Enzyme Catalysis

International Edition: DOI: 10.1002/anie.201507772
German Edition: DOI: 10.1002/ange.201507772

Stepwise Versus Concerted Mechanisms in General-Base Catalysis by Serine Proteases

Neta Uritsky, Michael Shokhen, and Amnon Albeck**

Abstract: *General-base catalysis in serine proteases still poses mechanistic challenges despite decades of research. Whether proton transfer from the catalytic Ser to His and nucleophilic attack on the substrate are concerted or stepwise is still under debate, even for the classical Asp-His-Ser catalytic triad. To address these key catalytic steps, the transformation of the Michaelis complex to tetrahedral complex in the covalent inhibition of two prototype serine proteases was studied: chymotrypsin (with the catalytic triad) inhibition by a peptidyl trifluoromethane and GlpG rhomboid (with Ser-His dyad) inhibition by an isocoumarin derivative. The sampled MD trajectories of averaged pK_a values of catalytic residues were QM calculated by the MD-QM/SCRF(VS) method on molecular clusters simulating the active site. Differences between concerted and stepwise mechanisms are controlled by the dynamically changing pK_a values of the catalytic residues as a function of their progressively reduced water exposure, caused by the incoming ligand.*

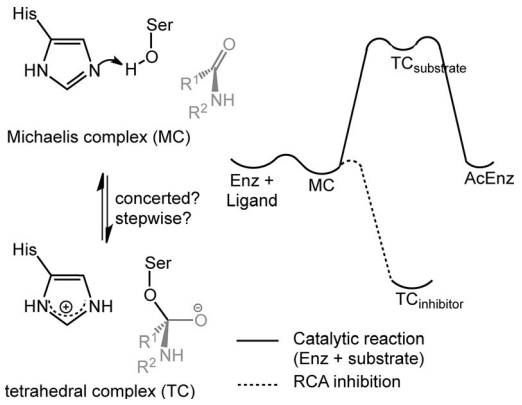

Figure 1. A) The first chemical steps in amide-bond hydrolysis catalyzed by serine proteases, from the Michaelis complex (MC) to the tetrahedral complex intermediate (TC). B) Comparative energy diagram of the transformation of the MC into the TC with a substrate and a RCA inhibitor.

The aim of this study is to shed light on the challenging problem of general-base catalysis in serine proteases. The "classical" Asp-His-Ser catalytic triad of chymotrypsin was the first protease catalytic machinery to be characterized almost 50 years ago.[1] By now, several new families of serine proteases have been discovered with catalytic residues other than the canonical triad (for example, Ser-His-Glu, Ser-His-His, Ser-Glu-Asp, Ser-Ser-Lys, and dyads like Ser-Lys and Ser-His).[2] The first chemical transformation in the multistep peptide hydrolysis by all serine proteases is the general-base catalyzed nucleophilic attack of serine on the substrate's scissile amide bond in the noncovalent Michaelis complex (MC). It results in an anionic covalent tetrahedral complex (TC; Figure 1 A). The catalytic His residue in most serine protease subfamilies plays a central role as a general-base/general-acid catalyst. In particular, this role is true for the chymotrypsin family (S1), having a His-Asp-Ser catalytic triad, and for the rhomboid protease family (S54), having a His-Ser catalytic dyad.[2c,3]

The recently discovered family of membrane-embedded rhomboid serine proteases,[4] which hydrolyze membrane-bound proteins within the membrane environment, is ubiquitous in all forms of life.[5] Rhomboids can be inactivated by irreversible covalent inhibitors, but are insensitive to common serine protease peptidic inhibitors.[6,7]

In this work, we present a computational study of the general-base-catalyzed transformation of the MC into the TC for the acylation step in water-solvated chymotrypsin which features the Ser-His-Asp catalytic triad, and in *E. coli* GlpG membrane-embedded rhomboid protease which features a Ser-His dyad.

Modern QM/MM techniques,[8] pioneered by Warshel and Levitt,[9] have become the main computational tool for simulation, analysis, and interpretation of enzyme mechanisms. Despite such extensive studies, the mechanism of the transformation of the MC into the TC, even with the classical Asp-His-Ser catalytic triad, is still under debate. Warshel and co-workers suggested a stepwise reaction with prior proton transfer in the MC.[10] Others support a concerted proton transfer from Ser to His and nucleophilic attack of Ser on the substrate.[11] In fact, Warshel's old question of "How do serine proteases really work?",[12] is still open.

