Three Regioselectively Acylated Flavonoid Aglycone Derivatives in Equimolar Yield at One Blow

Eleni Kyriakou,[a] Androniki D. Kostagianni,[b] Tahsin F. Kellici,[b,d] Efstathia Giannopoulou,[c] Konstantinos E Siatis,[c] Nisar Sayyad,[b] Haralabos P Kalofonos,[c] Thomas Mavromoustakos,[d] Haralambos Stamatis,[e] and Andreas G. Tzakos\*[b]

[a] *E. Kyriakou*

Institute of Structural Biology, Helmholtz Zentrum München,

85764 Neuherberg, Germany.

[b] *A.D. Kostagianni, Dr. T.F. Kellici, Dr. N. Sayyad, Prof. Dr. A.G. Tzakos*  
Department of Chemistry  
University of Ioannina  
Ioannina 45110, Greece  
E-mail: agtzakos@gmail.com

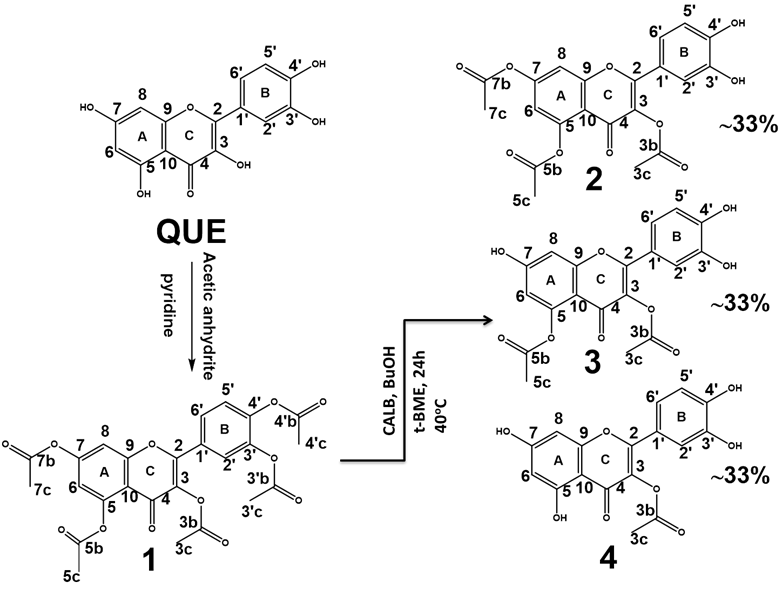
[c] *Dr. E. Giannopoulou, K.E. Siatis, Prof. Dr. H.P. Kalofonos*  
Department of Medicine,  
University of Patras  
Rio, 26504, Greece

[d] *Dr. T.F. Kellici, Prof. Dr. T. Mavromoustakos*  
Department of Chemistry  
National and Kapodistrian University of Athens  
Panepistimiopolis Zografou, Greece

[e] *Prof. Dr. H. Stamatis*  
Department of Biological Applications and Technologies  
University of Ioannina

Supporting information for this article is given via a link at the end of the document

