

# Translocation of knotted proteins through a pore

P. Szymczak<sup>a</sup>

Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Hoza 69, 00-681 Warsaw, Poland

Received 30 June 2014 / Received in final form 21 July 2014

Published online 22 September 2014

**Abstract.** We report the results of molecular dynamics simulations of translocation of knotted proteins through pores. The protein is pulled into the pore with a constant force, which in many cases leads to the tightening of the knot. Since the radius of tightened knot is larger than that of the pore opening, the tight knot can block the pore thus preventing further translocation of the chain. Analyzing six different proteins, we show that the stuck probability increases with the applied force and that final positions of the tightened knot along the protein backbone are not random but are usually associated with sharp turns in the polypeptide chain. The combined effect of the confining geometry of the pore and the inhomogeneous character of the protein chain leads thus to the appearance of topological traps, which can immobilize the knot and lead to the jamming of the pore.

## 1 Introduction

In less than 1% of the proteins the polypeptide chain adopts a knotted configuration [1–3]. Compared with ordinary polymers of comparable length, compactness, and flexibility proteins have fewer knots than would be expected for a random distribution of conformations [4] suggesting that nature finds it expedient on the whole to eliminate knots. What is it then about the knotted proteins that makes them so rare in living matter? One possibility, proposed in [2, 5, 6] is that the presence of a knot may affect the ability of proteins to be degraded in proteasome or translocated through the intercellular membranes, e.g. during import into mitochondria. The smallest constrictions in the mitochondrial pores or proteasome openings are 12–14 Å in diameter [7, 8], too narrow to accommodate folded structures, thus translocation must be coupled to protein unfolding. Unfolding and import of proteins into mitochondria or proteasome are facilitated by molecular motors that act with the forces of the order of 30 pN [9]. However, as it was shown in a number of studies, both experimental and numerical [5, 10, 11], the protein knots tend to tighten under the action of the force. The radius of gyration of the tight knot has been estimated to be around 7–8 Å for the simplest knot (a trefoil) and correspondingly larger for more complicated knots. This means that the knot seems to be a shade too large to squeeze through the pore

<sup>a</sup> e-mail: Piotr.Szymczak@fuw.edu.pl



**Fig. 1.** Pulling the rope with a trefoil knot into the model pore (the opening of a rope float). The initial configuration (left) and the tightened configuration after a sharp tug at the rope (right).

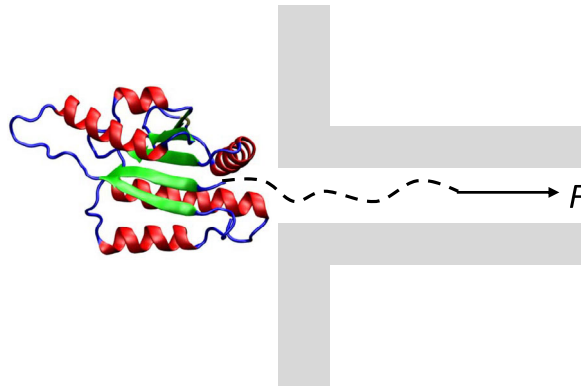


**Fig. 2.** Same as in Fig. 1 but for a figure-of-eight knot.

openings. This leaves us with two possibilities: either the knot diffuses towards the end of the chain and simply slides away, or it gets tightened and jams the opening, as suggested by the numerical studies reported previously in [6]. It is easy to convince oneself that it might indeed be so by conducting a simple macro-scale experiment: tying a knot on a piece of rope and then pulling it through a cylindrical hole (e.g. the one in the rope float, as shown in Fig. 1). If pulled sharply, the knot invariably tightens; however if tugged slowly – it might be able to squeeze its way through the hole without tightening. Below, we analyze the jamming process in more detail, performing the numerical simulations of the translocation process for six different knotted proteins.

