

# Exploring Biological Electron Transfer Pathway Dynamics with the *Pathways* Plugin for VMD

Ilya A. Balabin,<sup>[a,b]\*</sup> Xiangqian Hu,<sup>[a]</sup> and David N. Beratan<sup>[a,c,d]</sup>

We describe the new *Pathways* plugin for the molecular visualization program visual molecular dynamics. The plugin identifies and visualizes tunneling pathways and pathway families in biomolecules, and calculates relative electronic couplings. The plugin includes unique features to estimate the importance of individual atoms for mediating the coupling, to analyze the

coupling sensitivity to thermal motion, and to visualize pathway fluctuations. The *Pathways* plugin is open source software distributed under the terms of the GNU's Not Unix (GNU) public license. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.22927

## Introduction

Biological electron-transfer (ET) reactions are essential for the conversion of energy from food or sunlight into the universal energy currency, adenosine triphosphate.<sup>[1–4]</sup> ET is performed within and between molecules, where tunneling may be mediated in part by water.<sup>[1,2]</sup> ET may also be coupled to proton or other ion transfer processes, or may be gated by conformational change.<sup>[5–7]</sup> Because of relatively large donor-to-acceptor distances (5–25 Å), these reactions typically occur in the weak coupling (or nonadiabatic) limit, where tunneling is mediated by virtual electronic states of the intervening medium (the bridge), and the reaction rates are described by Fermi's golden rule (high-temperature limit)<sup>[8]</sup>

$$k_{\text{ET}} = \frac{2\pi}{\hbar} \frac{1}{\sqrt{4\pi\lambda k_{\text{B}}T}} |T_{\text{DA}}|^2 \exp\left[-\frac{(\Delta G + \lambda)^2}{4\lambda k_{\text{B}}T}\right], \quad (1)$$

where  $\Delta G$  is the reaction free energy,  $\lambda$  is the reorganization energy, and  $T_{\text{DA}}$  is the electronic donor-to-acceptor tunneling coupling mediated by the intervening molecule(s). We focus on the electronic coupling  $T_{\text{DA}}$  that depends on the structure of the medium between the donor and the acceptor.

Biological ET reaction rates are known to decrease approximately exponentially with tunneling distance, with the exponential factor specific to the tunneling medium structure.<sup>[9,10]</sup> While proteins present lower barriers than water, and water is a better mediator than empty space, the exponential decay factors for different proteins are substantially different.<sup>[9,10]</sup> The origins in 3D structures of these electronic coupling differences were first addressed by the *Pathways* model.<sup>[11]</sup> This model is based on the fact that the energy barrier for tunneling through covalent bonds is lower than for tunneling through hydrogen bonds, and the latter barrier is lower than for tunneling through vacuum.<sup>[11,12]</sup> As such, the *Pathways* model treats tunneling propagation as a sequence of steps, each of which is mediated by a covalent bond, a hydrogen bond, or vacuum, and the electronic coupling is proportional to a product of penalties for each step:

$$T_{\text{DA}} \propto \prod_i \epsilon_i^{\text{C}} \prod_j \epsilon_j^{\text{H}} \prod_k \epsilon_k^{\text{S}}. \quad (2)$$

The *Pathways* strategy uses graph theory to search for the sequence of steps that maximizes the product in eq. (2). In most parameterizations, the penalty for a covalent bond-mediated step  $\epsilon^{\text{C}} = 0.6$  is based on experiments of Closs and Miller,<sup>[13]</sup> the penalty for a through-space jump is  $\epsilon^{\text{S}} = \epsilon^{\text{C}} \exp[-\beta^{\text{S}}(R^{\text{S}} - 1.4)]$ , where  $R^{\text{S}}$  is the jump distance in Å, and the penalty for a hydrogen bond-mediated step is  $\epsilon^{\text{H}} = (\epsilon^{\text{C}})^2 \exp[-\beta^{\text{H}}(R^{\text{H}} - 2.8)]$ , where  $R^{\text{H}}$  is the hydrogen bond length (heavy atom to heavy atom) in Å.<sup>[11]</sup> Early estimates of the decay factor for empty space  $\beta^{\text{S}} = 1.7 \text{ \AA}^{-1}$  were later reduced to  $1.1 \text{ \AA}^{-1}$  to reflect typical 5 eV binding energies of redox cofactors.<sup>[14]</sup> With this parameterization, the *Pathways* model has successfully described ET reactions in many redox proteins.<sup>[3,9–11]</sup>