Proton transfer, and therefore pK_a values of catalytic residues, play a central role in this transformation. We previously applied our QM/SCRF(VS) method to calculations on pK_a values, and it treats a biomolecule and its environment as a two-layer system (see Figure S1 in the Supporting Information (SI)).[13,14] The inner layer is the enzyme active site, simulated by a quantum mechanically calculated QM molecular cluster. The outer layer includes the rest of the protein and its environment, including lipid molecules, bulk solvent water molecules, and counter ions.

[*] N. Uritsky, Prof. M. Shokhen, Prof. A. Albeck
The Julius Spokojny Bioorganic Chemistry Laboratory
Department of Chemistry, Bar Ilan University
Ramat Gan 5290002 (Israel)
E-mail: michael.shokhen@biu.ac.il
amnon.albeck@biu.ac.il

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201507772.

The outer layer is accounted for in QM self-consistent reaction field (SCRF) calculations as a uniform dielectric continuum, "virtual solvent" (VS), which is characterized by a dielectric constant ε. We previously derived an analytical relationship implying that this ε value is a measure of the degree of water exposure of the catalytic residues within the active site.[13] Thus, the gradual reduction of ε values in QM/SCRF(VS) modeling simulates decreasing the water exposure of the enzyme active site, a decrease which is caused by ligand binding to the free enzyme in the process of the MC and TC formation. The reduced water solvation of the catalytic His is the dominant factor which elevates its pK_a value and therefore triggers the general-base catalysis in serine proteases.[13] The Ser to His proton transfer (general-base catalysis) can be realized when

$$\text{His } pK_a(\varepsilon) \geq \text{Ser } pK_a(\varepsilon) \quad (1)$$

Therefore knowledge of the p$K_a(\varepsilon)$ values of catalytic residues is essential for interpretation of the details of the enzymatic transformation of the MC into the TC. The QM/SCRF(VS) method provided excellent results in quantitative p$K_a(\varepsilon)$ calculations of enzyme catalytic residues.[13,14] In our analysis we accepted the physiological pH 7.4. In combination with calculated pK_a values as a function of ε, the pH value serves to selection of the reliable protonation states of catalytic residues in both the MC and TC states (see Tables S4 and S7–S10 in the SI).

TCs are short-lived intermediates in serine protease catalyzed peptide hydrolysis, and their three-dimensional structures cannot be identified by current experimental methods. In contrast, a thermodynamically stable TC is formed between serine proteases with the Asp-His-Ser triad and reaction coordinate analogue (RCA) inhibitors (Figure 1 B).[2c,3,15] The crystal structure of the TC formed by chymotrypsin with a peptidyl trifluoromethane (TFK; 1gg6.pdb)[16] is used in this work as a starting point for the MD simulated MC and TC. The superposition of the MD simulated TC with the crystal structure is presented in Figure S2 (see the SI), thus demonstrating a good geometrical match in both the backbone and the catalytic residues. The averaged interatomic distances of active-site residues (see Figure S7 in the SI) in the ground states of the TC and MC are presented in Tables S1, S2, and S3 (see the SI).

The TC formed between a rhomboid protease and an isocoumarin inhibitor is also a short-lived intermediate. We used the 2xow.pdb crystal structure[17] to reconstruct the MC and TC of *E. coli* GlpG rhomboid with 7-amino-4-chloro-3-methoxyisocoumarin. The coincidence of the MC and TC superimposed structures with each other and with the alkylated acyl enzyme crystal structure are presented in Figure S8 (see the SI). The averaged interatomic distances of key interactions between active-site residues (see Figure S13 in the SI) in the ground states of the TC and MC are presented in Tables S5 and S6 (see the SI). Most of the in vitro kinetics studies and crystal structure determinations of rhomboid inhibitor complexes were carried out in either detergent- or lipid-solubilized (bicelle) enzymes. Phosphatidylethanolamine (POPE) constitutes about 70% of the *E. coli* membrane,[18] therefore our molecular dynamics simulations of the MC and TC of GlpG were conducted on the enzyme surrounded by POPE lipid annulus and solubilized in water (see Figure S14 in the SI). It should be noted that the catalytic mechanism of the enzyme acylation step by the isocoumarin inhibitor is analogous to that of the proteolytic hydrolysis of esters (see Figure S15 in the SI).