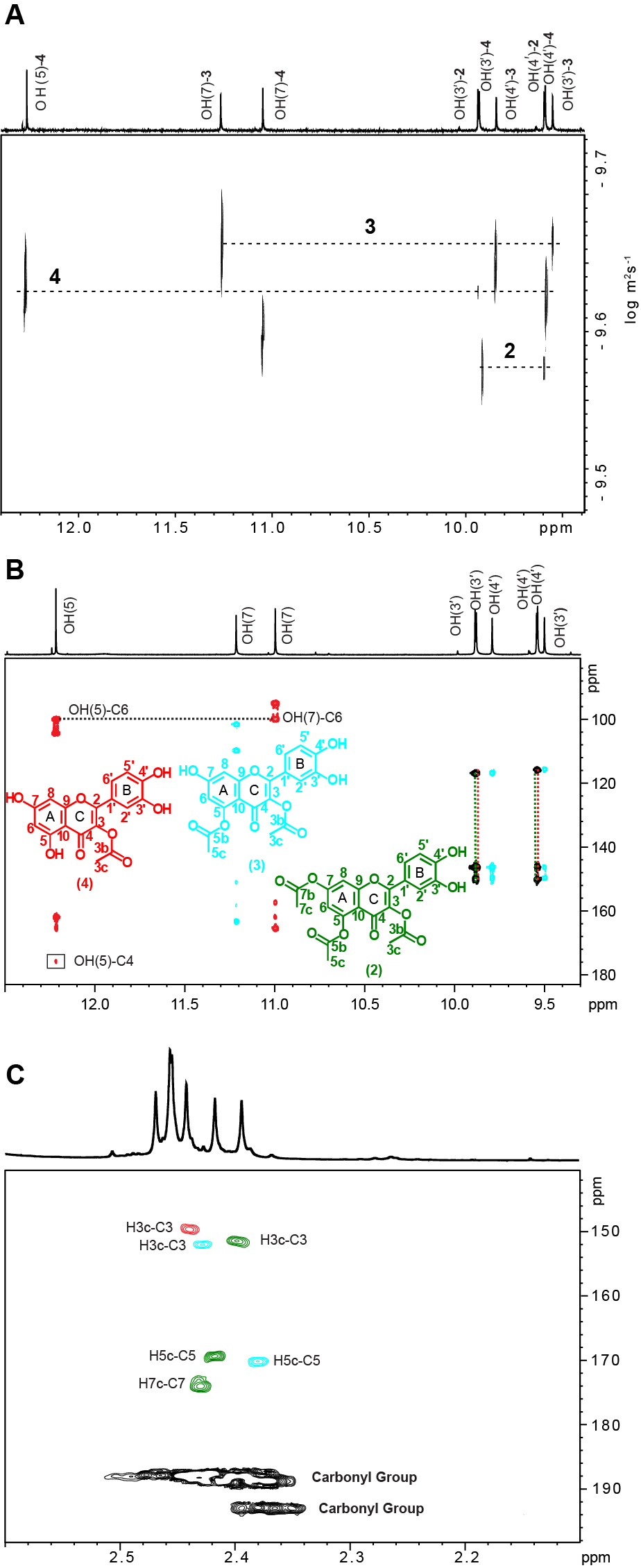
**Abstract:** A simple and time-effective chemoenzymatic process to obtain three regioselective acylated quercetin analogues in a 1:1:1 ratio in one pot is described. This process overcomes the inherent scaffold intricacy and synthetic complexity of quercetin. Cell proliferation experiments in three breast cancers cell lines pinpoint the high potency of the generated compounds.

The necessity for generating novel chemical space in a rapid way to strive to deliver potent anticancer agents is conspicuous. This is the case for breast cancer due to its complexity and heterogeneity. Although, natural products (NPs) have served as a rich source of anticancer drugs, this chemical space has been abandoned, due to their synthetic complexity. Phenol-containing NPs, the so called phenolome, such as flavonoids,[1] can serve as noteworthy anticancer agents. Thinking of the different possible combinations for decorating the NP phenol core with different chemotypes, especially for flavonoids containing up-to 5 phenol –OH groups, reveals the capacity to largely expand this unexploited chemical space. However, targeted derivatization of NPs, in an effort to enhance their bioactivity, is tedious and particularly generating regioselective substitution of NPs is laborious and cumbersome. Quercetin (QUE), a flavonoid aglycone, has illustrated a broad spectrum of bioactivities, especially due to its potential to kill cancer cells, while sparing normal cells.[2] Recently, a QUE-(2-hydroxypropyl)-β-cyclodextrin (HP-β-CD) complex was reported by us with amplified aqueous solubility and significant bioactivity in T24 human bladder cancer.[3] After using a multidisciplinary approach employing biochemical and physicochemical assays we revealed that QUE binds directly to the BH3 domain of Bcl-xL and Bcl-2 proteins, thus promoting cancer cell apoptosis.[4] Recent studies evidenced that although the phenol-OH groups, decorating QUE’s core, are the most important chemotypes conferring bioactivity, their protection by O-acetylation or O-methylation led to enhanced anticancer activity, intestinal absorption and resistance to hepatic metabolism with respect to the parent compound.[5] Also, as we recently illustrated, conjugation with amino-acids to QUE enhances its selective cytotoxic activity in cancer cells through targeting peptide transporters overexpressed in several cancers.[6] Therefore, a regioselective protection of these groups is highly desirable.[7] To this end several studies have illuminated the amplified potency gained after acylating the hydroxyl groups of QUE,[5b, 8] thus, much effort has been committed for the selective insertion of different groups on the QUE’s core.[9] In most of these studies to produce a regioselectively acylated core is highly challenging due to the numerous reactive hydroxyl groups present in the parent molecule, thus requiring tedious protection/deprotection as also purification and characterization steps. This could be surmounted through tailoring chemoezymatic processes.

