2, is of dimension $n_2 = 2$; and that of M, the cartesian product called $U = U_1 \times U_2$, is of dimension $n_1 + n_2 = 3$. Into this space we embed a plane as a 2-dimensional manifold intersecting the energy wells and containing the connections between them.

Suppose that in the isolated segments S_1 and S_2, the rotations τ_1 and τ_3 each have two favored values, 90° and 270°, while τ_2 has three, at 60°, 180°, and 300°. The corresponding favored positions in the rotation spaces of the segments are labelled with capital letters, as in Fig. 2.

In the cartesian product $U_1 \times U_2$, the rotation space of M, the expected configurations would be pairs; AC = (90°, 60°, 90°), BC = (270°, 60°, 90°), ..., BH = (270°, 300°, 270°), a pair of letters corresponding to a triple of angles (τ_1, τ_2, τ_3) determining a configuration of M.

Now suppose that near each of these locations, there is in fact an energy well, the energies [9] near AC, BC, AE, BE, AF, BF, AH, and BH being 2 k; those near AD, BD, AG, and BG being 0 k. Finally, suppose that the connection barrier heights are as in Table 1. A picture of U is shown in Fig. 3, indicating the energy wells and the paths with low barrier heights.

In this oblique view, the path from BC to BG is quite clear, via AC and AG, with two 6 k barriers between. However, the usual representations of this space consist of at most three projections into dimension 2, as shown in Fig. 4.

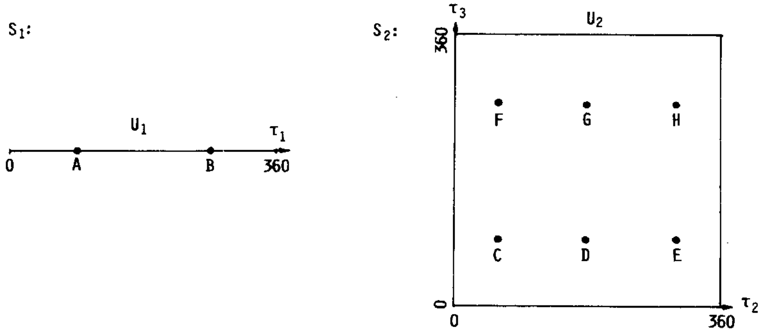

Fig. 2. Favored positions in the rotation spaces of S_1 and S_2, for molecule M

Table 1. Direct connection barrier heights for molecule M

3k:	AC–AG, AG–AE, BF–BD, BD–BH
6k:	AC–BC, AD–BD, AE–BE, AF–BF, AG–BG, AH–BH
10k:	all others

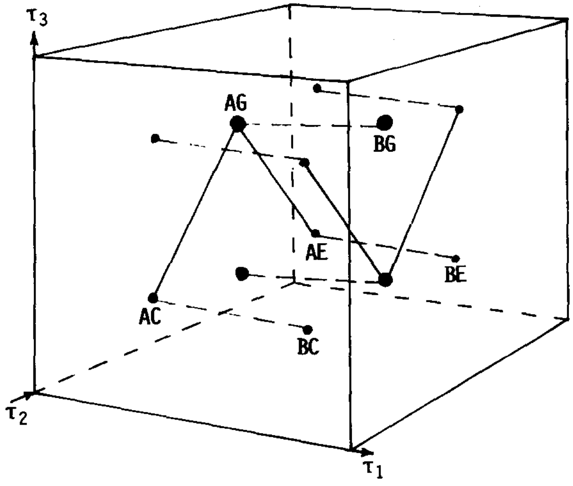

Fig. 3. Energy wells and connections for molecule M, (shown in perspective). Energy minima at 0 k: ●; at 2 k: •; Connections at 3 k: *solid*; at 6 k: *dashed*

The projections in Fig. 4 suggest that there is a direct path from BC to BG with barrier height 3 k, but this is not the case. The data in Table 1, from which the figures were constructed, show that the direct connection BC–BG has a barrier height of at least 10 k, while the minimal energy path is BC–AC–AG–BG, which has barrier height 6 k. Barriers such as this, which do not appear in coordinate projections such as Fig. 4, are called "hidden" barriers [10]. Their presence, which is always to be expected, constitutes the basic reason for devising and introducing new presentation methods.

