Supplementary information for:

High Level QM/MM Modelling Predicts an Arginine as the Acid in the Condensation Reaction Catalyzed by Citrate Synthase

Marc W. van der Kamp, Francesca Perruccio and Adrian J. Mulholland**

School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK

* adrian.mulholland@bristol.ac.uk or marc.vanderkamp@bristol.ac.uk

Contents

Computational procedures

System preparation and setup

A model of the enzyme substrate system was prepared as reported previously.[1] In short, the dimer form of the crystal structure of chicken CS co-crystallized with acetyl-CoA and R-malate (PDB entry code 4CSC) was taken and malate was replaced by OAA. The system was solvated (with TIP3P[2] water) and reduced to an 18 Å sphere centered on the terminal acetyl-CoA carbon. The sphere contains all catalytically important residues,[3] the whole of acetyl-CoA and OAA and all residues involved in binding them. 'Stochastic boundary' conditions[4] were applied with a 14 Å reaction region.

MM parameters

For the protein and water molecules, the CHARMM27[5, 6] all atom force field was used. Atom types and partial charges for acetyl-CoA and OAA were based on the types and partial charges used in the CHARMM27 parameter set for similar groups as reported previously[1]. Additional bond, bond angle and dihedral angle parameters were developed in analogy with existing CHARMM27 parameters[1]. In case of the O=C-COO^- dihedral angle in OAA, the dihedral term was fit to the RHF/6-31+G(d) QM energy profile for rotation. This level of theory and basis set have been shown to give good results for the OAA geometry and dihedral rotation barrier (e.g. in comparison to MP2 results)[7].

QM/MM calculations: AM1/CHARMM27 level

For QM/MM calculations, the 18 Å radius system was partitioned into a QM region treated by the semiempirical method AM1 and a MM region treated by the CHARMM27 potential. All atoms more than 16 Å away form the terminal methyl group of acetyl-CoA were kept fixed. The QM region always included the thioester part of acetyl-CoA (equivalent to methylthioacetate) and the side chain of Asp375 and OAA. Depending on the acid for the proton transfer in the condensation step, side chains of either Arg329, His320 or His238 was also included. In case Arg329 was included, then the side chain of Asp327 was also included. The total charge in the QM region was −3 in all cases. Link atoms[8] were used to model bonds crossing the

QM/MM boundary. MM charges of groups adjacent to link atoms were excluded from interaction with the QM region (using the EXGR command in CHARMM). Optimized Van der Waals parameters for OAA oxygens (when treated in QM region) and the thioester oxygen, sulfur and terminal carbon were used (based on B3LYP/6-311+G(d,p)(6-31+G(d)) results on model compounds, unpublished results).

Potential energy profiles were obtained by adiabatic mapping along the reaction coordinates defined. All systems were initially equilibrated using QM/MM MD for at least 80 ps. Initial conformations for the potential energy profiles were obtained from QM/MM umbrella sampling dynamics simulations with a harmonically restrained reaction coordinate value or of QM/MM dynamics simulations of the enolate intermediate without restraint, for at least 30 ps. Subsequently, conformations were optimized by 250 steps of steepest descents followed by adjusted-basis Newton Rapson (ABNR) minimization until a gradient of 0.001 kcal mol^{-1} $Å^{-1}$ was reached.

Using this starting conformation, a series of energy minimizations consisting of 250 steps of Steepest Descents followed by ABNR minimization to a gradient of 0.00001 kcal mol^{-1} $Å^{-1}$ was performed (equivalent to machine convergence). A harmonic force was applied to the reaction coordinate. Subsequent minimizations along the reaction coordinate were performed with a step size of 0.1 Å using the minimized conformation of the preceding target value as the starting point.