We applied the MD-QM/SCRF(VS) method, thus combining molecular dynamics with quantum mechanical calculations.[13b] The QM molecular cluster includes the catalytic residues and all residues interacting with them through hydrogen bonds and π stacking, and active-site residues identified as conserved residues in the sequence analysis. Water molecules forming hydrogen bonds with catalytic residues were also included. We generated QM molecular clusters simulating the MC and TC with zero, one, and two abstractable protons in all possible combinations of the protonation states at the catalytic residues (Figures 2 and 3). MD-QM/SCRF(VS) uses a sampling algorithm for the calculations of averaged heats of formations $\langle H_f(\varepsilon)\rangle$ and $\langle pK_a(\varepsilon)\rangle$ at different ε values on the QM clusters (see Method in the SI) for each protonation state in the MC and TC of

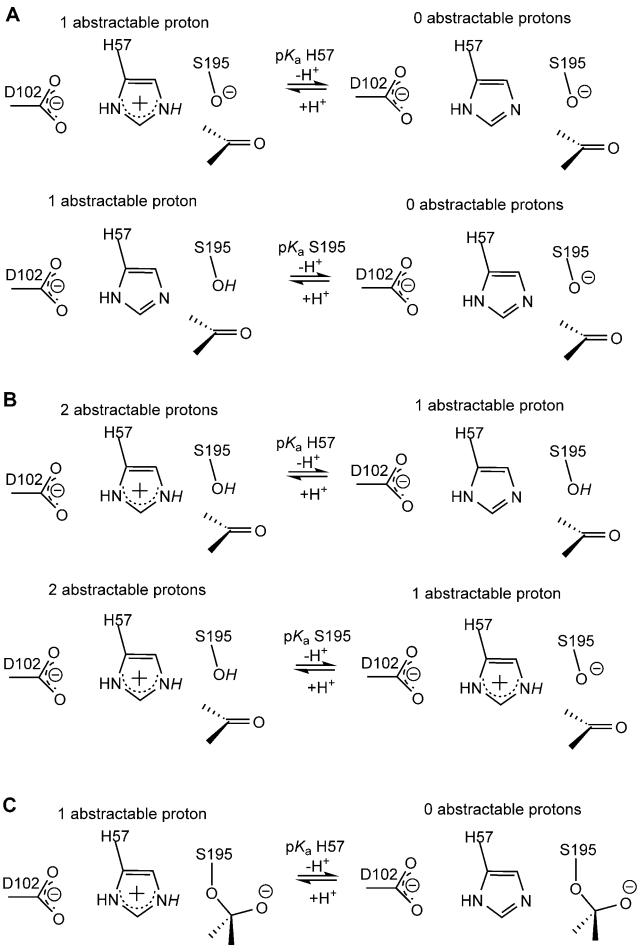

Figure 2. Proton abstraction schemes for $\langle pK_a(\varepsilon)\rangle$ calculations on the catalytic residues of chymotrypsin in the MC with one abstractable proton (A), and two abstractable protons (B), and in the TC with one abstractable proton. The abstractable protons are italicized.