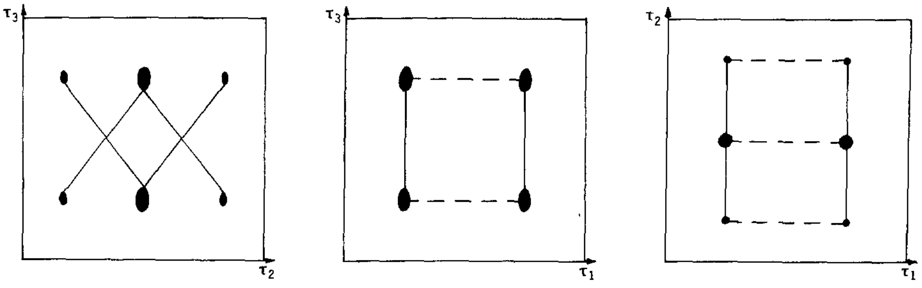

Fig. 4. Energy wells and connections for molecule M, shown in mutually orthogonal projections into dimension 2. Energy minima at 0 k: ●; at 2 k: •; Connections at 3 k: *solid*; at 6 k, *dashed*

This example was so constructed that Fig. 3 is easy to interpret, although it succeeds in showing the complexity of the whole rotation space with all significant wells and connections. That construction required finding an oblique projection in which all the wells and paths were resolved. Even then, all the energy contours had to be omitted, the energy information being carried by special symbols only. For the energy map of a real molecule with three rotation angles, the making of such a diagram is a formidable undertaking. With four rotation angles, the project fails completely and other methods must be employed [10].

For this example, the clarity of Fig. 3 renders unnecessary the dimension reduction. However, this technique is illustrated as follows: in Fig. 5, a 2-dimensional surface is shown embedded in the rotation space as a 2-dimensional submanifold. It virtually contains all the low-energy paths connected to BG and the wells which they connect. That is, it shows the connections of one component of the whole low-energy region (the other component being, by symmetry, like it). In Fig. 6a the surface is shown flattened out; here it is provided with the coordinate system shown by arrows. The corresponding arrows shown in Fig. 5 indicate what is actually a curvilinear coordinate system on the surface. That is, through each point of the surface there are two approximately perpendicular paths, one roughly parallel to the arc Γ_1 in U_1 which goes through A and B, the other roughly parallel to the arc Γ_2 in U_2, going through C, G, and E. The arcs Γ_1 and Γ_2 are defined in Fig. 6b, c. The surface closely approximates the cartesian product $\Gamma_1 \times \Gamma_2$. The wells and connections appearing in the surface constitute a graph with capacity [7], the nodes being the wells and the edges being connections.

A change in the symbols representing energy allows depiction of all possible energy values, from zero up to a cut-off energy. This representation is shown in Fig. 7, with cut-off energy 8 k. The result is referred to as a "balloon diagram" or *energy well connection graph*. The figure resembles a contour map showing six interconnected local minima and two isolated local minima. Only a single contour,

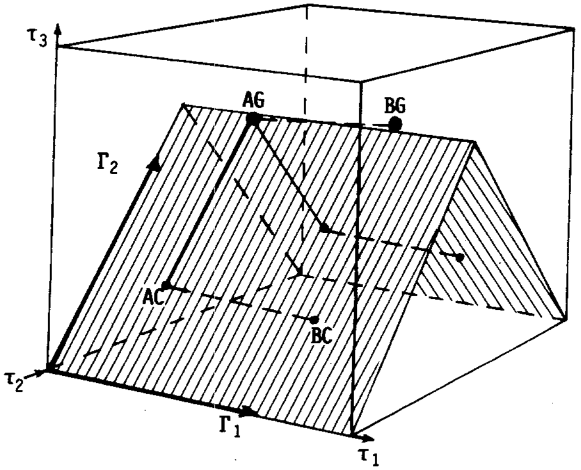

Fig. 5. The 2-dimensional surface embedded in $U_1 \times U_2$ as a 2-dimensional submanifold (molecule M)