**Figure 1:** Alcoholysis of pentaacetate quercetin (0.7 eq) catalyzed by CALB lipase in *t*-butyl methyl ether with n-butanol (1eq) at 40oC.

We have reported that *Candida antarctica* lipaseB (CALB) can transform flavonoid aglycones to their relevant esters in organic solvents using proper acyl donors,[10] whereas alcoholysis reactions have been reported for flavonolignans consisted of a flavonoid and a lignan part.[11] Along these lines, we developed a simple and time-effective process, the "NMR-tube bioreactor", allowing to monitor in situ the transformation efficacy and regioselectivity in multiple substrate biotransformations.[12] No enzymatic alcoholysis reaction has been reported thus far for esters of flavonoid aglycones. To explore the potential regioselective derivatization of QUE we first synthesized chemically pentaacetate quercetin **(1)** and then utilized CALB in alcoholysis reactions (Figure 1). Alcoholysis of **(1)** (0.7 eq) was catalyzed by CALB in *t*-butyl methyl ether with n-butanol (1eq).

We have formerly described a methodology to enhance the quality of the 1H NMR spectrum of a reaction mixture containing phenol chemotypes through tuning the proton chemical exchange by adding picric acid.[13] We thus applied this methodology in the alcoholysis reaction mixture we obtained using 30.24 nmol of picric acid [13]. After the addition of picric acid, nine peaks in the region of the phenol-OH groups appeared with equal integrals (1:1:1) pinpointing the presence of products with equimolar yield (Figure S1). To assign the derived products 2D 1H-13C HMBC (Figures 2B, C and Figure S2) and 1H-13C HSQC NMR experiments (Figure S3) were conducted on the crude reaction mixture. Several long range connectivities of 2-4J(1H,13C) couplings were observed, allowing the complete assignment of the carbon skeleton of the derivatives. However, assigning the –OH groups to each derivative was difficult due to peaks overlapping and the fact that position 3 is acetylated making it impossible to connect the benzopyrone system with ring B. To facilitate the assignment of the products of the crude enzymatic reaction, 2D Diffusion Ordered NMR Spectroscopy (DOSΥ) was used since as we have also reported it provides a way to separate the different compounds in a mixture based on their different translation diffusion coefficients.[14] As can be seen in (Figure 2A, Figure S4) the DOSY spectrum was highly resolved by the use of picric acid.[13] This spectrum clearly presents the existence of the three different reaction products. Specifically, in the selected –OH region three distinct sets of aligned proton cross peaks are evident. One of the products contains four OH groups and has a diffusion coefficient of DOH**(4)** ≈ 2.5 x 10-10 m2 s-1, a second product with DOH**(3)** ≈ 2 x 10-10 m2 s-1 contains three –OH groups and the third product with DOH**(2)** ≈ 3 x 10-10 m2 s-1 has two –OH groups. Using the information derived from the DOSY experiment, we assigned the HMBC spectrum of the reaction mixture without separating the products. More specifically, the OH groups at 12.21 ppm, 11 ppm, 11.93 ppm and 9.58 ppm correspond to analogue 3-acetoxy-tetrahydroxyflavone **(4)**, since the cross peak of the –OH group at 12.21 ppm to carbon C4 indicates that it is OH(5) group. The -OH at 11 ppm belongs to the OH(7) group of the same analogue, since both groups have a common cross peak to carbon C6 (Figure 2B). The OH groups at 11.22 ppm, 9.84 ppm and 9.55 ppm correspond to 3,5-diacetoxy-7,3΄,4΄-trihydroxyflavone analogue **(3)**, since the -OH group at 11.22 ppm is an OH(7) , and not OH(5), as it lacks cross peaks to the characteristic carbon C4. The –OH groups at 11.94 ppm and 9.59 ppm correspond to 3,5,7-triacetoxy-3΄,4΄-dihydroxyflavone **(2)** analogue, since the two OH groups appear in the characteristic region of the OH groups from the B ring of the flavonoids. Further assignment of the analogues was achieved following the cross peak connectivity in the aliphatic region where the characteristic acyl groups are showing up (Figure 2C). Having this information in hand it could be suggested that **(1)** undergoes initial alcoholysis at positions 3΄,4΄ to afford **(2)** followed by cleavage of the acetyl groups at position 7 and 5 to yield **(3)** and (**4)**,respectively.

