COMMUNICATION

Understanding substrate binding and the role of gatekeeping residues in PigC access tunnels

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Stefanie Brands,[a] Jarno G. Sikkens,[a] Mehdi D. Davari,[a] Hannah U. C. Brass,[b] Andreas S. Klein,[b] Jörg Pietruszka,[b,c] Anna Joëlle Ruff[a] and Ulrich Schwaneberg*[a,d]

Semi-rational redesign of the substrate binding pocket and access tunnels of prodigiosin ligase PigC enhanced the catalytic efficiency in synthesis of pyrrolic anti-cancer agents more than 45-times. A molecular understanding was gained on residues V333 and T334 relevant in substrate binding and translocation of small pyrroles through PigC access tunnels.

Prodigiosin ligase PigC catalyses the final step in the prodigiosin (**1a**) biosynthesis of *Serratia marcescens*.[1,2] The ATP-dependent condensation leads to the fusion of two pyrrolic prodigiosin precursors, 4-methoxy-2,2'-bipyrrole-5-carbaldehyde (MBC, **2**), and 2-methyl-3-amyl-1*H*-pyrrole (MAP, **3a**), yielding the red pigment and secondary metabolite prodigiosin (**1a**; **Figure 1**).

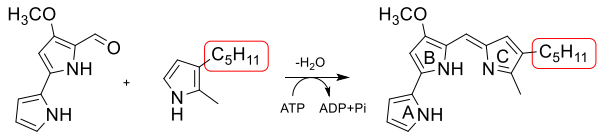

Figure 1 PigC catalysed reaction. Two pyrrolic precursors, MBC (**2**) and MAP (**3a**), are condensed to the secondary metabolite prodigiosin (**1a**) under ATP consumption. Shortening the circled aliphatic side chain at the C3 position in the pyrrole C-ring enhanced the prodiginines' anticancer bioactivity in autophagy tests.[3]

Prodigiosin (**1a**) derivatives form the compound class of prodiginines, which have attracted much attention due to their multiple bioactivities.[4,5] For instance, effects of prodiginines against plant pathogenic nematodes and fungi indicate their potential use as pesticides in agriculture,[6] while bioactivity against cancer cells hold potential for the pharmaceutical application of prodiginine compounds.[3,7] The impact of prodiginines on cancer cells in autophagy tests was most efficacious, when the length of the n-pentyl moiety shown in **Figure 1** is reduced to a n-propyl or shorter side chain.[3] Therefore, a special pharmaceutical interest lies in the synthesis of short-chain prodiginines. Because the multi-step chemical synthesis of tripyrrolic prodiginines is neither sustainable nor cost-efficient,[2,8,9] biosynthesis or a combination of chemical synthesis of prodiginine precursors with biocatalysis constitutes a promising bioeconomic alternative.[3,10,11] The bottleneck of enzymatic short-chain prodiginines synthesis is the limited substrate acceptance of the prodigiosin ligase PigC. The PigC substrate scope has already been reported in several studies,[12–17] which revealed a significant drop in PigC activity when the C3-substituent was lacking or shortened to a methyl group.[3]

In this study, we investigated and semi-rationally redesigned the PigC substrate binding pocket and access tunnels to enhance the acceptance of monopyrrole substrates with short aliphatic side chains to enable broad spectrum mutasynthesis experiments. First, target amino acid residues were selected based on their evolutionary conservation as well as their location in the substrate binding pocket or access tunnels. Subsequently, the selected target residues were saturated experimentally to determine amino acid substitutions with a beneficial effect on the acceptance of 2,3-dimethyl-1*H*-pyrrole (**3b**; **Figure S1**). Beneficial substitutions were finally analysed in *in silico* substrate docking and ligand translocation studies in a PigC structural model with the aim to gain molecular understanding and propose a role of the PigC access tunnels in the catalytic mechanism of the PigC reaction (**Figure S15**).