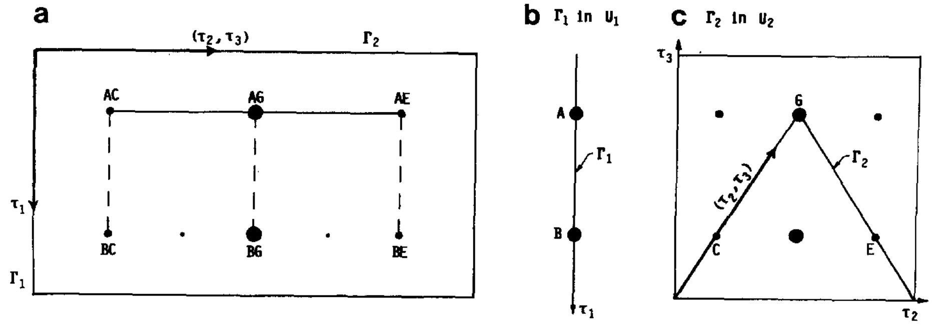

Fig. 6. a The 2-dimensional energy graph surface as developed in its own (curvilinear) coordinate space. b The curvilinear coordinates as defined in U_1, and c in U_2 (molecule M)

8 k, is given. It should be borne in mind that the only distances that have meaning in Fig. 7 are those depicting energy. The diameter across a disk (node) representing a well of minimum energy e is calculated $d = c(8 - e)$, where c is a scale factor. For a connecting bar (edge) the width, d, of the bar is calculated from the height of the barrier by the same formula, with the same value of c.

This example was chosen as introduction to the technique because Fig. 3 is sufficiently comprehensible to actually illustrate the embedding of the surface when redrawn as Fig. 5. For most real examples, no such figure will be available. However, the embedding was actually defined by Γ_1 and Γ_2, which are in turn defined in Fig. 6b, c. Thus, when the technique is applied in practice, the essential figures [11] will be the equivalents of Figs 1 and 6b, c, leading to the connection graph, which is equivalent to Fig. 7.

3. A real example: *m*-trifluoromethyl-N-ethyl-amphetamine (fenfluramine); reduction to dimension 2

The original report [10] on the conformational energies of fenfluramine contains contour maps showing some of the energy variation as a function of four variables.

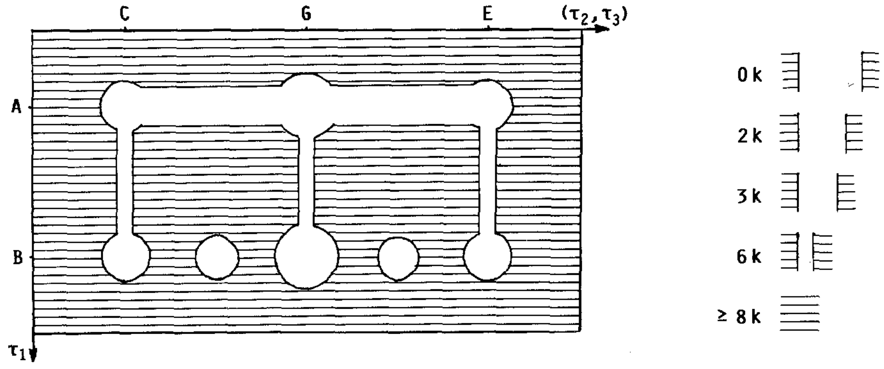

Fig. 7. Energy well connection graph for the hypothetical molecule M, shown as a graph with capacity in the form of non-metric contour map or "balloon diagram"

(See Fig. 8). The most important *in-vacuo*-energy well information given there will be summarized here in a 2-dimensional connectional connection graph, Fig. 10.

The four torsion angles (variables) are divided into two segments of two torsion angles each, (τ_1, τ_2) and (τ_3, τ_4). Of these, the first segment (pair) describes the connection of a chain to a ring, Y_1, τ_1 being the angle of the chain root relative to the plane of the ring. With respect to τ_1, the energy variation is, therefore, essentially periodic with period 180°. Thus, only half of the variation of τ_1 will be shown.

The second segment, (τ_3, τ_4), describes the configuration of the free outer end of a chain. Consequently, the map of energy as a function of τ_3 and τ_4 for fixed (τ_1, τ_2) has a very characteristic shape. For different choices of (τ_1, τ_2) the energies vary a great deal, but the locations of the wells (and to some extent the barriers) vary only little. As in the preceding example, the most prominent well locations are labelled with capital letters, A, B, and C. They are shown in Fig. 9b, a section, along with a path, Γ_2, which connects them relatively favorably for the choices of (τ_1, τ_2) that are of interest.