`

**Figure 2.** **(A)** 500 MHz 1H NMR DOSY of the crude enzymatic reaction in DMSO-*d6* (290 K, ∆= 100 ms, δg = 4.2 ms, ns = 16, total experimental time 27 min) after the addition of 30.24 nmol picric acid (3μl from a 10.18 mM stock solution). **(B**), **(C)** Selected regions of the 500 MHz 1H-13C HMBC NMR spectra of the mixture in 400 μl DMSO-d6 (T= 298 K, number of scans = 90**).** The assignment of the –OH area is illustrated in (B). The black dot lines are showing the connectivity between the OH(5) and OH(7) for compound **4**, while the vertical green and red dot lines are indicating to which compound (**4**, red, or **2**, green) corresponds each overlapped OH group. In (**C**) is illustrated the aliphatic area. The 2D structures of the three derivatives are also shown.

Based on the fact that the three products were present in the reaction mixture with the same yield (their integrals in the 1H-NMR were 1:1:1 (Figure S1) and on the basis that a predefined amount of picric acid was added in this mixture (Molar ratio 18.8:1 of each product: picric acid) we estimated that 0.57 μmol of each product were in the crude reaction mixture. The same three products were consistently received from this regioselective reaction (from 27 probable combinations) and modifying the amount of the enzyme, reaction temperature or time of the reaction led to different reaction yields for the products.

*In silico* studies were used to provide a potential rationale on the way that CALB resulted on the specific regioselective products. In the first step, compound **(1)** and the three analogues were docked in the active site of CALB. The ring B of pentaacetate quercetin is directed in the active site of the enzyme. The acetyl moiety attached at the 3΄ position develops three hydrogen bonds with Thr40 and one with Gln106, belonging to the oxyanion hole. A hydrogen bond, π-π and π-cation interactions are also formed between His224 and the acetyl moiety in the 4΄ position. The preferred docking position for the pentaacetated analogue is shown in Figure 3A and could explain the fact that hydrolysis occurs firstly in the 3΄ and 4΄ positions. To further explore the binding mode of compound **(1)** the atomic charges were calculated (Table S1). The carbons 3΄b and 4΄b have partial charges that correspond to 0.7323 and 0.7275 (C32 and C35 in Table S1). As a result a nucleophilic attack is favored in these two positions. Although carbon 7b has a slightly higher positive charge (C26 in Table S1) the nucleophilic attack does not happen there due to the unfavorable orientation it adopts in the active site.

The docking position of analogue **(2)** is shown in Figure 3B. The ring A is directed towards the active site. The acetyl group in position 7 is stabilized by a hydrogen bond with Thr40. The complex is further stabilized with a hydrogen bond with Ala282. The compound **(3)** is also positioned with the same direction in the active site of CALB (Figure 3C). The acetyl moiety in position 5 develops multiple hydrogen bonds with Thr40 and Gln106. π-π interactions are also developed between ring A and His224. A further hydrogen bond is formed between the hydroxyl group in position 3΄ and Leu140. Since the acetyl part is near Ser105 this could explain the fact that the subsequent hydrolysis happens in position 5 and not in position 3. The predicted position of the monoacetyl analogue **(4)** is shown in Figure 3D. Ring A is positioned in the catalytic site. The phenolic OH(5) develops hydrogen bonds with Gln106 and Thr40. Ring A develops π-π interactions with Trp104 and His224. Similarly, a hydrogen bond is formed between the OH(3΄) and Leu140. The acetyl moiety in position 3 remains far away from the catalytic serine and this is probably the reason why hydrolysis in this position does not occur.