## 2 Numerical model

The numerical model used here combines a coarse-grained protein model with a minimalist model of a repelling pore. For protein, we adopt a  $G\bar{o}$ -type model, in which individual amino acids are replaced by beads of uniform size whose positions correspond to the locations of the  $C_\alpha$  atoms. The effective potential of the interaction between these beads is then introduced, tailored to give lowest energy to the native state of a protein. A particular implementation of the  $G\bar{o}$ -type model followed here is by Cieplak and co-workers [10,12]. In short, the protein structure is represented by a chain of  $C_\alpha$  atoms tethered along the backbone by harmonic potentials with minima at  $l_p = 3.8 \text{ \AA}$ . Effective interactions between residues are split into native and nonnative interactions by checking for overlaps between the enlarged van der Waals surfaces of the residues [13]. Amino acids ( $i$  and  $j$ ) that overlap are endowed with the effective Lennard-Jones potential  $V_{ij}(r) = 4\epsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$  with energy scale  $\epsilon$  and pair-by-pair distances  $r_{ij}$ . The length parameters,  $\sigma_{ij}$ , are chosen such that the potential minima correspond pair by pair to the native state distance between the residues.



**Fig. 3.** The simulation setup: the protein (here Ins5), presequence (dashed) and the pore (grey).

Nonnative contacts are represented by hardcore repulsion to prevent entanglements. Correct chirality is imposed by the angle-dependent term in the Hamiltonian.

In the cell, the transport of proteins into mitochondria is usually mediated by a loosely folded presequence, which is modeled here as a loose piece of a peptide chain (10 amino acid long). One end of the presequence is attached to the terminus of the protein, while the other end is pulled with a constant force  $F$ . The overdamped motion of amino acids in solvent is mimicked using a standard Brownian dynamics algorithm at the temperature corresponding to  $kT = 0.3\epsilon$ . The characteristic timescale  $\tau = \sigma^2/6D_0$  is set by the time it takes for the amino acid to diffuse the typical contact distance  $\sigma \approx 5 \text{ \AA}$ . Here,  $D_0$  is a single particle diffusion coefficient. Experimentally, the time scale  $\tau$  is of the order of a nanosecond.

Next, the pore is modeled as a cylindrical structure interacting with the aminoacids by the potential

$$V_{pore}(\mathbf{r}_i) = V_0 \frac{1}{1 + \exp[1 - \rho_i^2/\rho_0^2]} \quad z_i > 0 \quad (1)$$

as proposed by J.M. Deutsch [14]. Here  $\mathbf{r}_i = \{x_i, y_i, z_i\}$  is the position of  $i$ th aminoacid,  $z$  is oriented along the pore axis and  $\rho_i = \sqrt{x_i^2 + y_i^2}$  is the distance from the axis. The potential is small within the radius  $\rho_0$  from the axis of the pore and then rises sharply. Additionally, to prevent the protein from entering the membrane except through the pore, a short range, repulsive membrane potential is introduced at its trans side ( $z < 0$ )

$$V_{mem}(\mathbf{r}_i) = V_0 \left(\frac{z_0}{z}\right)^9, \quad z_i < 0, \quad \rho_i > \rho_0. \quad (2)$$

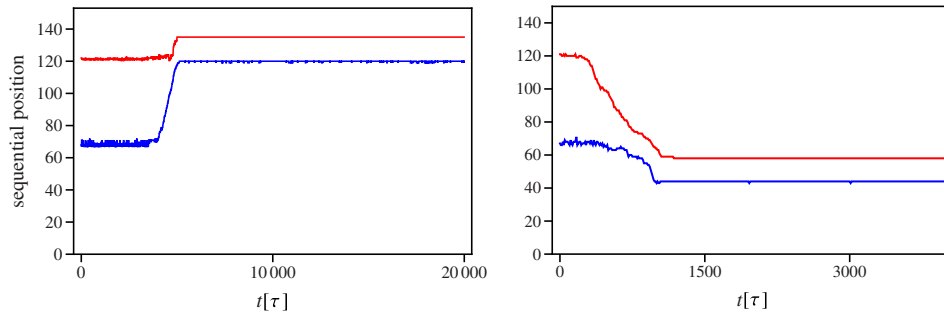
In the simulations reported here,  $V_0 = 10\epsilon$ ,  $z_0 = 0.5 \text{ \AA}$  and  $\rho_0 = 3 \text{ \AA}$ . Note that the pore potential acts on the centers of the particles. Since the van der Waals radii of amino acids are in the range of  $\sim 3.5 - 4.5 \text{ \AA}$ , the above value of  $\rho_0$  corresponds to the effective pore radius of about  $6.5 - 7.5 \text{ \AA}$ , which is consistent with the values reported for the narrowest constriction in the mitochondrial pores [8].

