Handling Cycles in Conformational Behaviour Visualization

Aleš Křenek

Faculty of Informatics
Masaryk University
Brno, Czech Republic
ljocha@fi.muni.cz

ABSTRACT

A simulation of conformational behaviour of molecules produces series of rapidly changing shapes of a particular molecule. Animated visualization of the process requires continuous morphing from one shape into another. Recently we proposed a method applicable on acyclic structures, in this paper we focus on smooth interpolation between different shapes of cycles.

Keywords: interpolation, conformational behaviour, cyclic structures

1 Introduction

Being an important motivation of computers' power growth, computational methods became popular in chemistry. Studying conformational properties (i. e. a capability of a molecule to change shape while no chemical reaction takes place) of medium to large sized biomolecules (peptides, proteins, enzymes, nucleic acids etc.) became a significant area of computational chemistry as analyzing those properties allows a better understanding of further biological activity of the molecules.

Extensive chemical calculations carried by methods called together *Travelling through conformational space* [Koča98] produce, in general, sequences of changing *geometries*, i. e. different 3-dimensional shapes of the particular molecule. The calculation does not output a continuous sequence, it stores only certain significant geometries (local energy minima and saddle points). Those geometries themselves are sufficient to describe the behaviour of the molecule, however, interpreting the sequence is still a non-trivial task.

Figures 1 and 2 show a representative example of the data involved (a DPDPE peptide, [Kříž99]). The images represent two consequent points in the generated sequence with no inbetweens directly available. Even if an experienced user faces such a sequence a lot of effort is required to understand what happened exactly in the single step. Therefore the principal task of the advanced visualization of the process is to provide the user with

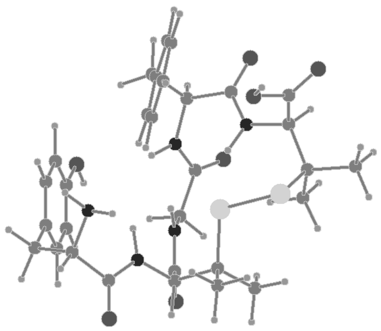

Figure 1: Example conformation of DPDPE peptide

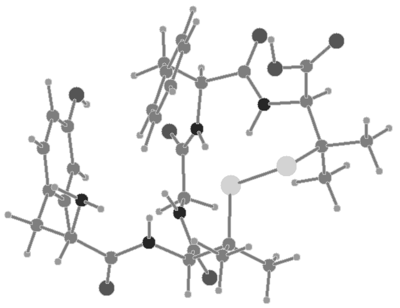

Figure 2: Subsequent conformation of DPDPE

an animated smooth 3D morphing between two given geometries. However, this process must not contradict chemical intuition, i. e. arbitrary deformations are not allowed.

Despite it has been reported recently that a similar chemical calculation can be optimized so that it is possible to run it on a powerful hardware on-line [Ai98], it is still very exhaustive computation and making the particular chemical software, that our users mostly use, run in this way would be rather hard. Therefore it makes sense to employ a heuristic, pure geometrical approach to the shape interpolation.

2 Basic Terms

The "chemical intuition" which we are trying not to contradict to make the visualization convincing, is mostly based on simple energetic constraints. In order to describe those more easily we have to introduce an *internal coordinate* system frequently used for chemical calculation. It defines the following classes of coordinates:

- *Bond length* is the distance between two atoms which are chemically bound together. It is usually measured in Ångstroms ($1\,\text{Å} = 10^{-10}\,\text{m}$).
- *Binding angle* is defined as the angle $\measuredangle ABC$ where A, B, C are atoms and there exist bonds AB and BC.
- *Torsion angle* can be defined roughly as the angle between planes ABC and BCD where A, B, C, D are atoms and there exist bonds AB, BC, CD.

  More precisely, taking the exact orientation in account, the absolute value of the angle is given by the angle between cross-products $BA \times BC$ and $CB \times CD$ (interpreting ordered pairs of atoms as vectors) and it is positive iff the system BA, BC, CD is right-handed.