QM/MM calculations: high level

B3LYP/6-31+G(*d*)/CHARMM27 geometry optimization was carried out using QoMMMa, a set of routines coupling output from QM and MM software developed by J. Harvey[9]. The *Jaguar* software package[10] was used for QM calculations and the TINKER code[11] using the CHARMM27 all-atom force field[6] for MM calculations. MM atoms more than 16 Å away from the carbon of the terminal methyl group of acetyl-CoA were held fixed. Full optimization of the geometry of the MM atoms was performed at each QM step. Electronic polarization of the QM system is accounted for by including the MM atomic partial charges in the QM Hamiltonian. The steric QM/MM interactions are described using MM van der Waals (Lennard-Jones) parameters for the QM atoms. Van der Waals parameters taken from the CHARMM27 force-field values for similar atoms were used, as they were found sufficiently accurate to describe hydrogen bonding between the QM and MM region

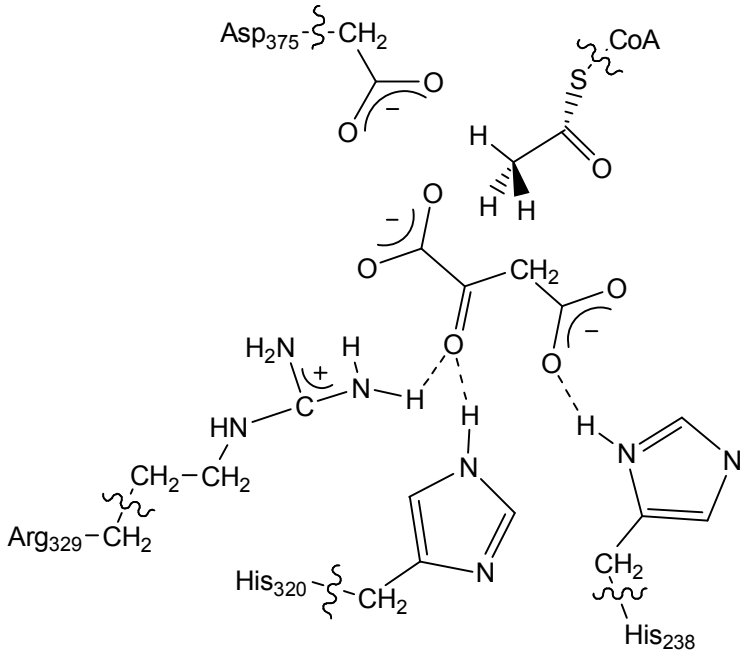
when tested in small model systems. As in the AM1/CHARMM27 QM/MM calculations, hydrogen link atoms were used to satisfy the valences of covalent bonds at the QM/MM boundary, and charges of MM atoms of groups adjacent to link atoms were set to zero in the QM calculations. No cut-off was applied for non-bonded interactions. Single point QM/MM calculations were performed on the B3LYP/6-31+G(d)/CHARMM27 optimized geometries using MP2/aug-cc-pVDZ for the QM region with the Molpro[12] program.

The starting conformation for B3LYP/6-31+G(*d*)/CHARMM27 geometry optimization was the AM1/CHARMM27 minimized conformation of the approximate transition state of the enolization step ($r1$ = 0.2 Å). From here, geometry optimization was performed backwards to the reactants and forwards to the enolate intermediate using $r1$. From the point with the lowest SCS-MP2/aug-cc-pVDZ/CHARMM27 energy for the enolate ($r1$ = 1.0 Å), $r2$ was used. At a value of $r2$ = 1.6 Å, the carbon-carbon bond is formed whereas the proton transfer from Arg329 is not complete. This conformation was optimized on B3LYP/6-31+G(*d*)/CHARMM27 level without any restraints. Thereafter, $r3$ was used to simulate the proton transfer from Arg329.

Test profiles for condensation with alternative proton donors

Introduction

The proton that has to be transferred during the condensation reaction to the (former) carbonyl carbon of OAA to arrive at the stable citryl-CoA intermediate, can come from 3 different residues (based on structural considerations, see Scheme 1). To test these possibilities, QM/MM potential energy profiles were obtained at the AM1/CHARMM27 level. The three different residues are Arg329, His320 and His238. Both histidine residues are modelled as singly protonated, as can be expected from the surrounding hydrogen bonding network and the proximity of positively charged arginine residues. Furthermore, treatment of the histidine residues as doubly protonated disrupts the active site geometry (see ref. 3 and unpublished results). To consider different levels of (a)synchronicity concerning the nucleophilic attack by the enolate onto the OAA carbonyl carbon and the proton transfer to the OAA carbonyl oxygen, different reaction coordinates were used for each alternative proton donor.