PigC access tunnels and the substrate binding cavity, which encloses the catalytic phosphohistidine H840, were identified by the Caver Web 1.0 tool[18] with standard settings and H840 as starting point. A large cavity was found in the core of the PigC substrate-binding domain (**Figures 2a and S11**) with a volume of 1612 Å^3 and 59 involved residues (relevance score: 100%; druggability score: 0.3; **Table S7**). Furthermore, three PigC tunnels were identified starting from the substrate binding

[a]. *Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 3, 52074 Aachen, Germany and Bioeconomy Science Center (BioSC).*
[b]. *Institute of Bioorganic Chemistry, Heinrich Heine University Düsseldorf located at Forschungszentrum Jülich, Stetternicher Forst, Building 15.8, 52426 Jülich, Germany and Bioeconomy Science Center (BioSC).*
[c]. *Institute of Bio- and Geosciences (IBG-1: Biotechnology), Forschungszentrum Jülich, 52426 Jülich, Germany.*
[d]. *DWI-Leibniz Institut für Interaktive Materialien, Forckenbeckstraße 50, 52074 Aachen, Germany*

Electronic Supplementary Information (ESI) available: Experimental and computational material and methods, catalytic PigC mechanism with access tunnels, *Michaelis-Menten* plots and substrate profiles of PigC variants, PigC ligand docking studies, conservation and evolutionary trace analysis, substrate tunnel, cavity and ligand translocation analysis. See DOI: 10.1039/x0xx00000x

pocket, two of which connected the active site with the outer solvent (tunnels 2/3 in **Figure 2b**), the first tunnel 1 being reversibly sealed by the phosphohistidine swivel domain of PigC in the depicted structural conformation (Figure S15). As second step, the short-chain monopyrrole substrate **3b** and the native PigC bipyrrole substrate MBC (**2**) were generated as ligands with the YASARA Structure 17.4.17 software suite.[19] Substrate docking studies in a 8 Å simulation cell around the catalytic phosphohistidine H840 revealed a list of 15 residues that were predicted to be involved in prodiginine formation (**Tables S4** and **S5**). By evolutionary conservation analysis (ConSurf and multiple sequence alignments),[20–22] conserved amino acid residues within the predicted substrate binding pocket and access tunnels were identified. Residues with high conservation confidence interval indices of [9,-] and ConSurf scores < -1.1 were discarded due to the risk of functional loss, and neighbouring residues selected instead (**Figure S7**). Based on these three approaches (tunnel/cavity analysis, substrate docking, and evolutionary conservation analysis), ten potential target residues were shortlisted (M327, T329, G330, V333, T334, F603, R674, A677, T678, and P680, **Figure 2b**), and cross-referenced by evolutionary trace and FuncLib analysis (**Figures S8**/**S9** and **Table S6**).[23–25] Finally, a consensus list was compiled, containing seven target residues for site-saturation mutagenesis (T329, G330, V333, T334, F603, R674, and P680; **Table 1** and **Figure 2b**, framed in cyan). All positions were located close to highly conserved regions with evolutionary importance (**Figure S9**).[23,24]