### 3 Results

At the beginning of the simulations the protein in its native conformation is positioned at the outside side of the membrane near the pore entrance, with the end of

**Table 1.** Proteins considered and the characteristics of their knots.

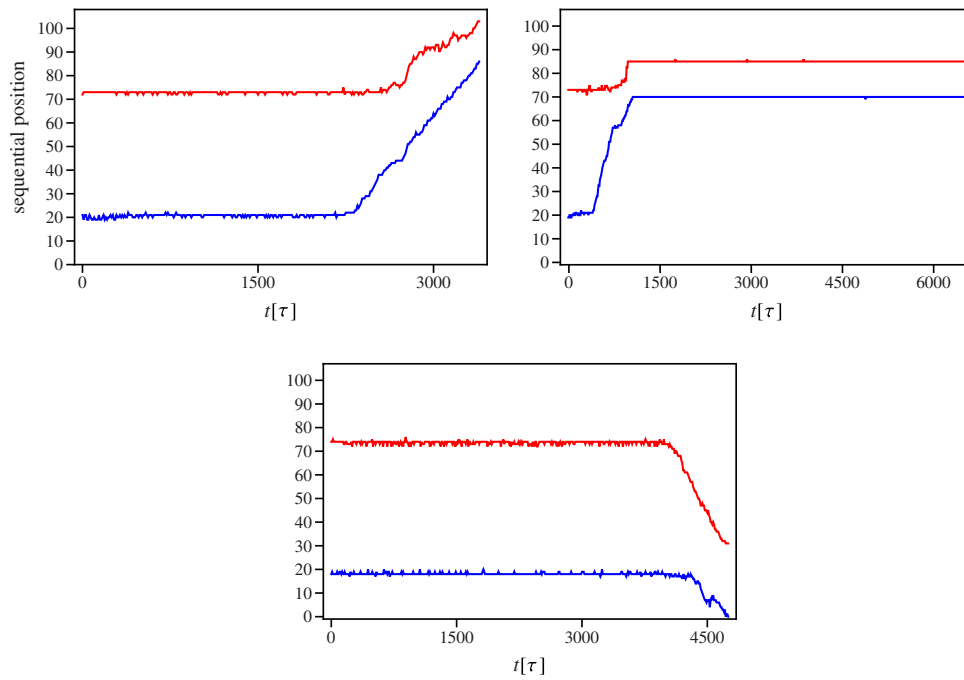
protein	pdb	length	knotted core	knot type
YbeA from <i>E. coli</i>	1ns5	1–153	67–121	$3_1$
zinc-finger motif	2k0a	–1–107	21–73	$3_1$
YibK methyltransferase	1j85	1–156	75–120	$3_1$
YbeA-like from <i>T.maritima</i>	1o6d	1–147	65–118	$3_1$
Ribbon-helix-helix protein (from <i>M.jannaschii</i> )	2efv	6–87	13–80	$3_1$
transcarbamylase from <i>X.campestris</i>	1yh1	3–336	172–254	$3_1$
FLIN2 chimaeric protein	1j2o(14)	1–114	42–95	$4_1$



**Fig. 4.** The movement of knotted core during the translocation process of the protein 1ns5 pulled by the N (left) terminus with the force  $F = 2.3\epsilon/\text{\AA}$  or C terminus (right) with  $F = 3\epsilon/\text{\AA}$ . The colors mark the two ends of the knotted core as they move along the sequence.

the presequence placed at the pore axis. The force is then switched on and the presequence is pulled into the pore. Several knotted proteins were studied, as summarized in Table 1. During the simulation we not only record the conformation of the protein but also track the position of the knotted core, i.e. the smallest region that will remain knotted when the residues are successively deleted from both ends [15]. Thereby we obtain the trajectories of knot's ends in the sequential space, such as those shown in Figs. 4 and 5. The former shows the movement of the knot for the YbeA protein from *E. coli* (pdb code: 1ns5). In this case, irrespectively which terminus is pulled into the pore, the knot always gets stuck in a tightened conformation, although the final positions of the knot differ. As listed in Table 2 for N-pulling it lands either between aminoacids 119 and 134 or between 136 and 151. On the other hand, for C-pulling the knot gets stuck in between amino acid 44 and 58. Invariably, when the knot is tightened, the knotted core reduces to 12–15 amino acids (for a trefoil knot) and the radius of gyration is of the order of 7–8 Å.