Any map of energy as a function of τ_1 and τ_2 for fixed (τ_3, τ_4) has also a characteristic shape. It shows one or more of three prominent energy wells, one for the molecule in the "extended" conformation, and two for the two "folded" conformations, whose locations vary little with the choice of (τ_3, τ_4). These wells we call E, F, and G. They are shown in the projection diagram Fig. 9a, along with a path connecting them, Γ_1. Paths near Γ_1 are relatively favorable connections among E, F, and G, no matter what the choice of (τ_3, τ_4).

The cartesian product of the paths Γ_1 and Γ_2 embeds a plane into $(\tau_1, \tau_2, \tau_3, \tau_4)$-space as a curved 2-dimensional submanifold. Very near it lie the wells EA, EB,

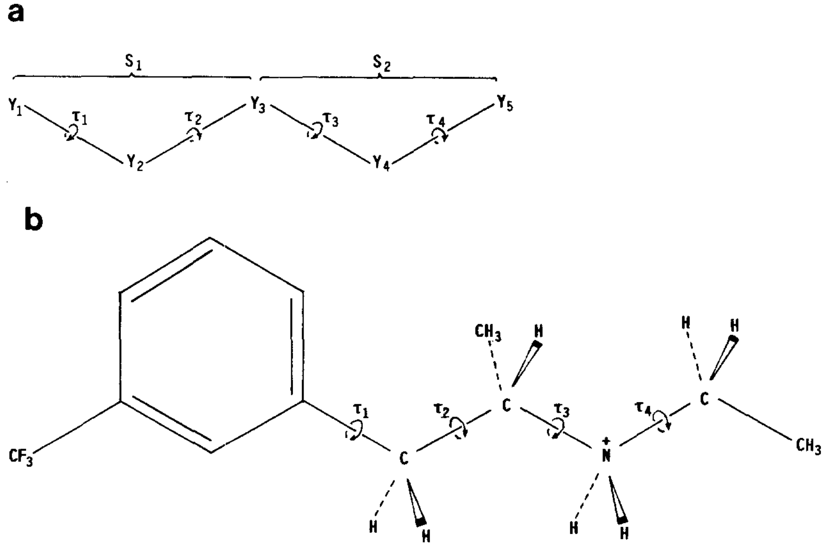

Fig. 8. **a** Segment structure of the fenfluramine molecule. **b** Molecular structure of fenfluramine

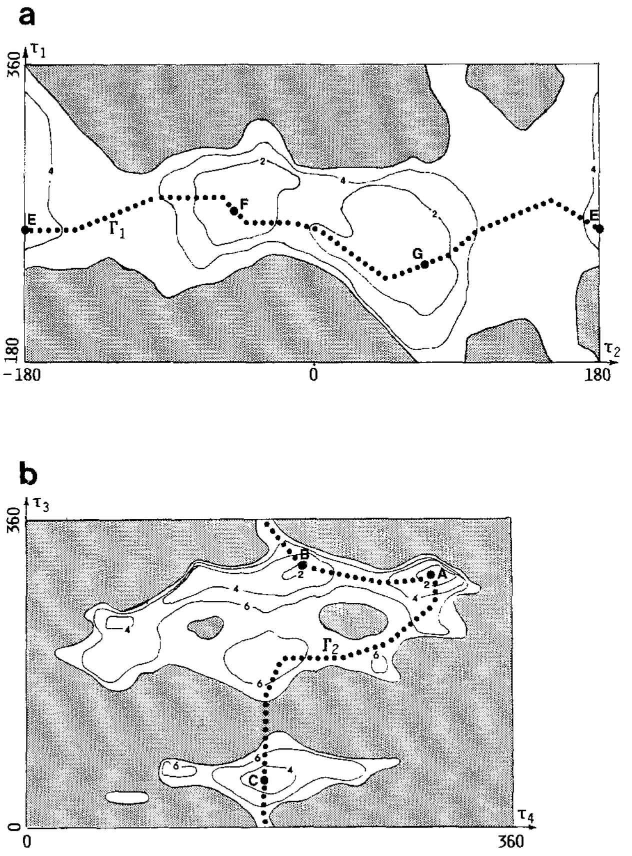

Fig. 9. Definition of the (curvilinear) coordinate paths Γ_1 and Γ_2 for fenfluramine. **a** Projection into (τ_1, τ_2)-space showing segment wells E, F, G, and path Γ_1 marked • • •. **b** Section for (τ_3, τ_4) taken through $(\tau_1, \tau_2) = (240°, 60°)$, showing segment wells A, B, C, and path Γ_2 marked • • •