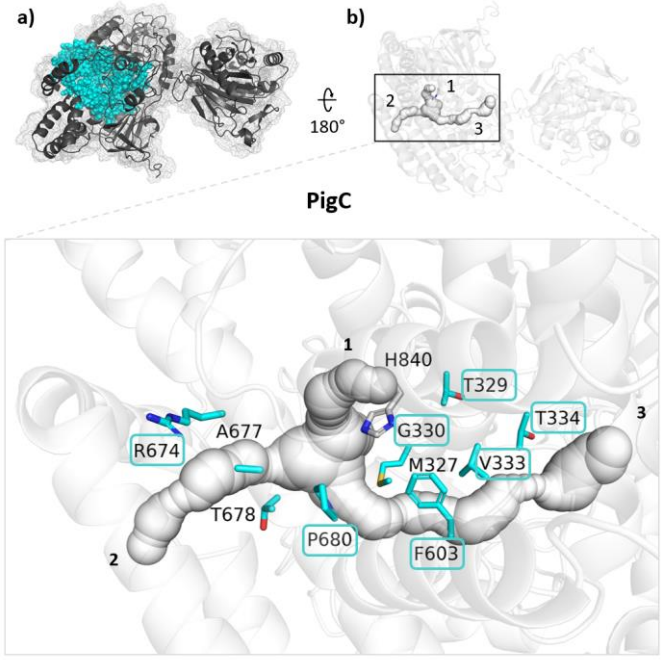

Figure 2: Identification of target residues for semi-rational engineering of PigC access tunnels. **a)** Residues forming the catalytic substrate pocket in the substrate-binding domain of PigC wild type. **b)** Access tunnels identified with the Caver Web tool 1.0.[18] Zoom: Target positions alongside PigC tunnels. Seven final positions that were selected for site-saturation mutagenesis are highlighted in frames.

Table 1 Final list of semi-rational target amino acid positions in access tunnels and the substrate binding pocket of PigC, including ConSurf[20–22] and rvET[23,24] scores.

Residue	Position	ConSurf[20–22] score* [confidence interval colour]ˣ		rvET[23,24] score†
T	329	-0.837	[8,7]	14.01
G	330	-0.472	[7,6]	16.66
V	333	-1.012	[8,7]	12.24
T	334	-0.985	[8,7]	10.46
F	603	-1.043	[8,8]	8.57
R	674	0.247	[5,4]	19.57
P	680	-0.661	[7,6]	12.98

***** Normalised ConSurf[20–22] conservation scores. Low scores imply conservation.
ˣ A confidence interval is assigned to each evolutionary conservation score by the bayesian method for calculating rates. The colour scale [cyan – 1 (low conservation) to magenta – 9 (high conservation); **Figure S7**] represents the lower and upper bounds of the confidence interval.
† Evolutionary scores for residues are calculated with real-value evolutionary trace (rvET) method.[23,24] Low scores (down to 1) imply evolutionary importance.

After selection of the seven target residues (**Table 1**), individual site-saturation mutagenesis (SSM) libraries were generated and screened with the short-chain substrate **3b** in a 96-well photometric assay in order to identify beneficial PigC variants.[26] In three of the seven SSM libraries, beneficial substitutions with an up to 6.9-times increased prodiginine formation were obtained (V333A, T334A, and R674Q/L; **Figure S2**). The single-substituted variants T334A and R674Q showed a clear preference for short-chain substrates (1.5-4.8-times higher prodiginine formation with pyrroles **3a–3c**; methyl to pentyl side chains; **Figure S5**), while variant V333A also produced slight improvements with bulkier substrates (up to 2.3-times increased prodiginine yields with pyrroles **3e–3g**). Double and triple recombination of substitutions V333A, T334A and R674Q in PigC variants V1–V4 did not further increase prodiginine yields (**Figure S5**). The latter is often the case if beneficial substitutions have an energy penalty (ΔΔG > +0.36 kcal/mol) comparable to the ones calculated for V333A and T334A (**Table S3**).[27]

PigC wild type and variants were subsequently overexpressed, quantified and kinetically characterised as previously published (**Table 2**, **Figure S3**).[10,26] Purified membrane fractions of the PigC variant V3 (T334A/R674Q) had the highest catalytic activity (k_{cat} = 3.1 ± 0.6 min^{-1}), which was 3.4-times increased over that of the PigC wild type (k_{cat} = 0.9 ± 0.1 min^{-1}). Improvement in catalytic efficiency of the triple recombination variant V4 (V333A/T334A/R674Q; k_{cat}/K_M = 27.3 mM^{-1} s^{-1}; **Table 2**, **Figure S4**) can mainly be attributed to a significant increase of affinity towards pyrrole **3b** [K_M (**2**) = 1.8 µM compared to 25.4 µM for the PigC wild type].