Given Cartesian coordinates of all the atoms all the internal coordinates can be computed. On the other hand, it can be shown that for a molecule of n atoms a careful choice of $3n - 6$ internal coordinates (including all the above classes) is sufficient to describe the geometry uniquely modulo its placement and orientation in space.

A relative stiffness (in other words, an energy required to change the coordinate) of the three classes of internal coordinates is significantly different with respect to the conformational behaviour:

- Bond lengths are very hard to modify. Actually, extending a bond length significantly means breaking the bond, and initiating a chemical reaction then. In general, initiation energies of reactions are several orders of magnitude higher than those of conformational behaviour on the same molecule.
- Binding angles are slightly more flexible. Therefore it is possible they change during conformational behaviour. However, the associated cost is quite high so the change of these coordinates is usually small.
- Torsion angles are, in many cases, free to change significantly. Therefore conformational behaviour mostly involves changes in torsion angles.

Statistical analysis of data describing conformational behaviour of some representative molecules confirm the characteristics above: changes in bond lengths are usually far bellow 5 %, bond angles may change by about

10 %, and, on contrary, a change of as much as π in a torsion angle is not an exception.

From here we can derive more exact requirements on the interpolation procedure between two consequent steps of conformational behaviour. These requirements more or less reflect the intuition we referred to:

1. Bond lengths should not visibly change at all.
2. Binding angles may change slightly — the user can accept it.
3. Torsion angles can change whenever required — the user even expects it.

Finally, there is one more general requirement with no direct relation to chemical aspects:

4. The change of any of the internal coordinates should be continuous in time, monotonous and preferably as close to linear as possible.

We still assume that the molecular geometries given by the chemical calculation are good key-frames with respect to the four requirements given above, i. e. whenever an interpolation between a pair of geometries follows the requirements it is visually acceptable (for example, does not contain artifacts like mutual penetration of bonds). This assumption is fulfilled by our data. The chemical calculation stores exactly such geometries.

3 Interpolation

3.1 Previous Work

Our problem belongs to the family of key-frame animation problems. It is possible to express the entire set of internal coordinates of a particular molecule in terms of rotations and scaling uniquely. Thus the particular shape of the molecule can be described by a set of quaternions. Known key-frame interpolation techniques derived from [Shoem85] could be applied then. Moreover, we would be able to express certain constraints (like not breaking a cycle) in a very comprehensive way as quaternion equations. However, the general advantage of this approach, i. e. unified handling of all the angles, turn into a drawback in our situation. Because of the chemical reasons described above we have to handle the two classes of angles in different ways. Therefore we propose a new interpolation technique.

3.2 Acyclic Structures

We have already proposed an interpolation method for acyclic structures [Křenek97]. In the following we describe it here very briefly, for completeness of the paper only.

The algorithm is based on breaking the acyclic molecule into fragments. In a given time t ranging from 0 (initial conformation) to 1 (final conformation) we interpolate the bond lengths linearly first. If interpolated the change in a bond length is almost unobservable, however, if ignored it may produce a visible artifact. Then for each atom an interpolated shape of it and its direct neighbourhood is computed by augmented linear interpolation on a unit sphere. This results in an interpolation of binding angles. As the difference between initial and final values of each binding angle is not very big this interpolation meets the fourth requirement from the previous section. Finally an interpolated averaged torsion angle is computed for each applicable bond (corresponding to the BC bond in the above definition of torsion angles). The interpolated geometry is reconstructed from the fragments then. It can be proven that, under certain regularity conditions, the reconstruction is unique.

3.3 Cycle Shape Interpolation

Unfortunately the method described in the previous section fails on cyclic structures. Let's consider a cycle $A_1, A_2, \ldots, A_m$. Trying to apply the acyclic approach we have to ignore one bond of the cycle, say

$A_m A_1$. Therefore we shall not linearly interpolate its bond length, binding angles $\measuredangle A_{m-1}A_mA_1, \measuredangle A_mA_1A_2$, and torsion angles $\measuredangle A_{m-2}A_{m-1}A_mA_1, \ldots, \measuredangle A_mA_1A_2A_3$. Ignoring exactly six internal coordinates corresponds to the fact that a geometry of such a cycle can be uniquely described by $3m-6$ internal coordinates.