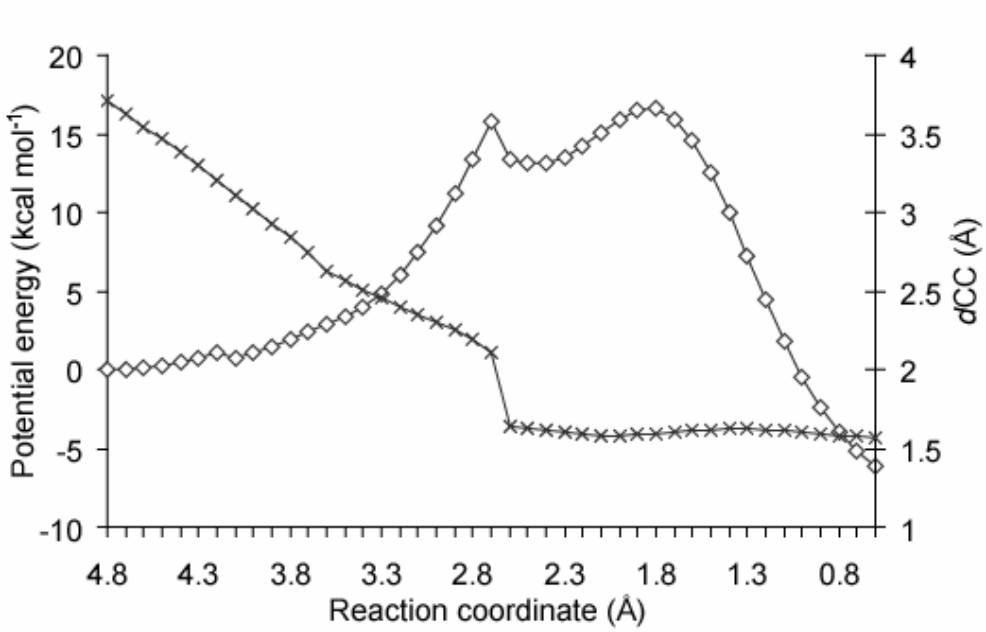


Scheme 1. Important residues in the CS active site with atom labels used for definition of reaction coordinates (*R*s) below. Wavy lines across bonds denote the location of the QM/MM barrier (link atom) when the sidechain is included in the QM region.