EC, FA, FB, FC, GA, GB, and GC, and also the most important connections among them. For presentation purposes, we take the bottoms of these wells and the barriers of these connections to actually lie in the submanifold. In Fig. 10 the submanifold is shown in its own coordinate space (flattened out), and equipped with a 8 k energy contour (similar to that in Fig. 7).

The figure readily shows the lack of very low energy connections between chain conformations, even in the "extended" form, E. The barrier of at least 6 k in every direction from tthe global minimum, GA, is also apparent. Although the chain in conformations B or C can cross the ring from F to G at energies of 3 k, in conformation A it does not, even in the vicinity of the global minimum.

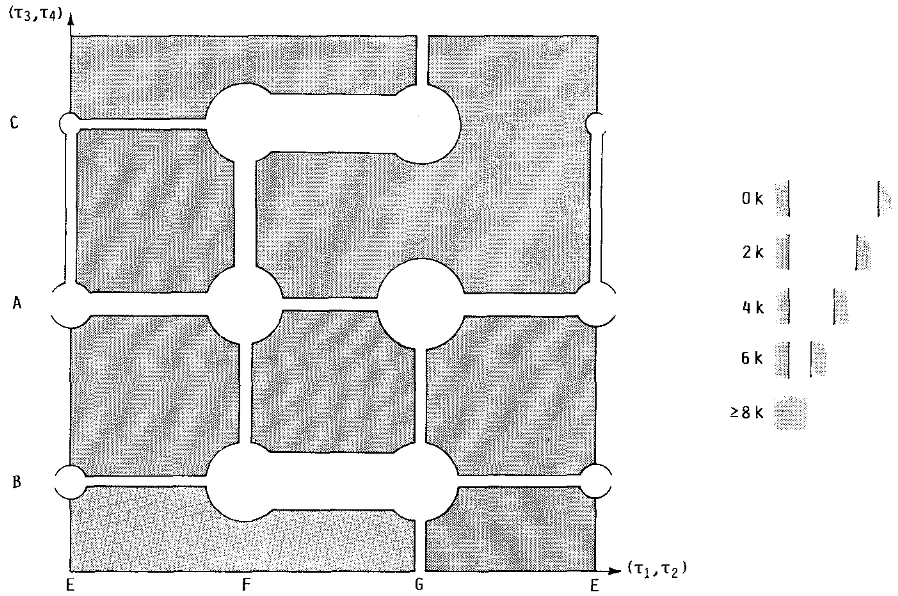

Fig. 10. Energy well connection graph for fenfluramine, showing 8 k energy contour. Curvilinear coordinates defined in Fig. 9

4. The major steps in preparing an energy well connection graph

1. Conformational analysis to find locations and energies of wells and barriers (connections).
2. Segmentation of the molecule.
3. Drawing the arcs (Γ) in the rotation spaces of the segments, determining the curvilinear coordinates and thus embedding the submanifold in the rotation space of the molecule.
4. Drawing the wells and connections (with their energies) on the coordinate grid for the submanifold, thus flattening the submanifold.

5. A more complex example: N-methyl-3-phenyl-3-(*o*-methoxyphenoxy)-propylamine(nisoxetine); reduction to dimension 3

The nisoxetine molecule has 7 rotation angles. One of these appears to have only one favorable value [12]. Hereafter it will be considered fixed rigidly at that value and ignored. The remaining rotation angles are numbered τ_1 to τ_6, as shown in Fig. 11. The molecule consists of three segments; the first is S_1, whose interest centers on the relationship between two aromatic rings, the rigid groups Y_1 and Y_4. U_1, the rotation space of S_1, has dimension 3. In U_1 there are six important