On almost all real data the interpolation produces visible artifacts, mainly violating the requirement on non-changing bond lengths. Therefore we developed a symmetric method that took in account all the internal coordinates of the cycle equally.

Our goal is, according to the fourth general requirement (Sect. 2), keeping the interpolation of all $3m$ internal coordinates as close as possible to a linear function. However, as stated above a geometry of the cycle is described uniquely by only $3m-6$ internal coordinates. Therefore it is not possible, in general, to reconstruct a cycle geometry that reflects a complete set of arbitrarily given internal coordinates. So in each interpolation step (a value of the time parameter t) we have to solve a minimization problem on a distance of a particular geometry to a desired but unrealizable complete set of internal coordinates given by a linear interpolation. Besides driving the resulting geometry of the particular step close to the ideal one (the 4^{th} requirement) the minimization should implement the first three requirements from Sect. 2. In the rest of this section we discuss design of a suitable distance function to be minimized.

We define the set of independent variables to be the Cartesian coordinates x_{ij} ($i = 1 \ldots m, j = 1, 2, 3$) of the cycle atoms. With respect to this set let l_i denote the bond length A_iA_{i+1} (also l_m denoting the length A_mA_1), ϕ_i the binding angle $\measuredangle A_{i-1}A_iA_{i+1}$, and ψ_i the torsion angle $\measuredangle A_{i-1}A_iA_{i+1}A_{i+2}$. Moreover, let $l_i^*, \phi_i^*, \psi_i^*$ be the ideal linearly interpolated values of the respective internal coordinates.

We specify the objective function to be minimized by the following formula

$$\sum_{i=1}^{m} \big(W_l(l_i - l_i^*)^2 + W_\phi(\phi_i - \phi_i^*)^2 \\ + W_\psi(\psi_i - \psi_i^*)^2 \big) \quad (1)$$

where W_l, W_ϕ, W_ψ represent weights of the three classes of internal coordinates. According to the requirements from the Section 2 the objective function should strictly penalize any deviation in bond length, it should tolerate small differences in binding angles, and it should allow free torsions (but still drive the geometry close to $l_i^*, \phi_i^*, \psi_i^*$). Therefore the weights must fulfill $W_l \gg W_\phi > W_\psi$ in general. We have not finished analyzing a sufficiently representative set of data yet, but probably we can consider $W_l = 20, W_\phi = 2, W_\psi = 1$ to be acceptable.

In experimental implementation we perform the minimization by a quasi Newton method [Gill78] implemented by the NAG library [NAG] using a Jacobian computation code automatically generated by the tool AdiC [Bisch97]. In order to achieve better convergence properties of the minimization process the objective function has to be extended by two rather technical terms.

First, the expression (1) is invariant to a rotation and translation of all the independent variables as a whole. Thus at any point of the minimization domain ($3m$-dimensional space) there exists a six-dimensional subspace of directions where the objective function is constant. This results in an ill-conditioned minimization problem. In order to improve performance of the minimization we let the minimized geometry be slightly attracted by a particular geometry. The choice of this geometry (y_{ij} in the formula below) is not very important, we use the starting point of the minimization which seems to add some little more stability. The attraction is implemented by extending the objective function to the following

$$\sum_{i=1}^{m} \big(W_l(l_i - l_i^*)^2 + W_\phi(\phi_i - \phi_i^*)^2 + W_\psi(\psi_i - \psi_i^*)^2\big) + \sum_{i=1}^{m}\sum_{j=1}^{3} W_x(x_{ij} - y_{ij})^2 . \qquad (2)$$

The weight W_x should not shift the minimum of (2) too far from the minimal region of (1) therefore $W_x \ll W_\psi$. However, experiments show that we get best results with approximately $W_x = 1/2$. This can be explained probably by the orthogonality of the subspaces where the two principal terms of Eq. 2 are constant.