In the second step the compounds were covalently bound in the active site of CALB by creating a bond with Ser105. This was done to determine the way that the tetrahedral intermediate is stabilized for each compound. The CovDock protocol was used to predict the pose of the covalently bound quercetin derivatives in the CALB active site.[15]

The docking pose for the tetrahedral intermediate of pentaacetate quercetin is shown in Figure S5A. The oxygen

|  |  |  |  |
| --- | --- | --- | --- |
| C:\Users\woland\Desktop\PhD\calb rev et al\new photos\fig 2 new.tif  **Figure 3:** Docking positions for: A) pentaacetate quercetin; B) 3,5,7-triacetoxy-3΄,4΄-dihydroxyflavone; C) 3,5-diacetoxy-7,3΄,4΄-trihydroxyflavone and D) 3-acetoxy- tetrahydroxyflavone. The ligands are shown in grey while the enzyme residues are shown in green. The hydrogen bonds are shown in blue dotted lines while π-π interactions are shown in turquoise.  anion created from the reaction process is stabilized through two hydrogen bonds with Gln106 and Thr40. The MMGBSA energy (prime energy using the Molecular Mechanics Generalized Born Surface Area)-69.478 kcal/mol (Table S2) and is the highest score with respect to the other three ligands.  In Figure S5B is shown the tetrahedral intermediate of compound **(2)**. The intermediate is stabilized by the same interactions as previously. The MMGBSA score was lowered to -61.401 kcal/mol. In the case of analogue **(3)** (Figure S5C) the MMGBSA score is 2.95 kcal/mol higher, when compared to analogue **(2)** (Table S2). This could explain the fact that the hydrolysis in position 7 happens before the one in position 5. The score for the monoacetyl analogue **(4)** is 14.1 kcal/mol higher than the one of analogue **(3).** The position that the compound **(4)** has to adopt to achieve the tetrahedral intermediate in the active site of the enzyme is stereochemically obstructed. This is probably the main reason that the hydrolysis of position 3 is never observed (Figure S5D).  To determine the biological activity of the three products we isolated them from the crude enzymatic reaction through HPLC. The three products could be easily separated with the first fraction being compound **(3)**, the second compound **(4)** and the last the compound **(2)** (Fig S6). The purity of each derivative was further validated by 1H-NMR (Figures S7-10). Having the three pure compounds in hand we evaluated their anti-tumor effect in various cancer cells. Based on our recent data demonstrating that the parent compound, QUE, targets Bcl-xl and Bcl-2,[16] and induced cancer cell death through ERα and ERβ dependent mechanisms,[17] we selected the breast cancer cell lines MCF-7, T47D and SKBR-3. These cell lines express different Bcl-2, ERα and ERβ protein levels, thus it is interesting to explore the ability of QUE and the other compounds to inhibit cell growth. All tested analogues were added to cells and 48 h later, cell number was calculated using the MTT assay. We found that the analogues exerted a better anti-proliferative effect in MCF-7 and T47D cells compared to SKBR-3 cells. The IC50 values were extrapolated from dose response analysis (Figures S11 A, B and C) as shown in Table 1. Table 1 illustrate that pentaacetate quercetin **(1)** and especially derivative **(2)** exert the most potent inhibitory effect on MCF-7 and T47D cell proliferation. Cells were classified according to Bcl-2, ERα and ERβ status. MCF-7 cells express high levels of Bcl-2 and contain a high ratio of ERα/ERβ. T47D cells express low Bcl-2 protein levels and low ratio of ERα/ERβ. SKBR-3 cells express medium levels of Bcl-2 and are ERα/ERβ negative.[18] Our data indicate that QUE as well as all tested analogues exerted a stronger effect in cells expressing high levels of Bcl-2 and ERα/ERβ ratio. To rationalize the different recorded bioactivities of the quercetin analogs they were docked in the structures of Bcl-2 (Figures S12, S13) and Bcl-xL (Figures S14, S15) and we found that compounds 1 and 2 presented the most favorable interactions. Calculations using Prime-MMGBSA (Table S3) further showed that compounds **1** and **2** adopt the most favorable binding energies for both Bcl-xL and Bcl-2. | | | |
| **Table 1**: The IC50 values of **1**, **2**, **3**, **4** and quercetin on cell proliferation of breast cancer cell lines MCF-7, SKBR-3 and T47D, as extrapolated from data analysis. | | | |
| **Compound Nr.** | **IC50 (μΜ)** | | |
|  | **MCF-7** | **SKBR-3** | **T47D** |
| **1** | **8,0 ± 2,0** | 20,1 ± 7,7 | **10,5 ± 1,9** |
| **3** | 24,7 ± 5,4 | > 80 | 45,3 ± 4,7 |
| **2** | **11,7 ± 3,9** | **12,7 ± 3,3** | **9,4± 6,2** |
| **4** | 16,1 ± 4,1 | > 80 | 31,2± 5,5 |
| **Quercetin** | 20,6 ± 6,7 | 26,5 ± 8,9 | 17,7 ± 6,6 |
|  | | | |