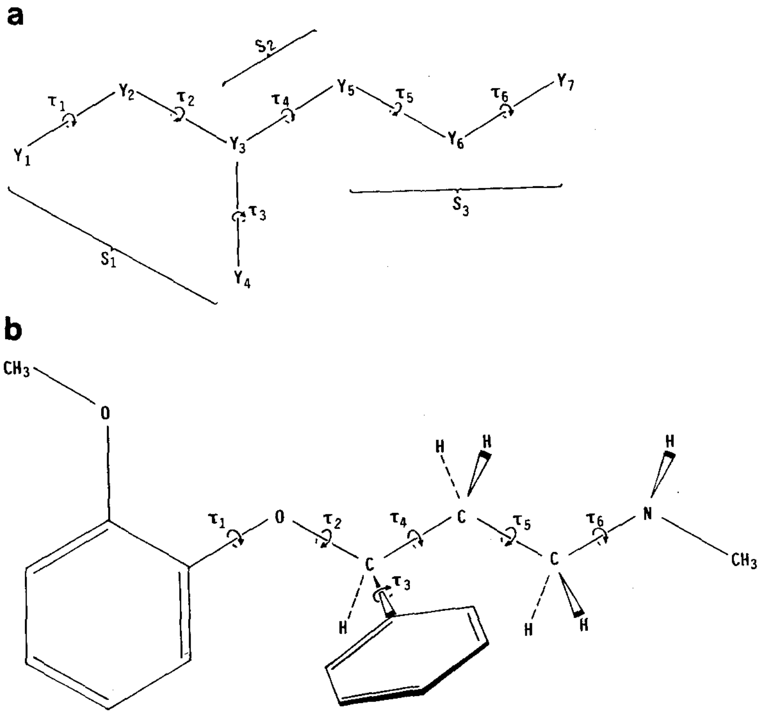

Fig. 11. a Segment structure of nisoxetine molecule. **b** Molecular structure of nisoxetine

Table 2. Energy well locations for nisoxetine, by segment

Segment	Variables	Approximate values	Conformation symbol
S_1	(τ_1, τ_2, τ_3)	(70°, 80°, 70°)	A
		(40°, 140°, 40°)	B
		(290°, 190°, 80°)	C
		(270°, 70°, 60°)	D
		(210°, 50°, 170°)	E
		(40°, 40°, 160°)	E′
S_2	τ_4	60°	60°
		180°	180°
		300°	300°
S_3	(τ_5, τ_6)	(60°, 60°)	P
		(300°, 120°)	Q
		(60°, 180°)	R
		(90°, 240°)	S
		(300°, 300°)	T

energy wells, which have been labelled, A, B, C, D, E, and E′ (see Table 2). The properties of E and E′ are very similar and both are separated from essentially all the other wells so that the curvilinear coordinate path ignores E′ but passes through the other five. Projection of U_1 into the (τ_1, τ_2)-plane resolves these wells and, for simplicity, only that projection of the path Γ_1 is given (Fig. 12a).

The sequence $\tau_5 - \tau_6$ defines a chain and it is designated as segment S_3 (leaving τ_4 to constitute S_2). Out of the nine expected energy wells in its 2-dimensional rotation space U_3, five are generally more prominent than the others, and Γ_3 is made to join these five along a path that is usually favorable: P, Q, R, S, T (Fig. 12c). (See Table 2).

There remains τ_4, the angle of the chain root. In some respects it is similar to τ_2