Despite its simplicity, the *Pathways* model is broadly used to identify coupling pathways and to estimate electronic coupling interactions. Although semiempirical<sup>[15,16]</sup> and *ab initio*<sup>[17,18]</sup> quantum chemistry methods may yield more accurate result when proper ensemble averaging is included, the *Pathways* model-based calculations are several orders of magnitude faster and provide an approach for identifying mediating residues and for estimating relative coupling interactions in large and complex biomolecular systems.<sup>[19–21]</sup> An example of how such calculations may remarkably well capture the electronic coupling decay with distance in the blue copper protein azurin is shown in Fig. 6 in ref. 10. Factors that influence the accuracy of the *Pathways* model are discussed in more details in ref. 3. In ET systems where tunneling is sensitive to thermal motion, the coupling that enters eq. (1) is the mean squared coupling. Indeed, early studies indicated that interactions computed

[a] I. A. Balabin, X. Hu, D. N. Beratan  
Department of Chemistry, Duke University, Durham, North Carolina 27708  
E-mail: [ilya.balabin@duke.edu](mailto:ilya.balabin@duke.edu)

[b] I. A. Balabin  
Lockheed Martin, Research Triangle Park, North Carolina 27709

[c] D. N. Beratan  
Department of Physics, Duke University, Durham, North Carolina 27708

[d] D. N. Beratan  
Department of Biochemistry, Duke University, Durham, North Carolina 27708

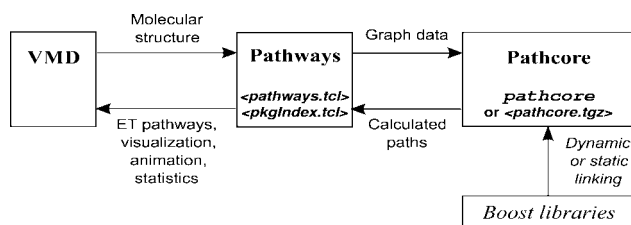
Contract/grant sponsor: NIH; Contract/grant number: GM-048043

© 2012 Wiley Periodicals, Inc.

with *Pathways* and multipath methods depended strongly on geometry because of fluctuations in through-space interaction strengths.<sup>[16,19,22]</sup> Recent studies of water-mediated coupling support this notion.<sup>[20,21,23]</sup> Structural fluctuations in biomolecular systems, including dynamics of transient water molecules that may mediate strong ET pathways,<sup>[23,24]</sup> are often sampled by all-atom classical molecular dynamics simulations. Currently available software is not capable of performing *Pathways* calculations along molecular dynamics trajectories without substantial modifications. As such, the development of software to compute and to analyze pathway fluctuations is timely, as we expand our understanding of biological ET reactions to regimes where protein dynamics plays a central role. The *Pathways* plugin for the molecular analysis and visualization program visual molecular dynamics (VMD)<sup>[25]</sup> was developed to provide these capabilities.