In some cases where the distance between the cycle geometries of the initial and final conformations is rather big the objective function becomes very rich in critical points. Therefore the choice of a good starting point of the minimization becomes rather hard. One solution is running the interpolation in very fine steps of the time variable t. However, this introduces a lot of otherwise unnecessary calculations and is usually computationally infeasible. Moreover, if in a rare ill-conditioned situation the minimization fails there is no starting point available for the next interpolation step.

A better approach (in fact a continuous version of the previous one) is introducing t as another independent variable. For the sake of consistency, we shall re-label the former t (i. e. the desired value of time for this interpolation step) to a new notation t^*. The quantities l^*, ϕ^*, ψ^* in formulae (1), (2) become functions of the new independent variable t now. We arrive at the final form of the objective function

$$\sum_{i=1}^{m} \big(W_l(l_i - l_i^*(t))^2 + W_\phi(\phi_i - \phi_i^*(t))^2 + W_\psi(\psi_i - \psi_i^*(t))^2\big) + \sum_{i=1}^{m}\sum_{j=1}^{3} W_x(x_{ij} - y_{ij})^2 + mW_t(t - t^*)^2 . \qquad (3)$$

We strictly require the minimization reaches the desired time t^* therefore the corresponding weight must be sufficiently high, currently we use $W_t = 50$. Multiplication by the size of the cycle m ensures the influence of this term to be independent on the size of the cycle.

If there is only a single cycle in the molecule or the cycles are mutually independent the interpolated cycle geometry can be included in the acyclic algorithm in a straightforward way. Let's assume there exists a bond A_iB for some atom A_i on the cycle and B outside the cycle. Then an acyclic interpolated geometry of the fragment $A_{i-1}, A_i, A_{i+1}, B, \dots$ (assuming there may be other atoms bound to B) is computed independently on the cycle. It is moved so that A_i's coordinates match exactly on the cycle and rotated so that A_{i-1} and A_{i+1} give minimal residual (see [Křenek96] or [Curti94] for details of this computation).

4 Results and Limitations

Figs. 3–5 show the 64-steps evolution of all the cycle internal coordinates between geometries in Fig. 1 and 2 (as the cycle consists of 14 atoms there are 14 coordinates of each class). It can be seen that the requirements specified in Sect. 2 are met. Bond lengths are very close to constant all the time. The binding angles that have to change evolve almost linearly. Those that should remain constant may slightly change in the middle of the interpolation. As stated above this is the acceptable cost for the constant bond lengths and linear evolution of torsion angles.

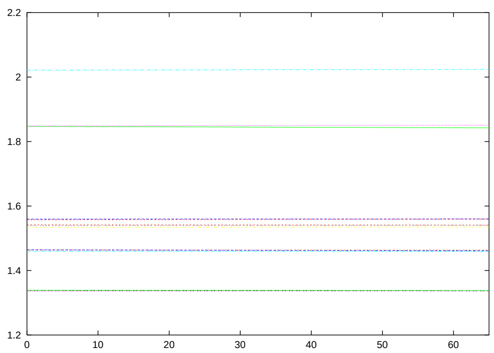

Figure 3: Bond lengths evolution

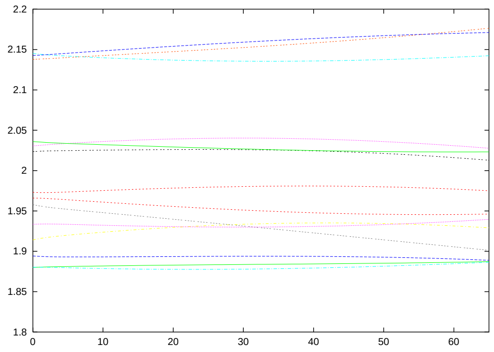

Figure 4: Binding angles evolution

As figure 5 shows the evolution of torsion angles is very close to linear (notice the different vertical scales of Fig. 4 and 5).