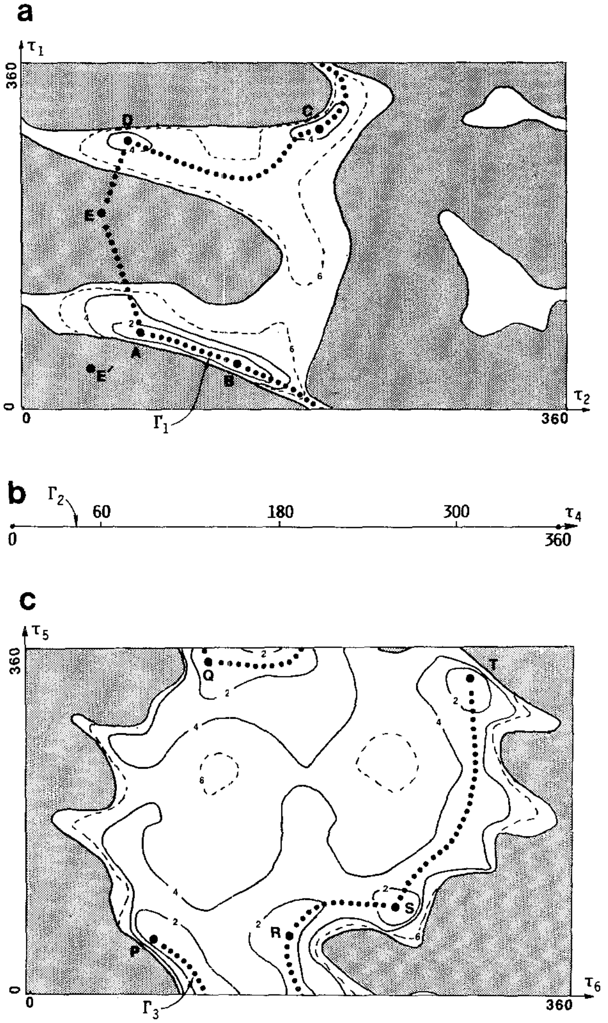

Fig. 12. Curvilinear coordinate paths for nisoxetine

in fenfluramine. It becomes segment S_2, with favorable locations 60°, 180°, 300° connected by Γ_2 (Fig. 12b). (See Table 2).

In the 6-dimensional space of nisoxetine, therefore, there are $5 \times 3 \times 5 = 75$ prominent locations for wells. Each well might connect with between 6 and 26 immediately adjacent wells, making between $6 \times 75/2 = 225$ and $26 \times 75/2 = 975$ possibly important barriers, of which 45 were found to have heights of less than 8 k. If the barriers were listed in a table, each entry would require values for seven variables: three coordinates $(\Gamma_1, \Gamma_2, \Gamma_3)$ for each of the two wells connected, and the barrier energy also. Even for the 45 significant barriers, such a table would be unwieldy and hard to comprehend. By contrast, the energy-well connection graph (Fig. 13) puts all the wells and connections into a pattern.

The connection graph provides an efficient method of presenting such information as the following:

The chain is rather flexible beyond the root for five choices of S_1 and S_2 conformation, specifically, for $\tau_4 = 180°$ or near the global minimum, which is the largest ball, low in the right rear at (A, 180°, P).