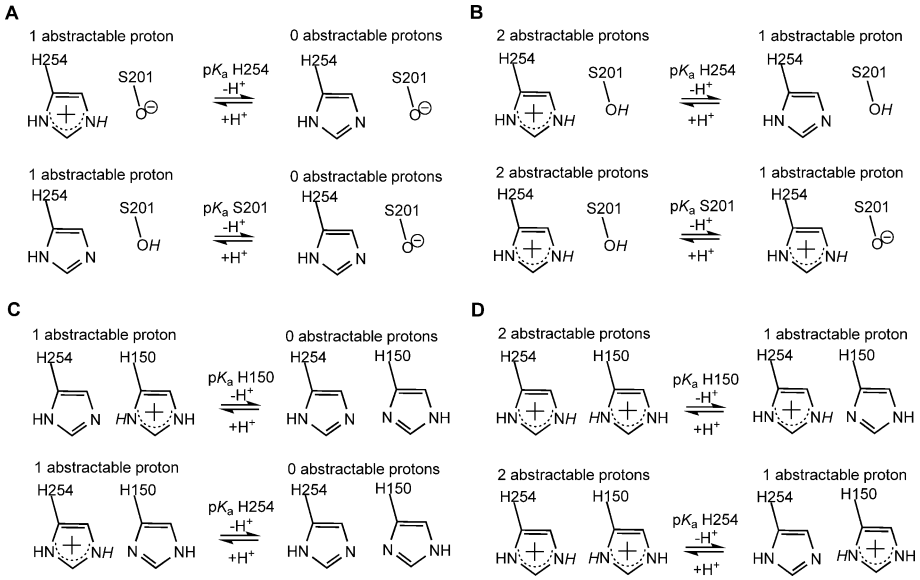

Figure 3. Proton abstraction schemes for $\langle pK_a(\varepsilon)\rangle$ calculations of catalytic residues in the active site of GlpG rhomboid for states with one (A) and two (B) abstractable protons (italicized) in the MC and with one (C) and two (D) abstractable protons (italicized) in the TC.

chymotrypsin (Figure 2; see Table S4 in the SI) and the rhomboid protease (Figure 3; see Tables S7–S10 in the SI). The QM clusters were collected from the end time interval of productive MD trajectories, thus satisfying the convergence criteria by total energy and RMSD(Cα) (see Protocol in the SI).

Experimental pK_a values for the TC formed by chymotrypsin with Ac-Leu-Phe-CF_3 and Ac-Phe-CF_3 at 5 °C are 12.0 and 10.8 (10.8 and 9.7 recalculated at 37 °C), respectively.[19] They were reproduced by computed $\langle pK_a(H57,\varepsilon,TC)\rangle$ at $\varepsilon \approx 6$ (see Table S4 in the SI). MD-QM/SCRF(VS) calculations reveal the exponential increase of $\langle pK_a(H57,\varepsilon,MC)\rangle$, $\langle pK_a(S195,\varepsilon,MC)\rangle$, and $\langle pK_a(H57,\varepsilon,TC)\rangle$ with decreasing ε values (Figure 4 and Table S4). The data show that the condition for general-base catalysis [$pK_a(\varepsilon)$ His $\geq pK_a(\varepsilon)$ Ser] is already satisfied in the MC at much higher values ($\varepsilon \leq 9$) than those in TC ($\varepsilon \leq 2$), according to the position of the intersection points of the curves in Figure 4, curves calculated with one abstractable proton (see Table S4 in the SI). The rate of proton transfer between catalytic residues is about five orders of magnitude faster than the diffusion-driven alignment of a ligand in an enzyme active site.[20] Together with the calculated His $pK_a(\varepsilon)$ value in the MC and TC as a function of water exposure, we are led to suggest that proton transfer from His to Ser in chymotrypsin could occur as a discrete first reaction step in the noncovalent enzyme–ligand MC, and the interatomic distance between the ligand electrophilic center and Ser is still too long for successful nucleophilic attack.

The acylation mechanism of an isocoumarin inhibitor is analogous to that of ester substrate hydrolysis by serine proteases (see Figure S15 in the SI). In the ground state of the MC, the only relevant protonation state of the S201/H254 pair at pH 7.4 is when both residues are neutral, with one abstractable proton located at S201 (see Table S8 in the SI). The Nε atom of H254 is within hydrogen-bond distance (2.81 Å) from Oγ S201 (see Figure S13 and Table S5 in the SI) and in an optimal position for general-base catalysis. Nevertheless, the proton transfer from S201 to H254 is energetically forbidden in the MC since the $\langle pK_a(\varepsilon)\rangle$ of the H254 Nε is considerably lower than the $\langle pK_a(\varepsilon)\rangle$ of S201 Oγ in the whole ε interval (see Table S9 in the SI and Figures 3 A and 5 A). In contrast, the $\langle pK_a(\varepsilon)\rangle$ values of H254 Nε become much higher in the TC than for Oγ of S201 in the MC with decreasing $\varepsilon \leq 10$ (Figure 5 B; see Tables S7 and S9 in the SI). This result supports general-base catalysis by H254 concerted with S201 nucleophilic attack during TC formation in rhomboid proteases with an isocoumarin inhibitor. Therefore we suggest that the low general-base catalytic efficiency of His in the MC is an intrinsic feature of the rhomboid His-Ser catalytic dyad, thus leading to a higher activation barrier for the transformation of the MC into TC than that for the classical water-solvated serine proteases having the Asp-His-Ser catalytic triad. The kinetics of the TatA substrate hydrolysis by ecGlpG at 37 °C in membrane supports this statement. Whereas the K_m value for this substrate (22–135 μM)[21,22] is similar to that for the hydrolysis of a 10–15