## Implementation

As shown in Figure 1, the *Pathways* plugin includes Tcl scripts `pkgIndex.tcl` and `pathways.tcl` as well as a binary executable `pathcore.pkgIndex.tcl` is a standard component of Tcl packages that informs the Tcl interpreter built into VMD about the existence and location of the *Pathways* plugin (<http://www.ks.uiuc.edu/research/vmd>). `pathways.tcl` implements key functionality of the plugin: analyzing the loaded molecular structure, preparing data for graph searches, analyzing graph search results, visualizing the pathways, animating the pathway dynamics, and calculating statistical descriptors of the electronic coupling fluctuations. This script includes two key functions, `pathways`, which implements pathway analysis for a single biomolecular conformation, and `pathtraj`, a wrapper for analyzing pathway dynamics over a molecular dynamics simulation trajectory that calls `pathways` for each trajectory frame. In addition, this script includes functions `addbond`, `delbond`, and `setseg` that control the list of covalent bonds, brief embedded help for all user-accessible functions, and numerous auxiliary functions used by `pathways` and `pathtraj` internally. The program `pathcore` performs graph searches using data generated by the script `pathways.tcl` and returns the results to the script. Its code, which is essentially based on Yan Qi's *K*-shortest path code (<http://code.google.com/p/k-shortest-paths>), is written in C++ and linked to Boost libraries (<http://www.boost.org>) that provide the Dijkstra strongest pathway search algorithm.<sup>[26]</sup>



**Figure 1.** Implementation of the *Pathways* plugin. Script `pkgIndex.tcl` provides integration of the *Pathways* plugin with VMD. Script `pathways.tcl` is the key component of the plugin that provides all its functionality except graph search. The latter task is performed by program `pathcore`, which is available for several popular platforms as a statically compiled binary executable. For other platforms, it has to be compiled from source code (provided as a compressed archive `pathcore.tgz`) and linked to Boost libraries.

The *Pathways* plugin is controlled by commands entered in the VMD command line interfaces, that is, the VMD console or the Tk console that provides additional command history and editing capabilities. The plugin output is printed in the same console. The plugin queries VMD to identify whether it runs in the graphics mode or in the text-only mode. In the graphics mode, the plugin visualizes identified pathways and may render 24-bit images using one of the VMD-compatible renderers. For molecular dynamics trajectories, the plugin calculates and renders pathways for each trajectory frame and may convert the images into MPEG2 video using the open-source utilities `convert` from Image Magick (<http://www.imagemagick.org> (for UNIX/Linux), <http://imagemagick.sourceforge.net/http/www/windows.html> (for Windows)) and `ffmpeg` from FFmpeg (<http://ffmpeg.org> (for UNIX/Linux), <http://ffmpeg.zeranoe.com/> (for Windows)). The plugin can also be run in a noninteractive mode, reading commands from a script file and writing its output into a log file.

Like any other Tcl plugin, *Pathways* needs to be loaded into VMD before using. Loading is performed by using the standard Tcl command `package require pathways`. Since the plugin receives hydrogen-bond information from VMD, hydrogen atoms need to be added to the Protein Data Bank (PDB) files (<http://www.pdb.org>).<sup>[27]</sup> The plugin also receives covalent-bond information from VMD, which VMD obtains from a molecular structure file (e.g., a PSF file), if one was loaded. Otherwise, VMD attempts to identify covalent bonds using atom types and interatomic distances. Since such an identification of bonds is only reliable for common biological structures (proteins, nucleic acids, and lipids), and not for many cofactors, the *Pathways* plugin provides commands that add or delete covalent bonds and set segment IDs in the loaded molecule as needed. The scope of these commands is limited to the *Pathways* plugin, and they do not alter the covalent-bond information that VMD generates or reads from a structure file.

## Capabilities

The *Pathways* plugin allows exploration of how the tunneling medium structure and dynamics influence the mean squared electronic couplings. The plugin includes the following capabilities:

1. Identify, visualize, and animate dominant ET pathways, and estimate relative electronic couplings mediated by each pathway.
2. Estimate the importance of individual atoms to ET pathways and visualize relevant atoms.
3. Analyze the statistics of electronic coupling fluctuations over a molecular dynamics trajectory.