Substrate docking studies revealed that substitution V333A in V1, V2, and V4 enabled a beneficial change in the docking pose of substrate **3b** in the active pocket (**Figure 3a**) with a slightly enhanced binding energy (-4.4 kcal/mol) compared to the PigC wild type and PigC variants without the V333A substitution (-4.0 kcal/mol; **Figure 3b** and **Table S4**).

Table 2 Kinetic characterisation of PigC wild type and semi-rationally designed variants in *E. coli* BL21(DE3) membrane fraction with substrates **3b** and **2**.

PigC	k_{cat} [min^{-1}]	Substrate **3b**		Substrate **2**	
		K_M (**3b**) [μM]	k_{cat}/K_M (**3b**) [mM^{-1} s^{-1}]	K_M (**2**) [μM]	k_{cat}/K_M (**2**) [mM^{-1} s^{-1}]
WT	0.9 ± 0.1	25.4 ± 3.0	0.6 ± 0.1	1.6 ± 0.2	9.3 ± 0.9
V333A	1.2 ± 0.2	3.3 ± 0.3	6.1 ± 0.8	4.7 ± 0.6	4.0 ± 0.5
T334A	1.5 ± 0.1	3.6 ± 0.3	6.9 ± 0.3	1.5 ± 0.4	16.3 ± 0.7
R674Q	1.5 ± 0.2	10.4 ± 1.0	2.5 ± 0.3	1.4 ± 0.6	17.7 ± 2.3
V1 (V333A/T334A)	1.9 ± 0.3	3.3 ± 1.1	9.4 ± 1.5	11.1 ± 3.3	2.8 ± 0.4
V2 (V333A/R674Q)	2.0 ± 0.6	9.9 ± 4.3	3.4 ± 1.1	3.5 ± 0.6	9.5 ± 3.0
V3 (T334A/R674Q)	3.1 ± 0.6	4.4 ± 0.7	11.7 ± 2.3	3.7 ± 0.4	14.0 ± 2.8
V4 (V333A/T334A/R674Q)	2.9 ± 0.5	1.8 ± 0.5	27.3 ± 4.5	5.8 ± 0.7	8.3 ± 1.4

The different orientation of pyrrole **3b** in the substrate binding pocket brought substrate **3b** into closer proximity to the catalytic phosphohistidine H840 (3.5 Å compared to 6.9 Å in PigC wild type), which might be a reason for lower K_M values, higher turnover frequencies (k_{cat}), and enhanced catalytic efficiencies with monopyrrole **3b** (**Table 2**).

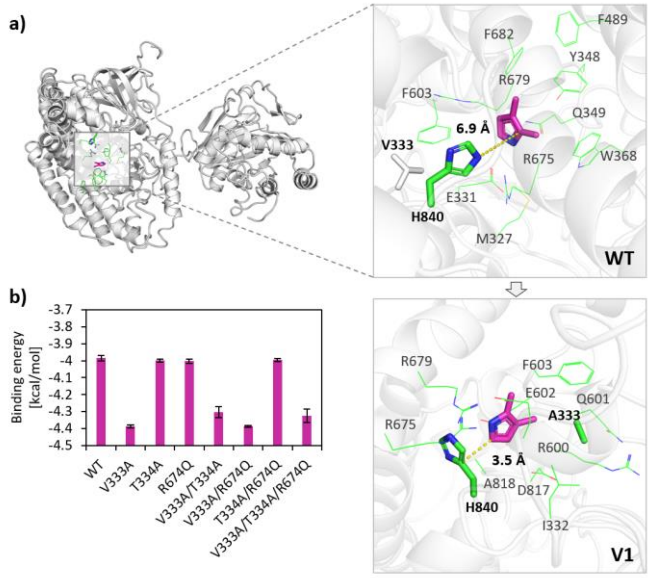

Figure 3 a) Preferred docking pose of 2,3-dimethyl-1*H*-pyrrole (**3b**) in a 8 Å simulation cell around phosphohistidine H840 in PigC WT and variant V1. **b)** Binding energies from substrate docking studies with **3b** in semi-rationally designed variants.