On the basis of these results we can suggest that regioselective protection of the phenol functional groups of QUE is required for the synthetic analogues to exert their biological activity. In addition, it was pinpointed that the –OH(5) group should be protected for maximal activity. We would expect that similar behavior could potentially be obtained for other flavonoids of the same family, but further research should be done to confirm this hypothesis.

Herein, we reported a novel chemoenzymatic synthetic process ebabling the production in one pot of three regioselectively acetylated QUE analogues at equimolar yield. To our knowledge this is the first reported example of enzymatic alcoholysis reaction for esters of flavonoid aglycones. Molecular docking experiments provided an explanation of the regioselectivity of the deacetylation and the three products were characterized extensively by 1H-NMR, 2D 1H-13C HMBC, 1H-13C HSQC and 1H NMR DOSY. CALB was found to regioselectively recognize more favorably the ester group at C-3΄, C-4΄, then the ester group at C-7 and finally the C-5 more favorably leading to three products, **(2)**, **(3)** and **(4)**. Importantly, the three products were obtained in a 1:1:1 ratio and their separation and isolation was easily performed through HPLC. The isolated compounds were evaluated for their potency to inhibit cell growth in three breast cancer cell lines MCF-7, T47D and SKBR-3 expressing different Bcl-2, ERα and ERβ protein levels. The product **(2)** exerted the most potent inhibitory effect on SKBR-3 and T47D cell proliferation even though this derivative shows equally important cytotoxicity in MCF-7 cells as also validated by *in silico* studies. This green-, time-, and cost-effective process provides a simple route from natural products to lead compounds for drug discovery.

Supporting Information Summary

All experimental details, compound characterization, NMR spectral data and Docking of the analogues to Bcl-2 and Bcl-xl are compiled in the supporting information.

**Keywords:** Biocatalysis • Cancer • Quercetin • In Silico Studies • NMR experiments •Regioselectivity

[1] K. A. Steinmetz, J. D. Potter,  *J. Am. Diet. Assoc.* **1996**, *96*, 1027-1039.

[2] A. V. Anand David, R. Arulmoli, S. Parasuraman, *Pharmacogn. Rev.* **2016**, *10*, 84-89.

[3] T. F. Kellici, M. V. Chatziathanasiadou, D. Diamantis, A. V. Chatzikonstantinou, I. Andreadelis, E. Christodoulou, G. Valsami, T. Mavromoustakos, A. G. Tzakos, *Int. J. Pharm*. **2016**, *511*, 303-311.

[4] A. Primikyri, M. V. Chatziathanasiadou, E. Karali, E. Kostaras, M. D. Mantzaris, E. Hatzimichael, J. S. Shin, S. W. Chi, E. Briasoulis, E. Kolettas, I. P. Gerothanassis, A. G. Tzakos, *ACS Chem. Biol*. **2014**, *9*, 2737-2741.

[5] a) R. Bernini, F. Crisante, M. C. Ginnasi, *Molecules* **2011**, *16*; b) S. Rubio, J. Quintana, J. L. Eiroa, J. Triana, F. Estévez, *Carcinogenesis* **2007**, *28*, 2105-2113.

[6] a) M. V. Chatziathanasiadou, E. G. Geromichalou, N. Sayyad, E. I. Vrettos, A. Katsikoudi, E. Stylos, S. Bellou, G. D. Geromichalos, A. G. Tzakos, *Amino Acids* **2018**, *50*, 279-291; b) T. F. Kellici, M. V. Chatziathanasiadou, M. S. Lee, N. Sayyad, E. G. Geromichalou, E. I. Vrettos, A. D. Tsiailanis, S. W. Chi, G. D. Geromichalos, T. Mavromoustakos, A. G. Tzakos, *Org. Biomol. Chem*. **2017**, *15*, 7956-7976.