The *Pathways* plugin accepts parameters for the ET-mediating molecules, *Pathways* model parameters, output parameters, and graphics parameters. Only two parameters are required: atom selection strings for the electron donor and acceptor atoms. Other parameters are preset to default values but can be readily changed. When starting pathways calculations, the plugin prints out the parameters as follows:

PATHWAYS) Starting pathway calculations:

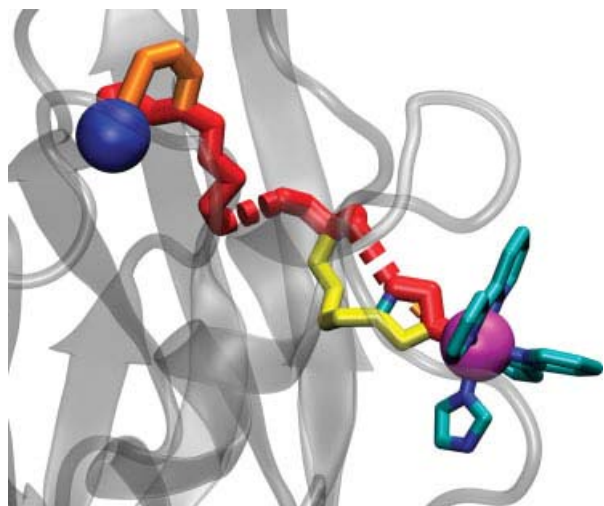
```
MOLECULE:
Molecule:          ionized.psf
Trajectory frame:   0
Donor selection:    name CU
Acceptor selection: name RU
Bridge selection:   protein or rename RBU
Number of pathways: 100
Atom importance:    0
Environment size:   4.0
Collective D/A:    1
```

```
PATHWAYS:
Include hydrogens: 0
H-bond cutoff:     3.0 A
H-bond angle:      30 deg.
TS jump cutoff:    6.0 A
Protein pruning:   7.0 A
Epsilon_C:         0.6
Epsilon_H:         0.36
Exp. H-bond:       1.7
Offset H-bond:     2.8 A
Epsilon_TS:        0.6
Exp. TS jump:      1.7
Offset TS jump:    1.4 A
```

```
OUTPUT:
Output prefix:     testrun
Debugging mode on: 1
```

```
GRAPHICS:
Max. pathway radius: 0.5
Pathway resolution: 30
Pathway material:   Opaque
Renderer:           TachyonInternal
```

In the graphics mode, calculated pathways are displayed as shown in Figure 2. Each pathway is represented by a sequence of solid cylinders (covalent bond-mediated steps) and dashed cylinders (hydrogen bonds and through-space steps). The cylinder diameters indicate the pathway's strength. In addition, pathways



**Figure 2.** The 10 strongest pathways calculated for ruthenated wild-type azurin (PDB code 1BEX)<sup>[10]</sup> using standard *Pathways* parameters ( $\epsilon_c = 0.6$ ,  $\beta^S = 1.7 \text{ \AA}^{-1}$ ). Blue and magenta spheres indicate copper and ruthenium ions, respectively. Solid cylinders indicate pathway steps mediated by covalent bonds, and dashed cylinders indicate pathway steps mediated by hydrogen bonds or through-space jumps. Cylinder diameters indicate the pathway strengths.

are color-coded by strength using the repeating sequence of red, orange, yellow, green, cyan, blue, and magenta. For clarity, visual representations of all loaded molecules are preserved, and the calculated pathways are shown as a new loaded molecule. In the graphics mode as well as in the text mode, pathways, ordered by strength, are printed in the following format (output truncated):