Substitutions T334A and R674Q had no influence on the K_M of MBC (**2**), while V333A significantly reduced MBC (**2**) affinity (K_M values increased from 1.6 μM for the PigC wild type to up to 11.1 μM for variant V1; **Table 2**). Compared to the double recombination variant V3, V4 had a slightly lower k_{cat} with monopyrrole **3b** and bipyrrole **2** (k_{cat} = 3.1 min^{-1} for V3 and 2.9 min^{-1} for V4) next to an obvious decrease in catalytic efficiency with MBC (**2**; from 14.0 to 8.3 mM^{-1} s^{-1}).

In variants V333A, V1 and V2, a slight substrate inhibition was observed at high concentrations of pyrrole **3b** over 50 μM (**Figures S3** and **S4**; K_i = 1.12 mM for V333A, 334.9 μM for V1 and 334.0 μM for V2), which was neither observed with the PigC wild type nor with variants that did not contain the substitution V333A (T334A, R674Q, and V3). Possibly, V333A promoted a competition between substrates **2** and **3b** for binding to the substrate pocket by preferring substrate **3b** over **2**, which inhibited the catalytic activity of PigC. A competitive inhibition has been observed before for other monopyrrole substrates, so far mainly reported for pyrroles with long side chains. The latter might fit into and subsequently block the MBC (**2**) binding site.[10,28] It has further already been postulated that MBC (**2**) needs to bind prior to the monopyrrole substrate to the PigC active site to initialise the reaction.[12,16] In variants harbouring V333A, the access of the larger MBC (**2**) substrate to the active pocket might be blocked by already bound **3b** molecules at higher **3b** concentrations in the new catalytically competent docking pose enabled by substitution V333A (**Figure 3**).

A second effect of the substitutions V333A and T334A in the PigC structural model was the site-specific widening of substrate tunnel 3 that connects the outside with the buried PigC active pocket in the substrate-binding domain (**Figure 4**). The PigC substrate access tunnel 3 was widened by 0.2 Å in the disk radius and by an area of ca. 4 $Å^2$ after simultaneous substitutions of V333 and T334 by the sterically less demanding alanine residues in variants V1 and V4 (**Figures 4 and S14**). The tunnel widening supposedly improved the translocation of substrates through tunnel 3 at the respective site of positions 333 and 334, which is indicated by the decreased binding energy at the respective tunnel position (approx. 15 Å; **Figure 4b**). While the hydrophobicity is unchanged in V333A, in the substitution T334A the polar residue threonine was exchanged to the inert residue alanine. The latter substitution of threonine by alanine therefore may also have had a positive influence on transport of hydrophobic substrates, because the hydroxy group of T334 could have acted as a polar barrier and gatekeeper in the tunnel, limiting substrate translocation.

Upon our analysis of the three PigC access tunnels and ligand translocation (**Tables S8 and S9**),[18,29,30] we suggest a crucial role of the three PigC access tunnels in the PigC catalytic mechanism (**Figure S15**), building on previously proposed mechanisms by Chawrai *et al*. (2012) and Picott *et al*. (2020).[16,28]