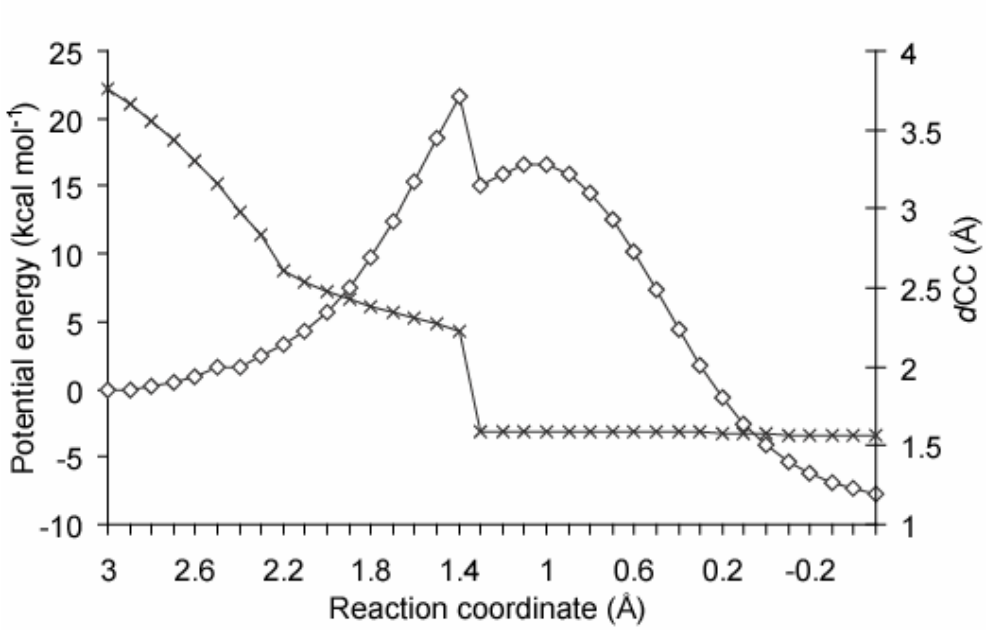
Profiles with Arg329 as proton donor

The starting conformation for the three profiles below was obtained from the potential energy profile of the enolization reaction. The QM region for that profile is identical to the one used for the condensation profiles and includes OAA, the thioester part of acetyl-CoA, sidechains of Asp375 and Asp327 from CB and the sidechain of Arg329 from CG. A harmonic restraint of 500 kcal mole^{-1} Å^{-2} was applied in each case to the reaction coordinate *R*.

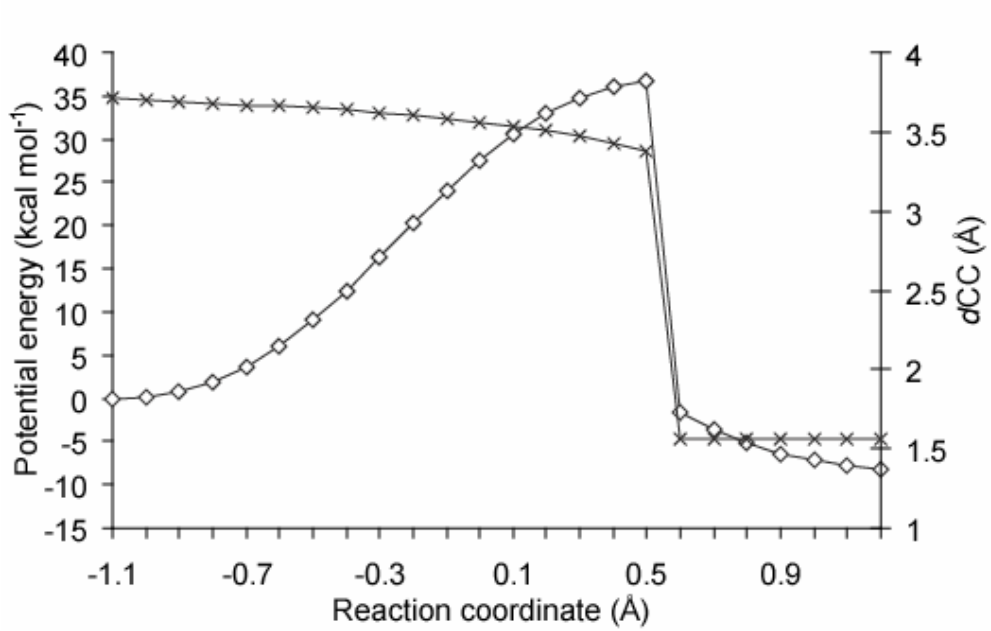
Profile 1 – $R = d(C_{acetyl\text{-}CoA}C_{OAA}) - d(N_{Arg329}H) + d(O_{OAA}H)$



Profile 2 – $R = 0.5*d(C_{acetyl\text{-}CoA}C_{OAA}) - d(N_{Arg329}H) + d(O_{OAA}H)$