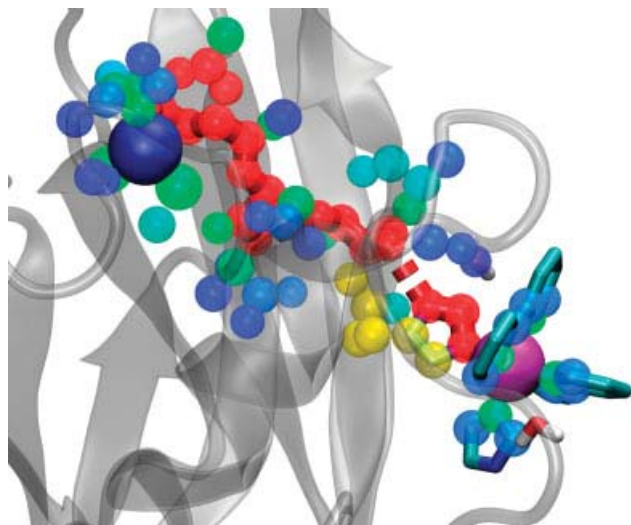
```
PATHWAYS) Found 100 pathway(s)
PATHWAYS) Sorted 100 pathway(s): max coupling 2.43041e-05
PATHWAYS) ELECTRON TRANSFER PATHWAYS:

PATH 1: net decay 2.43041e-05
1005 CU CU 129 CU --> 339 A GLY 45 O : CB 3.00 3.00
339 A GLY 45 O --> 338 A GLY 45 C : CB 1.22 4.22
338 A GLY 45 C --> 347 A HSP 35 N : CB 1.32 5.54
346 A HSP 46 N --> 658 A ILE 87 O : HB 2.87 8.41
658 A ILE 87 O --> 657 A ILE 87 C : CB 1.23 9.65
657 A ILE 87 C --> 652 A ILE 87 CA : CB 1.54 11.19
652 A ILE 87 CA --> 651 A ILE 87 N : CB 1.46 12.64
651 A ILE 87 N --> 649 A LEU 86 C : CB 1.32 13.97
649 A LEU 86 C --> 644 A LEU 86 CA : CB 1.46 15.48
644 A LEU 86 CA --> 643 A LEU 86 N : CB 1.47 16.95
643 A LEU 86 N --> 641 A LYS 85 C : CB 1.33 18.28
641 A LYS 85 C --> 635 A LYS 85 CA : CB 1.53 19.81
635 A LYS 85 CA --> 634 A LYS 85 N : CB 1.46 21.26
634 A LYS 85 N --> 632 A THR 84 C : CB 1.33 22.59
632 A THR 84 C --> 633 A THR 84 O : CB 1.23 23.82
633 A THR 84 O --> 623 A HSP 83 ND1 : HB 2.67 26.48
623 A HSP 83 ND1 --> 624 A HSP 83 CE1 : CB 1.32 27.81
624 A HSP 83 CE1 --> 617 A HSP 83 NE2 : CB 1.32 29.13
617 A HSP 83 NE2 --> 975 RBU RBU 130 RU : CB 2.11 31.24

PATH 2: net decay 1.45824e-05
1005 CU CU 129 CU --> 339 A GLY 45 O : CB 3.00 3.00
```

The columns on either side of the arrows show amino acid number, chain, amino acid name, atom number, and atom name for the initial and the final atom in each step along the pathway. The last three columns show the step type (CB for covalent bonds, HB for hydrogen bonds, and TS for through-space jumps), step distance, and the running distance along the pathway from the donor to the end of the current step, respectively.

Some single-point mutations may dramatically alter ET reactions in redox proteins,<sup>[9,10]</sup> and *Pathways* model-based calculations have successfully captured the influence of individual protein atoms on the electronic coupling in many redox proteins.<sup>[3,9–11]</sup> As such, the *Pathways* plugin includes capabilities for estimating the importance of individual atoms for mediating the electronic coupling. The importance of the *i*th atom is defined as  $I_i = T_{Di}T_{iA}/T_{DA}$ , where  $T_{Di}$  or  $T_{iA}$  are pathway couplings between the *i*th atom and the donor or the acceptor, respectively, and  $T_{DA}$  is the strongest donor-to-acceptor pathway coupling. Atom importance *I* can vary from 0 (for atoms not involved in mediating the electronic coupling) to 1 (for atoms on the strongest path). Figure 3 shows atom importance values near the strongest pathway in ruthenium-modified wild-type azurin.<sup>[9,10]</sup> Atom importance values can be indicated by radius (default) or color-coded from red to blue, as shown in Figure 3. Atoms with importance values less than 1 are those that mediate secondary pathways (Figure 2). As such, atom importance values provide a method of describing ET mediating elements in proteins, and these values may be viewed along molecular dynamics trajectories.