[7] a) N. Dai, Y. N. Teo, E. T. Kool, *Chem. Commun*. **2010**, *46*, 1221-1223; b) F. Torres, J. Quintana, F. Estévez, *Mol. Carcinog*. **2010**, *49*, 464-475.

[8] a) I. Milackova, L. Kovacikova, M. Veverka, J. Gallovic, M. Stefek, *Interdiscip. Toxicol*. **2013**, *6*, 13-17; b). K. Sakao, M. Fujii, D. X. Hou, *BioFactors* **2009**, *35*, 399-405.

[9] a) S. L. M. Bouktaib, A. Atmani and C. Rolando, *Tetrahedron* **2002**, *58*, 10001-10009; b) Z. Cao, Y. Qu, J. Zhou, W. Liu and G. Yao, *J. Carbohydr. Chem*. **2015**, *34*, 28-40; c) V. Kumar, F. Jahan, R. V. Mahajan, R. K. Saxena, *Bioresour. Technol.* **2016**, *218*, 1246-1248; d) E. Vavříková, F. Langschwager, L. Jezova-Kalachova, A. Křenková, B. Mikulová, M. Kuzma, V. Křen, K. Valentová, *Int. J. Mol. Sci.* **2016**, *17*, 899.

[10] E. Kyriakou, A. Primikyri, P. Charisiadis, M. Katsoura, I. P. Gerothanassis, H. Stamatis, A. G. Tzakos, *Org. Biomol. Chem.* **2012**, *10*, 1739-1742.

[11] E. Vavříková, P. Gavezzotti, K. Purchartová, K. Fuksová, D. Biedermann, M. Kuzma, S. Riva, V. Křen, *Int. J. Mol. Sci.* **2015**, *16*, 11983-11995.

[12] A. V. Chatzikonstantinou, M. V. Chatziathanasiadou, E. Ravera, M. Fragai, G. Parigi, I. P. Gerothanassis, C. Luchinat, H. Stamatis, A. G. Tzakos, *Biochim Biophys Acta* **2018**, *1862*, 1-8.

[13] A. P. P. Charisiadis, V. Exarchou, A. Tzakos, I. P. Gerothanassis, *J. Nat. Prod* **2011**, *74*, 2462-2466.

[14] A. Primikyri, E. Kyriakou, P. Charisiadis, C. Tsiafoulis, H. Stamatis, A. G. Tzakos, I. P. Gerothanassis, *Tetrahedron* **2012**, *68*, 6887-6891.

[15] a) K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid, E. Harder, *J Chem Inf Model* **2014**, *54*, 1932-1940; b) T. F. Kellici, T. Mavromoustakos, D. Jendrossek, A. C. Papageorgiou, *Proteins* **2017**, *85*, 1351-1361.

[16] L. G. Costa, J. M. Garrick, P. J. Roquè, C. Pellacani, *Oxid. Med. Cell. Longevity* **2016**, *2016*, 2986796.

[17] a) E. Elstner, E. A. Williamson, C. Zang, J. Fritz, D. Heber, M. Fenner, K. Possinger, H. P. Koeffler, *Breast Cancer Res. Treat.* **2002**, *74*, 155-165; b) J. Sastre-Serra, M. Nadal-Serrano, .D. G. Pons, A. Valle, J. Oliver and P. Roca, *Cell. Physiol. Biochem.* **2012**, *29*, 261-268.

**Entry for the Table of Contents** (Please choose one layout)

Layout 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| COMMUNICATION | | | | |
|  |  |  |  | Eleni Kyriakou,Androniki D. Kostagianni, Tahsin F. Kellici, Efstathia Giannopoulou, Konstantinos E Siatis, Nisar Sayyad, Haralabos P Kalofonos, Thomas Mavromoustakos, Haralambos Stamatis, and Andreas G. Tzakos\*  Page No. – Page No.  Three Regioselectively Acylated Flavonoid Aglycone Derivatives in Equimolar Yield at One Blow |

A chemoenzymatic process to generate three regioselectively acylated quercetin derivatives in equimolar yield at one pot is reported. *In silico* studies suggest a rationale on the enzymatic formation of the three products. Cell proliferation experiments in three breast cancers cell lines unveiled the high potency of the generated compounds.