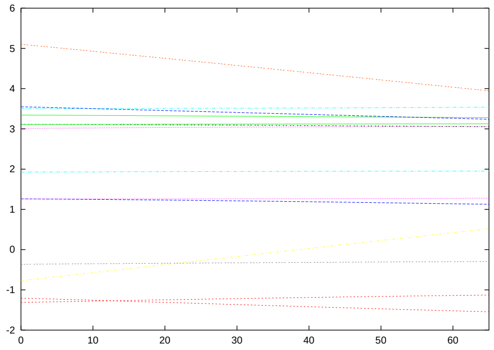

Figure 5: Torsion angles evolution

The method as proposed in the previous section is applicable on a single cycle or on more cycles which share at most one atom. Otherwise, the interpolated geometries of the cycles cannot be computed independently. A multiple cycle extension of the method would be straightforward probably. However, in the data we have access to, mutually dependent cycles are solely aromatic structures which are very stable so that no conformational changes occur on those cycles. Therefore we have no experience with fine tuning the objective function for this case.

5 Conclusions

Continuous interpolation between consequent but rather different steps of conformational behaviour of molecules adds significant value to the visualization of the process. The user can observe the behaviour and visually extract its principal features more easily from a continuously animated visualization.

The part of the interpolation method which handles cyclic structures properly was presented in this paper. The method involves relatively simple geometric calculations only which can be performed in real time (unlike computationally exhaustive chemical calculations).

We formalized several fairly simple geometric requirements which the interpolation should follow in order to be convincing. We have shown that our method is designed to meet those requirements. Evaluation on a testing set of real data looks promising, we get smooth animation with no visible artifacts or asymetric behaviour. The more formal evaluation of internal coordinate evolution confirms the observed results. In this paper we presented a representative example.

Capability of proper cycle handling makes this visualization of conformational behaviour complete — we can handle virtually all reasonable data now.

REFERENCES

[Ai98] Z. Ai and T. Frölich. Molecular dynamics simulation in virtual environments. *J. Eurographics Assoc.*, 17(3):267–273, 1998.

[Bisch97] C. Bischof, L. Roh, and A. Mauer-Oats. ADIC: an extensible automatic differentiation tool for ANSI-C. Technical Report ANL/MSC-P626-1196, Argonne National Laboratory, 1997.

[Curti94] W. Dan Curtis and Karel Zikan. Singular-value decomposition, optimal transformations and applications. Technical Report BCSTECH 94-038, Boeing Computer Services, Research and Technology, 1994.

[Gill78] P. E. Gill and W. Murray. Algorithms for the solution of the nonlinear least-squares problem. *SIAM J. Numer. Anal.*, 15:977–992, 1978.

[Kříž99] Z. Kříž, P. H. J. Carlsen, and J. Koča. Conformational features of linear and cyclic enkephalins. A computational study. *J. Comp. Aid. Mol. Design*, 1999. in press.

[Koča98] J. Koča. Traveling through conformational space: an approach for analyzing the conformational behaviour of flexible molecules. *Progress in Biophys. and Mol. Biol.*, 70:137–173, 1998.

[Křenek96] Aleš Křenek. Molecular models superposition and visualization. In V. Skala and N. Magnenat Thalmann, editors, *Proc. WSCG'96*, volume 2, pages 271–277, 1996.

[Křenek97] Aleš Křenek. An algorithm on interpolating between two shapes of molecule. In W. Straßer, editor, *Proc. SCCG '97*, pages 77–84. Comenius University, Bratislava, 1997.

[NAG] NAG Fortran library. http://www.nag.com/numeric.html.

[Shoem85] Ken Shoemake. Animating rotation with quaternion curves. In B. A. Barsky, editor, *Computer Graphics (SIGGRAPH '85 Proceedings)*, volume 19, pages 245–254, July 1985.