It is interesting to analyze which amino acids act as pinning centers for the knots in different proteins. Similarly to what was reported in knot tightening simulations without the pore [10], the pinning centers are mostly associated with very tight turns in the protein backbone, where the polypeptide chain changes its direction. There are several aminoacids usually residing in such turns [16]: glycine, proline, aspartic acid, glutamic acid, and serine. Proline forces a sharp turn, since its side chain is connected to the protein backbone twice: to the backbone nitrogen as well as to carbon. Glycine is unique as it contains a single hydrogen as its side chain, which allows for a considerable conformational flexibility. Similarly, serine is often found at the tight turns due to its small size. Finally, aspartic acid and glutamic acid prefer to



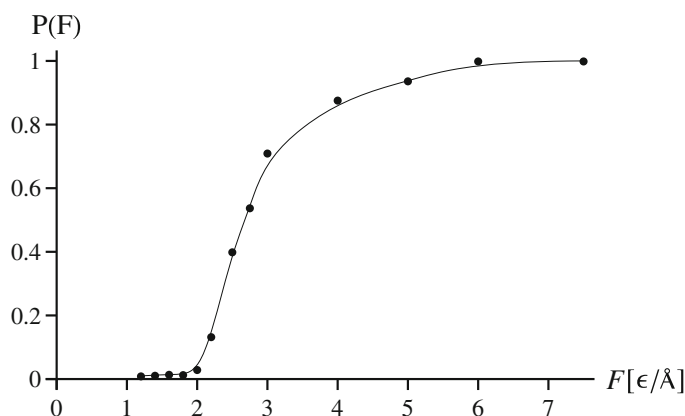
**Fig. 5.** The movement of knotted core during the translocation process of the protein 2k0a pulled by the N (the upper left and upper right panel) or C terminus (lower). The colors mark the two ends of the knotted core as they move along the sequence. The panels correspond to the forces  $F = 2\epsilon/\text{\AA}$  (upper left),  $F = 2.25\epsilon/\text{\AA}$  (upper right) and  $F = 3\epsilon/\text{\AA}$  (lower). In the upper left and lower panel the knot slides off the chain, in the upper right one it gets pinned.

**Table 2.** The scenarios observed during the translocation of knotted proteins. The knot either freely slides off the chain (“free”) or it gets tightened blocking the pore (“stuck”). The intermediate case in which the pore gets stuck with a certain force-dependent probability,  $P(F)$ , is tagged as “sigmoidal”. Additionally, the positions of the ends of the pore in a stuck conformation are given, marking the cases in which they involve turn-prone residues (Gly,Pro,Ser,Glu,Asp). Lower case “t” marks the positions in the immediate neighborhood of an actual turn in the protein structure (as determined by the visual inspection).

protein	N-puling	C-pulling
1ns5	stuck: 119(Ser)-134 or 136-151(Pro,t)	stuck: 44(Gly,t)-58(Gly)
2k0a	sigmoidal: 70-85(Asp,t)	free
1j85	free	stuck: 68(Glu)-82(Gly,t)
1o6d	sigmoidal: 118(Ser)-135(Glu)	stuck: 62(Pro,t)-47(Glu)
2efv	free	free
1yh1	stuck: 253(Ser)-268(Pro) or 268(Pro)-282	stuck: 159-174
1j2o	free	sigmoidal: 32(Ser)-52(Gly,t)

expose their charged side chains to solvent and thus they tend to reside at sharply turning regions on the surface of the protein. As shown in Table 2 indeed many of the above-mentioned residues are among those which stop the moving knot.