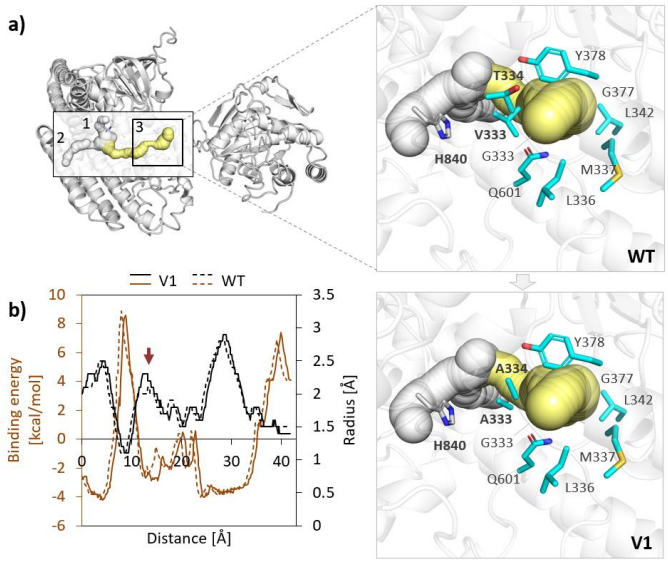

Figure 4 a) Access tunnel analysis of the PigC wild type (WT) and variant V1 (V333A/T334A), using Caver Web 1.0.[18] The zoom shows the tunnel 3 bottleneck with gatekeeper residues V333 and T334 (cyan), and the catalytic phosphohistidine H840 (grey). **b)** Ligand transport analysis by Caver analyst 2.0[29,30] of pyrrole **3b** traveling through substrate tunnel 3 of the PigC WT (dashed lines) and variant V1 (solid lines). The red arrow marks the widened tunnel site around the positions 333 and 334.

In conclusion, the PigC variant V3 (T334A/R674Q) enables synthetic access to pharmaceutically interesting short chain prodiginines (k_{cat} enhanced 3.4-times over the PigC wild type, from 0.9 min^{-1} to 3.1 min^{-1}). Molecular understanding of the gatekeeper role of residue T334 as polar barrier in substrate access tunnel 3 and of the key position 333 in the substrate binding pocket further enlightens the PigC catalytic mechanism and empowers rational PigC engineering studies to broaden the substrate profile and enhance the catalytic performance.

The scientific activities of the Bioeconomy Science Center were financially supported by the Ministry of Culture and Science within the framework of the NRW Strategieprojekt BioSC (No. 313/323-400-00213).

Conflicts of interest

There are no conflicts to declare.