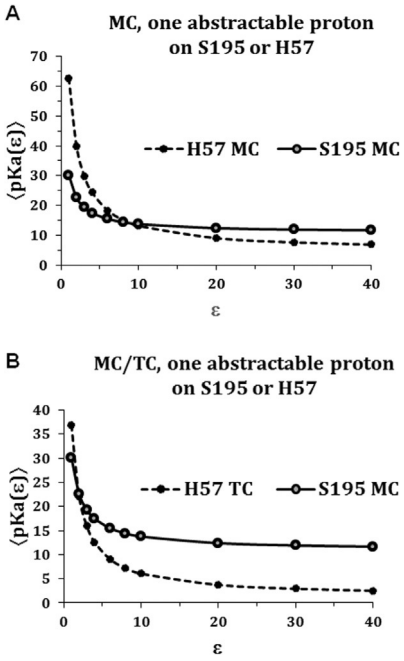

Figure 4. The calculated average $\langle pK_a(\varepsilon)\rangle$ values of a chymotrypsin–TFK complex. A) $\langle pK_a(H57,\varepsilon,MC)\rangle$ vs. $\langle pK_a(S195,\varepsilon,MC)\rangle$. B) $\langle pK_a(H57,\varepsilon,TC)\rangle$ vs. $\langle pK_a(S195,\varepsilon,MC)\rangle$.

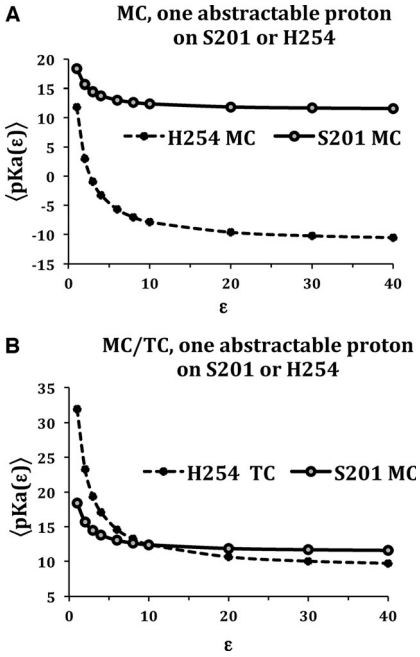

Figure 5. The calculated average $\langle pK_a(\varepsilon)\rangle$ values of a GlpG–isocoumarin complex. A) H254 $\langle pK_a(\varepsilon)\rangle$ vs. S201 $\langle pK_a(\varepsilon)\rangle$ in MC. B) H254 $\langle pK_a(\varepsilon)\rangle$ in TC vs. S201 $\langle pK_a(\varepsilon)\rangle$ in MC.

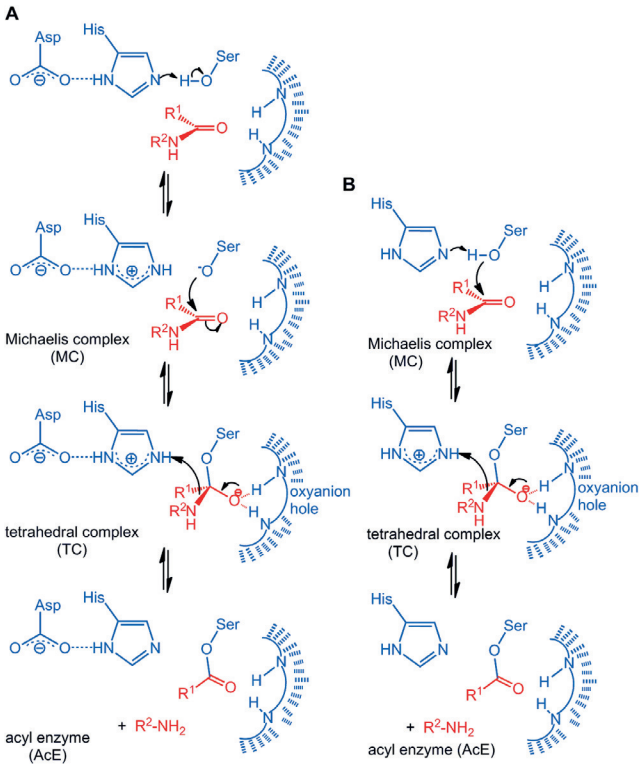

Figure 6. The first chemical steps catalyzed by serine proteases. A) A stepwise mechanism, involving a $HisH^+/SerO^-$ ion pair in the Michaelis complex, as suggested for chymotrypsin with a Ser-His-Asp catalytic triad. B) A concerted mechanism, involving a simultaneous proton transfer and nucleophilic attack, thus characterizing a rhomboid protease with a Ser-His catalytic dyad.