**Figure 3.** The importance  $I_i$  of atoms surrounding (within 4 Å of) the strongest path for mediating the electronic coupling in ruthenated wild-type azurin (PDB code 1BEX).<sup>[10]</sup> Atoms are shown as spheres, and their importance is color-coded from red (highest importance,  $I_i \sim 1$ ) to blue (lowest importance,  $I_i \ll 1$ ).

To perform pathway calculations for multiple frames of a molecular dynamics trajectory, Pathways uses a built-in wrapper that accepts all the above arguments as well as additional optional arguments for the first frame, the last frame, and the increment between frames. An example 300-ps animation of the 10 strongest pathways in ruthenated wild-type azurin (PDB code 1BEX; Ref. [28]) is available on the web (<http://www.chem.duke.edu/~ilya/Media/pathways.mpg>). In addition to calculating pathways and estimating relative electronic couplings for the specified frames, Pathways also calculates the statistical characteristics of coupling fluctuations: the average pathway coupling, the standard deviation of the pathway coupling, the mean squared pathway coupling, and a pathway coherence-like parameter  $C_P$  that quantifies the extent of the electronic coupling fluctuations

$$C_P = \langle |T_{DA}| \rangle^2 / \langle T_{DA}^2 \rangle. \quad (3)$$

The parameter  $C_P$  is related to the coherence parameter  $C$  introduced earlier.<sup>[15]</sup>  $C_P$  describes the sensitivity of the pathway coupling to thermal atomic motion. We refer to it as  $C_P$  because the Pathways model does not explicitly include effects of multiple coupling pathway interference. As in Ref. [15], values of  $C_P$  vary between 0 and 1. Values close to 1 indicate that the strongest pathway is weakly sensitive to thermal motion, while values close to 0 indicate strong coupling dependence on conformation. Statistical characteristics calculated for the above trajectory are:

```

PATHTRAJ) Created the movie
PATHTRAJ) PATHWAY COUPLING STATISTICS:
PATHTRAJ) <T_DA>: 1.5952911063122922e-5
PATHTRAJ) sigma(T_DA): 1.1210076963218966e-5
PATHTRAJ) <T_DA^2> 3.8016119690920234e-10
PATHTRAJ) "Coherence" 0.6694406832075852
    
```

Recent examples of the key role that pathway dynamics can play include water-mediated ET studies.<sup>[3, 20, 21, 23, 24]</sup>

## Acknowledgments

The Pathways plugin and its User's Guide are freely available from the Pathways web page at Duke University (<http://www.chem.duke.edu/~ilya/Software/Pathways/docs/pathways.html>). They are distributed under the terms of the GNU General Public License version 3 or later (<http://www.gnu.org/copyleft/gpl.html>). The User's Guide includes a brief introduction of the Pathways model, installation instructions, usage tips, and a description of Pathways parameters. In addition, a complete parameter reference is built into in the plugin itself.

The Pathways plugin was created by Ilya Balabin (project design and management, integration with VMD, user interfaces, graphics, animation), Xiangqian Hu (graph search wrapper), and David Beratan (overall guidance and development of the Pathways theory). We thank J. N. Betts, J. J. Regan, and J. N. Onuchic for contributions to the Pathways model and earlier code. I.A.B. also thanks John Stone and Klaus Schulten for valuable comments and suggestions.