It is, however, not every time that the knot passing a potential pinning center gets stopped. In fact, as summarized in Table 2 for some proteins there is only a finite probability to get stuck. An example of such behaviour for the protein 2k0a is

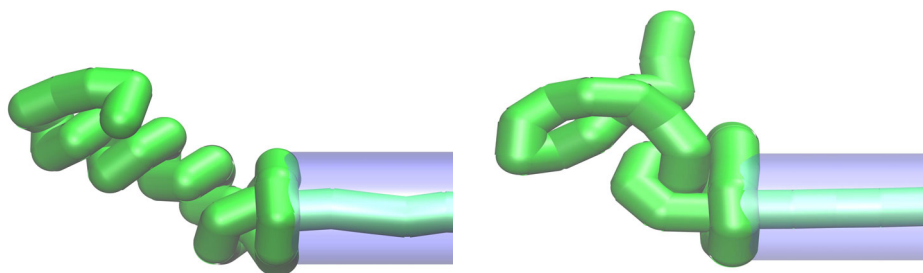


**Fig. 6.** Stuck probability as a function of force for 2k0a protein with a trefoil knot.

presented in Fig. 5: in the upper left panel the knot slides off the chain whereas in the upper right one it gets pinned between aminoacids 71 and 85. In such cases, the stuck probability is invariably force-dependent. Although one might have thought that a sharper pull will make it more probable for a knot to get to the other side, in fact it is just the opposite: large forces tend to tighten the knot and block the pore. In fact, as the experiments with the rope and a float would tell us the successful tactics is different: “Wisely and slowly, they stumble that run fast” as the Shakespearean quote goes. Entering tightened configuration involves crossing an energy barrier [11,17], which is the easier to overcome the higher the force is; hence the dramatic increase of pore blockage probability with the force,  $P(F)$  (*cf.* Fig. 6). Once tightened, the knot is highly unlikely to get loose again, unless the force is relaxed – thus the tightening inevitably results in the blockage of the pore. The data presented in Fig. 6 was obtained for the N-pulling of zinc-finger motif protein 2k0a, but similar sigmoidal shape of  $P(F)$  is also found in the case of N-pulling of 1o6d and C-pulling of 1j2o, as summarized in Table 2. This behavior is similar to that reported by Rosa et al. [18] for force-induced translocation of knotted polyelectrolyte chain, such as ssDNA. On the other hand, no jamming was observed in the passive ejection of DNA out of a spherical cavity [19,20]. Note however that one needs to be careful when drawing a parallel between the behaviour of translocating proteins and DNA, precisely because of the presence of the potential pinning centers in the protein backbone in contrast to (nearly) uniform properties of the DNA chain. For the same reason, the rope and float example is more relevant to DNA than to proteins. In the latter case, instead of a rope, one should take the old garden hose with a lot of kinks, which could potentially pin the translocating knot<sup>1</sup>.

In some cases, however, the knot translocates no matter how sharply we pull it in. Within the set of proteins considered, this takes place for 2efv protein (both for N- and C-terminus pulling) and for C-pulling of 2k0a and N-pulling for both 1j85 and 1j2o. In all of these cases the distance between the free end of the protein and the boundary of the knotted core is relatively short (less than 10 amino acids for both sides of 2efv, about 20 for 2k0a and 1j2o) with an exception of 1j85 where the end of the knot is 36 residues away from the end of the chain. A more detailed analysis of the latter case reveals that the knot gets temporarily pinned between Gly (152) and Ser (136), but the pinning is apparently too weak and eventually the knot gets released and slides off the chain.

<sup>1</sup> The kinked garden hose as a parallel for a translocating protein is due to Gregory Buck.



**Fig. 7.** Example tight-knot configurations blocking the entrance to the pore. Left: trefoil knot in *E. coli* methyltransferase (1ns5), right: figure-of-eight knot in FLIN2 chimaeric protein (1j2o).

Finally, there are cases in which – at least in the force range considered (i.e.  $1 - 8 \epsilon/\text{\AA}$ ) the knot always tightens and gets stuck at the entrance to the pore (*cf.* Table 2). This happens as a rule for deep knots – i.e. in cases where the distance between the end of the knot and the protein terminus is at least 30 residues. Such long stretches of protein backbone contain many potential trapping sites, which effectively stop the knot. It might be, however, that one needs to go to much lower forces to see the sigmoidal behaviour. Unfortunately, smaller force regimes are computationally inaccessible due to the large translocation times involved. Note that the depth of the knot does not need to correlate with the length of the protein. A good example here is carbonic anhydrase, the first knotted protein discovered [21], which is relatively long (260 amino acids), but the knot there is extremely shallow (with one of the ends positioned just two residues away from C terminus).