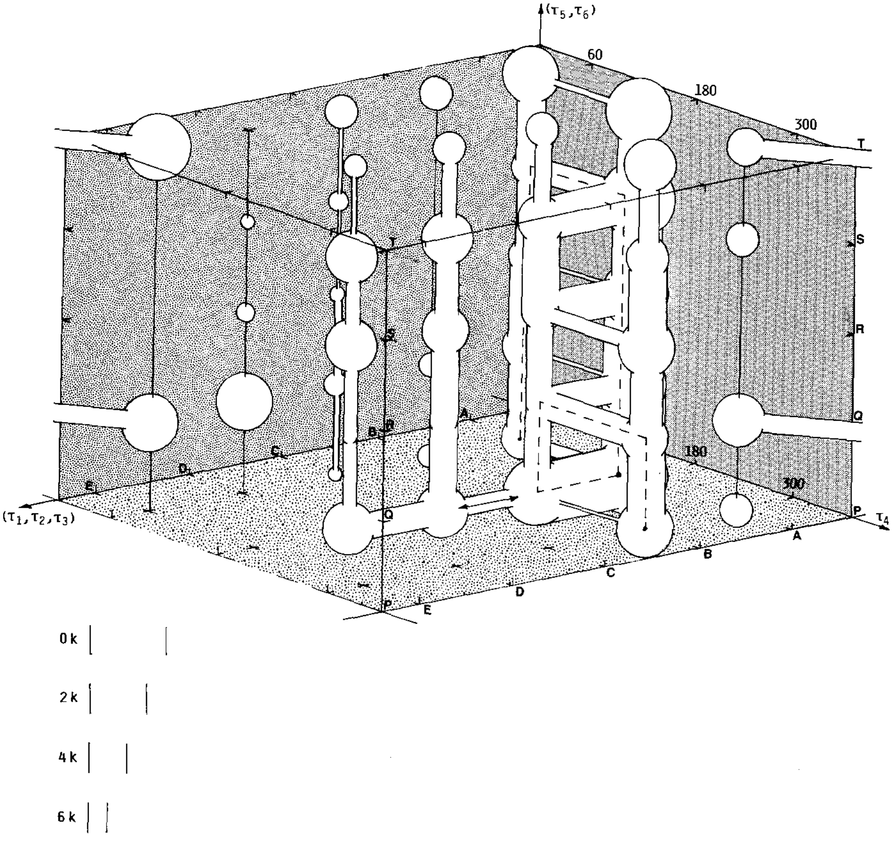

Fig. 13. Energy well connection graph for nisoxetine

The global minimum connects readily (even if not directly) with an adjacent energy well in each curvilinear coordinate direction. Since the two most favored ring positions, A and B, are actually fairly close and similar [12], chain root (S_2) and the outer chain (S_3) rotations provide most of the variety near the global minimum. From the global minimum (A, 180°, P) there is a possible τ_4 (S_2) rotation to (A, 60°, P), but not to (A, 300°, P). From the adjacent point (B, 180°, P) the possible τ_4 rotation goes in the opposite direction to (B, 300°, P) but not to (B, 60°, P). In each case the most favored path requires some accommodation in the outer chain. It is shown with a dashed line.

Looked at differently, the connection graph shows barrier heights. Between B and C there is a notable barrier (even for the only favorable S_2-S_3 configuration) whose low-energy crossing is marked ↔. This corresponds primarily to the rotation of the substituted ring Y_1 in the vicinity of the other ring, Y_4. (The axis of rotation is the τ_1 bond.) In Fig. 12a it is the passage from the lower half plane downward along the coordinate path Γ_1 to the top of the adjacent upper half plane. The barrier is about 6.5 k. Along the broad and direct route upward from B to C, in Fig. 12a, (not on the coordinate path) the energy barrier is slightly higher. Fig. 13 shows the lower energy path and (in this way) the lack of very low-energy connections between $0° < \tau_1 < 180°$ and $180° < \tau_1 < 360°$.

Overall, the connection graph shows very clearly the compact region of flexibility, and the single connection, marked ↔, into that small portion of the remainder of the rotation space in which there is still some chain flexibility.

Unlike the 2-dimensional energy-well connection diagrams in Figs. 7 and 10, any 3-dimensional diagram will itself have presentation problems. The view selected for Fig. 13 was chosen after all the data were collected. The global minimum was then placed relatively far from the viewer, where it could remain visible without obscuring smaller (high energy) features. The vertical connections (chain rotations) were resolved only just far enough to avoid obscuring vertical connections in the rear, while leaving in the clear the horizontal (S_1, S_2) connections around the global minimum.

6. Conclusion

In three examples, including two real molecules, the energy-well connection diagram is presented in the form of a graph with capacity. The connection graph shows wells and barriers that are significant because of low energy (up to 8 k). The wells, barriers, and energies are visible and readily comprehensible from the connection graph because of the analogue representation of the energy, and the reduction of dimension from that of the rotation (conformation) space to dimension 2 or 3.

Acknowledgments. This work was supported in part by U.S. Public Health Service Research Grants GM 22988 and HL 34193.

References

1. Olson WK, Flory PJ (1972) Biopolymers, 11:1
2. Flory PJ (1980) Pure Appl Chem 52:241
3. Hopfinger AJ, Battershell RD (1973) Adv Pestic Sci Plenary Lect Symp Pap Int Congr Pestic Chem 1987:196
4. For a contrast, see the study of shapes and barriers in [5]
5. Mezey PG (1977) Prog Theor Org Chem 2:127
6. Balaban AT (1976) Chemical applications of graph theory. Academic Press, London
7. Wilson RL, Beineke LW (1979) Applications of graph theory, Academic Press, London
8. Mezey PG (1982) Theor Chim Acta 62:133
9. Numerical values for energies indicated with the abbreviation k, are in kcal/mole above the global minimum (taken as 0 kcal/mole) for the system in question
10. Creese MW, Grunewald GL (1983) J Am Chem Soc 105:2463
11. In practice, Figs. 2, 3, 4, and 5 will be non-essential. The content corresponding to Fig. 2 may be given more conveniently by a description or table; views corresponding to Figs. 3 and 5 will be unobtainable; views corresponding to Fig. 4 are often obtainable, but must be handled with care because projections are misleading
12. Grunewald GL, Creese MW (1986) Drug Des Delivery 1:23; this paper includes ORTEP drawings of the sequence of favored ring orientations discussed below