References

1 N. R. Williamson, P. C. Fineran, F. J. Leeper and G. P. C. Salmond, *Nat. Rev. Microbiol.*, 2006, **4**, 887–899.
2 D. X. Hu, D. M. Withall, G. L. Challis and R. J. Thomson, *Chem. Rev.*, 2016, **116**, 7818–7853.
3 A. S. Klein, A. Domröse, P. Bongen, H. U. C. Brass, T. Classen, A. Loeschcke, T. Drepper, L. Laraia, S. Sievers, K.-E. Jaeger and J. Pietruszka, *ACS Synth. Biol.*, 2017, **6**, 1757–1765.
4 N. Darshan and H. K. Manonmani, *J. Food Sci. Technol.*, 2015, **52**, 5393–5407.
5 Z. You, S. Zhang, X. Liu, J. Zhang, Y. Wang, Y. Peng and W. Wu, *Appl. Microbiol. Biotechnol.*, 2019, **103**, 2873–2887.
6 S. S. Habash, H. U. C. Brass, A. S. Klein, D. P. Klebl, T. M. Weber, T. Classen, J. Pietruszka, F. M. W. Grundler and S. Schleker, *Front. Plant Sci.*, 2020, **11**, 579807.
7 M. L. Arellano, G. Borthakur, M. Berger, J. Luer and A. Raza, *Clin. Lymphoma Myeloma Leuk.*, 2014, **14**, 534–539.
8 H. Rapoport and K. G. Holden, *J. Am. Chem. Soc.*, 1962, **84**, 635–642.
9 A. Fürstner, *Angew. Chem. Int. Ed.*, 2003, **42**, 3582–3603.
10 H. U. C. Brass, A. S. Klein, S. Nyholt, T. Classen and J. Pietruszka, *Adv. Synth. Catal.*, 2019, **361**, 2659–2667.
11 A. Domröse, A. S. Klein, J. Hage-Hülsmann, S. Thies, V. Svensson, T. Classen, J. Pietruszka, K.-E. Jaeger, T. Drepper and A. Loeschcke, *Front. Microbiol.*, 2015, **6**, 972.
12 Q. Long, D. E. Jeffries, S. Lin, X. Chen, W. He, Y. Wang, Z. Deng, C. W. Lindsley and M. Tao, *Appl. Environ. Microbiol.*, 2020, **86**, e02331-19.
13 Z. You, X. Liu, S. Zhang and Y. Wang, *Biochem. Eng. J.*, 2018, **134**, 1–11.
14 A. S. Klein, H. U. C. Brass, D. P. Klebl, T. Classen, A. Loeschcke, T. Drepper, S. Sievers, K.-E. Jaeger and J. Pietruszka, *ChemBioChem*, 2018, **19**, 1545–1552.
15 M. Couturier, H. D. Bhalara, S. R. Chawrai, R. Monson, N. R. Williamson, G. P. C. Salmond and F. J. Leeper, *ChemBioChem*, 2020, **21**, 523–530.
16 S. Chawrai, N. Williamson, T. Mahendiran, G. P. Salmond and F. Leeper, *Chem. Sci.*, 2012, **3**, 447–454.
17 S. Chawrai, N. Williamson, G. P. Salmond and F. Leeper, *Chem. Commun.*, 2008, **16**, 1862–1864.
18 J. Stourac, O. Vavra, P. Kokkonen, J. Filipovic, G. Pinto, J. Brezovsky, J. Damborsky and D. Bednar, *Nucleic Acids Res.*, 2019, **47**, W414–W422.
19 E. Krieger and G. Vriend, *Bioinformatics*, 2014, **30**, 2981–2982.
20 H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2016, **44**, W344–W350.
21 H. Ashkenazy, E. Erez, E. Martz, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2010, **38**, W529–W533.
22 G. Celniker, G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko and N. Ben-Tal, *Isr. J. Chem.*, 2013, **53**, 199–206.
23 I. Mihalek, I. Reš and O. Lichtarge, *J. Mol. Biol.*, 2004, **336**, 1265–1282.
24 O. Lichtarge, H. R. Bourne and F. E. Cohen, *J. Mol. Biol.*, 1996, **257**, 342–358.
25 O. Khersonsky, R. Lipsh, Z. Avizemer, Y. Ashani, M. Goldsmith, H. Leader, O. Dym, S. Rogotner, D. L. Trudeau, J. Prilusky, P. Amengual-Rigo, V. Guallar, D. S. Tawfik and S. J. Fleishman, *Mol. Cell*, 2018, **72**, 178-186.e5.
26 S. Brands, H. U. C. Brass, A. S. Klein, J. Pietruszka, A. J. Ruff and U. Schwaneberg, *Chem. Commun.*, 2020, **56**, 8631–8634.
27 H. Cui, H. Cao, H. Cai, K.-E. Jaeger, M. D. Davari and U. Schwaneberg, *Chem. – Eur. J.*, 2020, **26**, 643–649.
28 K. J. Picott, J. A. Deichert, E. M. deKemp, V. Snieckus and A. C. Ross, *ChemBioChem*, 2020, **20**, 1036–1042.
29 A. Jurcik, D. Bednar, J. Byska, S. M. Marques, K. Furmanova, L. Daniel, P. Kokkonen, J. Brezovsky, O. Strnad, J. Stourac, A. Pavelka, M. Manak, J. Damborsky and B. Kozlikova, *Bioinformatics*, 2018, **34**, 3586–3588.
30 E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLOS Comput. Biol.*, 2012, **8**, e1002708.