amino-acid peptide by trypsin (115 μM),[23] the k_{cat} values are markedly different: 0.006 s^{-1} for the rhomboid reaction[21] versus 137 s^{-1} for trypsin.[23] The k_{cat} value of trypsin is five orders of magnitude higher and is aslo manifested in a much higher catalytic efficiency (k_{cat}/K_m).

To summarize, we demonstrated that comparative analysis of the $\langle pK_a(\varepsilon)\rangle$ of the catalytic residues as a function of ε, calculated by MD-QM/SCRF(VS), could serve as a tool for interpretation of enzyme catalytic mechanisms. We found that in a rhomboid protease general-base catalysis is a concerted process of proton transfer from Ser to His and nucleophilic attack of the former on the electrophilic carbonyl of the isocoumarin inhibitor. In contrast, the mechanism of general-base catalyzed chymotrypsin inactivation by a TFK inhibitor could be described in terms of a stepwise process, with an initial proton transfer from Ser to His in the MC, with subsequent nucleophilic attack of the Ser alkoxide on the carbonyl group of TFK. This mechanism will be operative if His $pK_a(\varepsilon) \geq$ Ser $pK_a(\varepsilon)$ in the MC, conditions which result from ligand binding. Otherwise, the classical concerted mechanism will dominate. Both mechanisms are controlled by a dynamic change in the $\langle pK_a(\varepsilon)\rangle$ of the catalytic residues as a function of their progressively reduced water exposure, caused by the ligand approach into the enzyme active site. It is intriguing to suggest that these two mechanisms are also operative in peptide substrate hydrolysis, a stepwise mechanism for chymotrypsin (Figure 6A) and a concerted mechanism for rhomboid protease (Figure 6B).

Experimental Section

All molecular graphics, molecular modeling, and MD simulations were conducted by the YASARA Structure software package.[24] Quantum mechanical calculations were performed by the semi-empirical PM6 Hamiltonian[25] implemented in MOPAC2012.[26] See the Supporting Information for detailed computational protocols.

Acknowledgments

This research was supported by the Marcus Center for Medicinal Chemistry at Bar Ilan University.

Keywords: acidity · enzyme catalysis · molecular modeling · proteases · reaction mechanisms

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 1680–1684
Angew. Chem. **2016**, *128*, 1712–1716

[1] B. W. Matthews, P. B. Sigler, R. Henderson, D. M. Blow, *Nature* **1967**, *214*, 652–656.
[2] a) M. J. Page, E. Di Cera, *Cell. Mol. Sci.* **2008**, *65*, 1220–1236; b) O. D. Ekici, M. Petzel, R. E. Dalbey, *Protein Sci.* **2008**, *17*, 2023–2037; c) *Handbook of Proteolytic Enzymes*, *Vol. 3*, 3rd ed. (Eds.: N. D. Rawlings, G. S. Salvesen), Elsevier, Amsterdam, **2012**.
[3] L. Hedstrom, *Chem. Rev.* **2002**, *102*, 4501–4524.
[4] S. Urban, J. R. Lee, M. Freeman, *Cell* **2001**, *107*, 173–182.