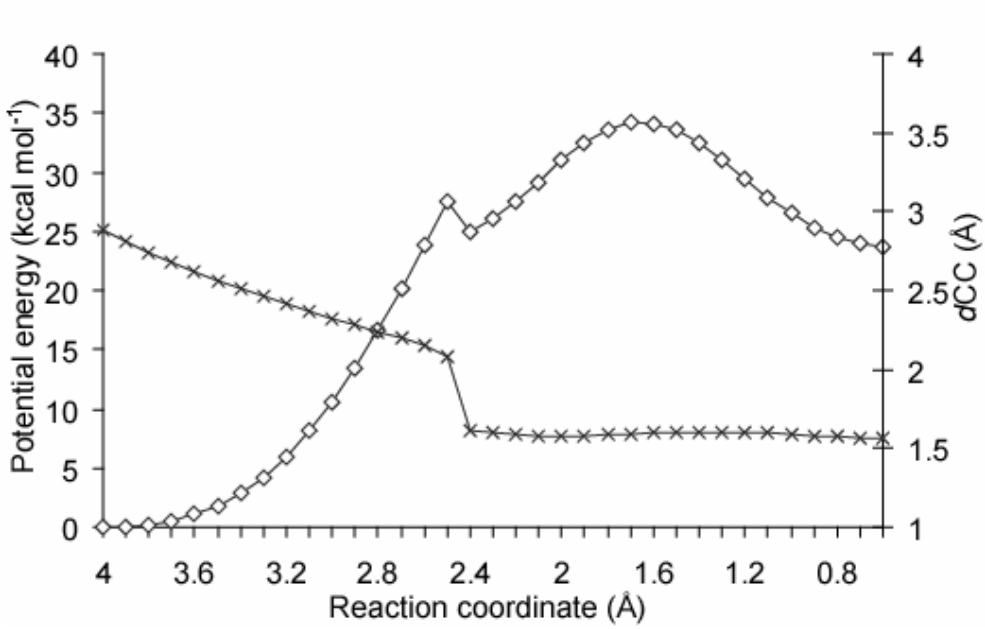
Profile 3 – $R = d(N_{Arg329}H) - d(O_{OAA}H)$



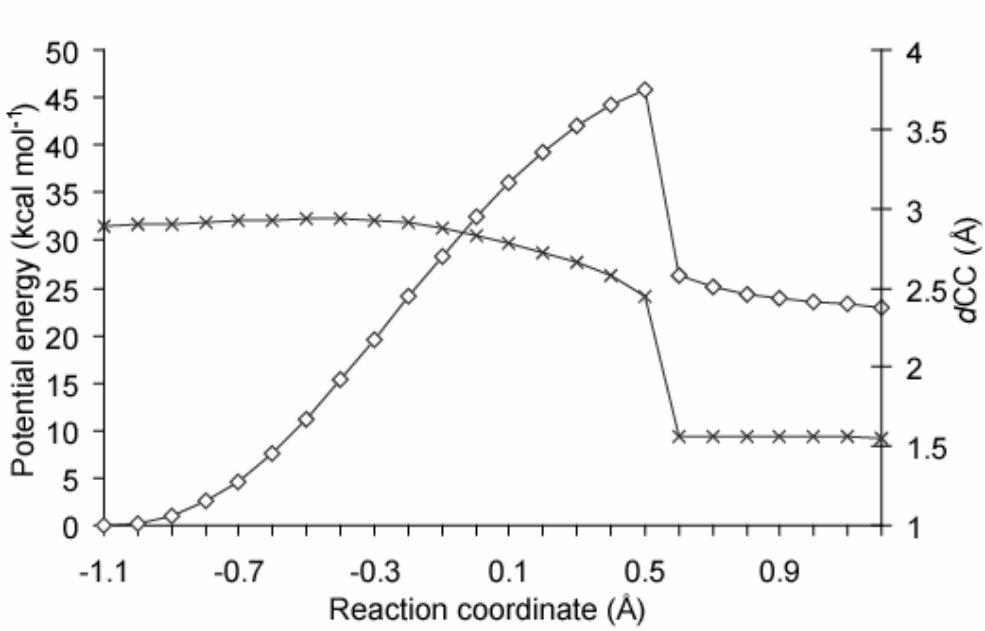
Profiles with His320 as proton donor

The starting conformation for the two profiles below was obtained after 30ps QM/MM MD and minimization applied to an enolate conformation. The QM region consists of the following: OAA, the thioester part of acetyl-CoA and the sidechains of Asp375 and His320 from CB (41 QM atoms, 3 link atoms). A harmonic restraint of 1000 kcal $mole^{-1}$ $Å^{-2}$ was applied in each case to the reaction coordinate *R*.