**Keywords:** bioenergetics • electron transfer • redox proteins • tunneling pathways • VMD

How to cite this article: I. A. Balabin, X. Hu, D. N. Beratan, *J. Comput. Chem.* **2012**, *33*, 906–910. DOI: 10.1002/jcc.22927

- [1] J. Berg, J. Tymoczko, L. Stryer, *Biochemistry*, 6th ed.; W. H. Freeman and Company: New York, NY, **2008**.
- [2] I. A. Balabin, S. S. Skourtis, D. N. Beratan, *Wiley Encyclopedia of Chemical Biology*; Wiley-VCH Verlag GmbH, Hoboken, NJ, **2008**.
- [3] D. N. Beratan, S. S. Skourtis, I. A. Balabin, A. Balaeff, S. Keinan, R. Venkatramani, D. Xiao, *Acc. Chem. Res.* **2009**, *42*, 1669.
- [4] S. S. Skourtis, D. H. Waldeck, D. N. Beratan, *Ann. Rev. Phys. Chem.* **2010**, *61*, 461.
- [5] S. Hammes-Schiffer, *Acc. Chem. Res.* **2001**, *34*, 273.
- [6] J. A. Weil, J. R. Bolton, J. E. Wertz, *Electron Paramagnetic Resonance: Elementary Theory and Practical Applications*; Wiley, Hoboken, NJ, **1994**.
- [7] V. I. Davidson, *Acc. Chem. Res.* **2008**, *41*, 730.
- [8] R. A. Marcus, N. Sutin, *Biochim. Biophys. Acta* **1985**, *811*, 265.
- [9] H. B. Gray, J. R. Winkler, *Biochim. Biophys. Acta Biomembr.* **2010**, *1797*, 1563.
- [10] H. B. Gray, J. R. Winkler, *Q. Rev. Biophys.* **2003**, *36*, 341.
- [11] D. N. Beratan, J. N. Betts, J. N. Onuchic, *Science* **1991**, *252*, 1285.
- [12] J. N. Betts, D. N. Beratan, J. N. Onuchic, *J. Am. Chem. Soc.* **1992**, *114*, 4043.
- [13] G. L. Closs, J. R. Miller, *Science* **1988**, *240*, 440.
- [14] M. L. Jones, I. V. Kurnikov, D. N. Beratan, *J. Phys. Chem. A* **2002**, *106*, 2002.
- [15] I. A. Balabin, J. N. Onuchic, *Science* **2000**, *290*, 114.
- [16] S. S. Skourtis, I. A. Balabin, T. Kawatsu, D. N. Beratan, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3552.
- [17] T. R. Prytkova, I. V. Kurnikov, D. N. Beratan, *Science* **2007**, *315*, 622.
- [18] J. N. Gehlen, I. Daizadeh, A. A. Stuchebrukhov, R. A. Marcus, *Inorg. Chim. Acta* **1996**, *243*, 271.
- [19] J. J. Regan, S. M. Risser, D. N. Beratan, J. N. Onuchic, *J. Phys. Chem.* **1993**, *97*, 13083.
- [20] A. de la Lande, S. Mart, O. Parisel, V. Moliner, *J. Am. Chem. Soc.* **2007**, *129*, 11700.

- [21] A. de la Lande, N. S. Babcock, J. Rezac, B. C. Sanders, D. R. Salahub, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11799.
- [22] J. Wolfgang, S. M. Risser, S. Priyadarshy, D. N. Beratan, *J. Phys. Chem. B* **1997**, *101*, 2987.
- [23] J. Lin, I. A. Balabin, D. N. Beratan, *Science* **2005**, *310*, 1311.
- [24] I. A. Balabin, S. S. Skourtis, D. N. Beratan, *Phys. Rev. Lett.* **2008**, *101*, 158102.
- [25] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
- [26] E. Dijkstra, *Numer. Math.* **1959**, *1*, 269.
- [27] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acid Res.* **2000**, *28*, 235.
- [28] S. Faham, M. W. Day, W. B. Connick, B. R. Crane, A. J. Di Bilio, W. P. Schaefer, D. C. Rees, H. B. Gray, *Acta Crystallogr. Sect. D* **1999**, *55*, 379.

---

Received: 24 February 2011

Revised: 12 October 2011

Accepted: 11 December 2011

Published online on 1 February 2012