[5] L. N. Kinch, N. V. Grishin, *Biochim. Biophys. Acta Biomembr.* **2013**, *1828*, 2937 – 2943.
[6] S. Urban, *Biochem. J.* **2010**, *425*, 501 – 512.
[7] a) O. A. Pierrat, K. Strisovsky, Y. Christova, J. Large, K. Ansell, N. Bouloc, E. Smiljanic, M. Freeman, *ACS Chem. Biol.* **2011**, *6*, 325 – 335; b) Y. Xue, S. Chowdhury, X. Liu, Y. Akiyama, J. Ellman, Y. Ha, *Biochemistry* **2012**, *51*, 3723 – 3731; c) O. Vosyka, K. R. Vinothkumar, E. V. Wolf, A. J. Brouwer, R. M. Liskamp, S. H. Verhelst, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2472 – 2477; d) E. V. Wolf, A. Zeißler, O. Vosyka, E. Zeiler, S. Sieber, S. H. Verhelst, *PLoS One* **2013**, *8*, e72307.
[8] a) H. Hu, W. Yang, *Annu. Rev. Phys. Chem.* **2008**, *59*, 573 – 601; b) H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198 – 1229; *Angew. Chem.* **2009**, *121*, 1220 – 1254; c) S. C. L. Kamerlin, M. Haranczyk, A. Warshel, *J. Phys. Chem. B* **2009**, *113*, 1253 – 1272; d) M. W. van der Kamp, A. J. Mulholland, *Biochemistry* **2013**, *52*, 2708 – 2728; e) F. Duarte, B. A. Amrein, D. Blaha-Nelson, S. C. L. Kamerlin, *Biochim. Biophys. Acta Gen. Subj.* **2015**, *1850*, 954 – 965.
[9] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227 – 249.
[10] a) A. Warshel, *Computer modeling of chemical reactions in enzymes and solutions*, Wiley, New York, **1991**; b) J. Bentzien, R. P. Muller, J. Florián, A. Warshel, *J. Phys. Chem. B* **1998**, *102*, 2293 – 2301.
[11] a) I. Ishida, S. Kato, *J. Am. Chem. Soc.* **2003**, *125*, 12035 – 12048; b) Y. Zhou, Y. Zhang, *Chem. Commun.* **2011**, *47*, 1577 – 1579.
[12] A. Warshel, G. Naray Szabo, F. Sussman, J.-K. Hwang, *Biochemistry* **1989**, *28*, 3629 – 3637.
[13] a) M. Shokhen, N. Khazanov, A. Albeck, *Proteins Struct. Funct. Bioinf.* **2008**, *70*, 1578 – 1587; b) N. Uritsky, M. Shokhen, A. Albeck, *J. Chem. Theory Comput.* **2012**, *8*, 4663 – 4671.
[14] a) M. Shokhen, N. Khazanov, A. Albeck, *ChemBioChem* **2007**, *8*, 1416 – 1421; b) M. Shokhen, N. Khazanov, A. Albeck, *Proteins Struct. Funct. Bioinf.* **2009**, *77*, 916 – 926; c) M. Shokhen, N. Khazanov, A. Albeck, *Proteins Struct. Funct. Bioinf.* **2011**, *79*, 975 – 985.
[15] M. Shokhen, M. Hirsch, N. Khazanov, R. Ozeri, N. Perlman, T. Traube, S. Vijayakumar, A. Albeck, *Isr. J. Chem.* **2014**, *54*, 1137 – 1151.
[16] D. Neidhart, Y. Wei, C. Cassidy, J. Lin, W. W. Cleland, P. A. Frey, *Biochemistry* **2001**, *40*, 2439 – 2447.
[17] K. R. Vinothkumar, K. Strisovsky, A. Andreeva, Y. Christova, S. Verhelst, M. Freeman, *EMBO J.* **2010**, *29*, 3797 – 3809.
[18] S. Urban, M. S. Wolfe, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1883 – 1888.
[19] C. S. Cassidy, J. Lin, P. A. Frey, *Biochemistry* **1997**, *36*, 4576 – 4584.
[20] V. L. Schramm, *Annu. Rev. Biochem.* **2011**, *80*, 703 – 732.
[21] S. W. Dickey, R. P. Baker, S. Cho, S. Urban, *Cell* **2013**, *155*, 1270 – 1281.
[22] E. Arutyunova, P. Panwar, P. M. Skiba, N. Gale, M. W. Mak, M. J. Lemieux, *EMBO J.* **2014**, *33*, 1869 – 1881.
[23] G. S. Coombs, A. T. Dang, E. L. Madison, D. R. Corey, *J. Biol. Chem.* **1996**, *271*, 4461 – 4467.
[24] YASARA Structure, http://www.yasara.com.
[25] J. J. P. Stewart, *J. Mol. Model.* **2007**, *13*, 1173 – 1213.
[26] J. J. P. Stewart, MOPAC 2012; Computational Chemistry; Colorado Springs, CO, **2008**. http://OpenMOPAC.net.

Received: August 19, 2015
Published online: December 21, 2015