## 4 Summary

It has been shown that when the knotted protein is pulled with a constant force into the pore there are two scenarios possible: the knot either slides off the chain or it gets tightened and blocks the pore. Note that this is in contrast to the results of Huang and Makarov [22], in which the parameters of the pore allowed the knotted chain to enter inside it. The snapshots of tightened conformations are presented in Fig. 7. Interestingly, they both involve a fastened loop around the entrance of the pore, quite similar to what takes place in the case of the rope (right panels of Figs. 1 and 2).

The positions along the protein sequence at which the knot gets stuck are not random, but are usually associated with sharp turns in the polypeptide backbone. This is similar to what happens during knot tightening in stretched proteins in the absence of the pore [10]. If the knot is relatively shallow then it might happen that there are no potential pinning centers between its end and the end of the chain – in such cases it would slide off. In other cases, the pinning centers trap the passing knot with a certain probability,  $P(F)$ , which increases with the applied force. Finally, if the knot is deep and there are many potential trapping sites, the tightening probability is almost one and the knot invariably blocks the pore, at least in the range of forces considered.

This work was supported by the Polish Ministry of Science and Higher Education under research Grant No. N N202 055440. The author benefited from discussions with Marek Cieplak, Joanna Sułkowska and Piotr Sułkowski. The Brownian dynamics code used in the simulations was based on the protein-modeling package by Marek Cieplak.

## References

1. W.R. Taylor, K. Lin, *Nature* **421**, 25 (2003)
2. P. Virnau, L. Mirny, M. Kardar, *PLoS Comput. Biol.* **2**(9), e122 (2006)
3. D. Bölinger, J.I. Sulkowska, H.P. Hsu, L.A. Mirny, M. Kardar, J.N. Onuchic, P. Virnau, *PLoS Comput. Biol.* **6**(4), e1000731 (2010)
4. R.C. Lua, A.Y. Grosberg, *PLoS Comput. Biol.* **2**(5), e45 (2006)
5. T. Bornschlöggl, D.M. Anstrom, E. Mey, J. Dzubiella, M. Rief, K.T. Forest, *Biophys. J.* **96**, 1508 (2009)
6. P. Szymczak, *Biochem. Soc. Trans.* **41**, 620 (2013)
7. A. Förster, C. Hill, *Trends Cell. Biol.* **13**, 550 (2003)
8. P. Rehling, K. Brandner, N. Pfanner, *Nat. Rev. Mol. Cell Biol.* **5**, 519 (2004)
9. N.N. Alder, S.M. Theg, *Trends Biochem. Sci.* **28**, 442 (2003)
10. J. Sulkowska, P. Sulkowski, P. Szymczak, M. Cieplak, *Phys. Rev. Lett.* **100**, 058106 (2008)
11. J. Dzubiella, *Biophys. J.* **96**, 831 (2009)
12. M. Cieplak, T.X. Hoang, M.O. Robbins, *Proteins: Struct. Funct. Bio.* **56**, 285 (2003)
13. J. Tsai, R. Taylor, C. Chotchia, M. Gerstein, *J. Mol. Biol.* **290**, 253 (1999)
14. J.M. Deutsch, arXiv preprint [[arXiv:1303.0453](https://arxiv.org/abs/1303.0453)] (2013)
15. W.R. Taylor, *Nature* **406**, 916 (2000)
16. M.J. Betts, R.B. Russell, in *Bioinformatics for Geneticists*, edited by M.R. Barnes, I.C. Fray (Wiley Publishers, 2003), p. 289
17. R. Metzler, W. Reisner, R. Riehn, R. Austin, J.O. Tegenfeldt, I.M. Sokolov, *Europhys. Lett.* **76**, 696 (2006)
18. A. Rosa, M. Di Ventra, C. Micheletti, *Phys. Rev. Lett.* **109**, 118301 (2012)
19. R. Matthews, A. Louis, J. Yeomans, *Phys. Rev. Lett.* **102**, 088101 (2009)
20. D. Marenduzzo, E. Orlandini, A. Stasiak, L. Tubiana, C. Micheletti, et al., *Proc. Natl. Acad. Sci. USA* **106**, 22269 (2009)
21. M.L. Mansfield, *Nat. Struct. Mol. Biol.* **1**, 213 (1994)
22. L. Huang, D.E. Makarov, *J. Chem. Phys.* **129**, 121107 (2008)