Profile 1 – $R = d(C_{acetyl\text{-}CoA}C_{OAA}) - d(N\delta_{His320}H) + d(O_{OAA}H)$



Profile 2 – $R = d(N\delta_{His320}H) - d(O_{OAA}H)$

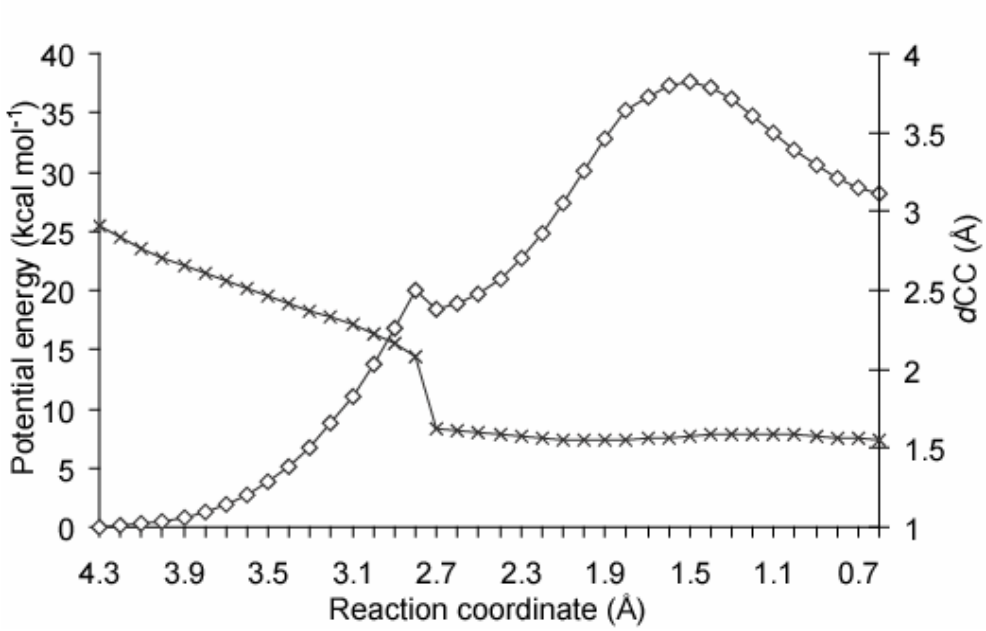


Profiles with His238 as proton donor

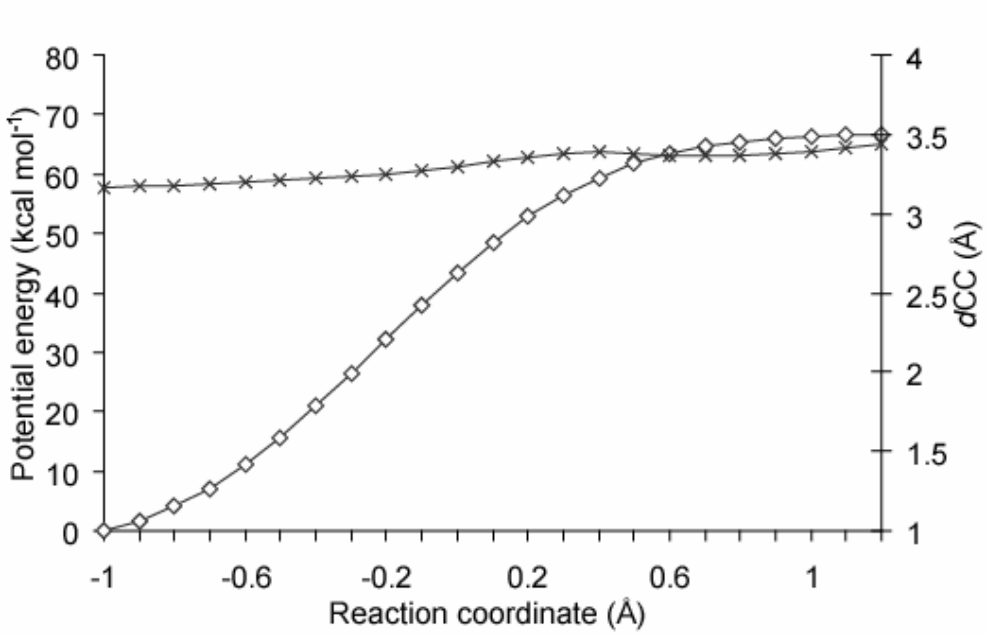
The starting conformation for the two profiles below was obtained after 60ps QM/MM MD and minimization applied to an enolate conformation. The QM region consists of the following: OAA, the thioester part of acetyl-CoA and the sidechains of

Asp375 and His238 from CB (41 QM atoms, 3 link atoms). A harmonic restraint of 1000 kcal mole^{-1} Å^{-2} was applied in each case to the reaction coordinate *R*.

Profile 1 – $R = d(C_{acetyl\text{-}CoA}C_{OAA}) - d(N\delta_{His238}H) + d(O_{OAA}H)$



Profile 2 – $R = d(N\delta_{His238}H) - d(O_{OAA}H)$



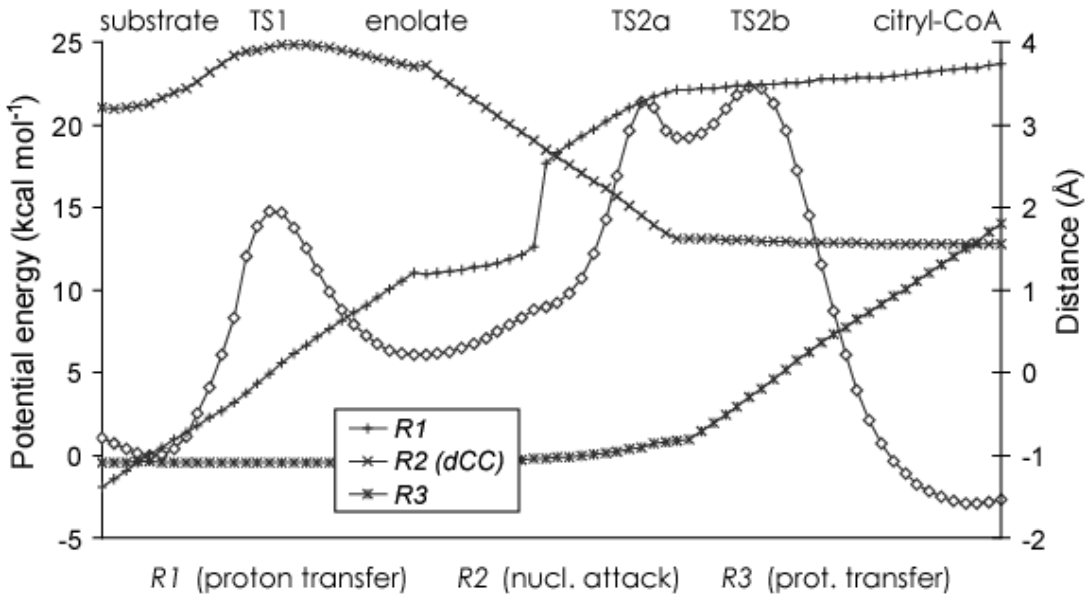
Total profile with Arg329 as proton donor at AM1/CHARMM27 level

Three reaction coordinates were sequentially applied (similar as the high-level profile shown in the main publication):

R1 = $d(C_{acetyl\text{-}CoA}H)$ - $d(O_{Asp375}H)$

R2 = $d(C_{acetyl\text{-}CoA}C_{OAA})$

R3 = $d(N_{Arg329}H)$ - $d(O_{OAA}H)$



References

1. M. W. Van Der Kamp, F. Perruccio and A. J. Mulholland, *Proteins: Struct., Funct., Bioinf.*, 2007, **69**, 521.
2. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
3. A. J. Mulholland and W. G. Richards, *Proteins: Struct., Funct., Genet.*, 1997, **27**, 9.
4. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
5. A. D. Mackerell, N. Banavali and N. Foloppe, *Biopolymers*, 2000, **56**, 257.
6. A. D. Mackerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-Mccarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586.
7. A. J. Mulholland and W. G. Richards, *THEOCHEM - J. Mol. Struc.*, 1998, **429**, 13.
8. M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700; N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720.
9. J. N. Harvey, *Faraday discussions*, 2004, **127**, 165.
10. *Jaguar 5.0*, Schrodinger, LLC, Portland, Oregon, 2002.
11. J. W. Ponder, *TINKER: Software Tools for Molecular Design, v4.0* Saint Louis, MO, 2003.
12. *MOLPRO 2002.10*, University College Cardiff Consultants Limited, Cardiff, 2004.