Regulating effect of β-ketoacyl synthase domain of fatty acid synthase on fatty acyl chain length in *de novo* fatty acid synthesis

Wei Cui[a], Yan Liang[b], Weixi Tian[c], Mingjuan Ji[a*], Xiaofeng Ma[c*]

[a]School of Chemistry and Chemical Engineering, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

[b]School of Kinesiology and Health, Capital University of Physical Education and Sports, No. 11 Beisanhuanxi Road, Beijing 100191, China.

[c]College of Life Sciences, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

E-mail:

W.C. cuiwei@ucas.ac.cn

Y.L. yanliang@cupes.edu.cn

W.T. tianweixi@ucas.ac.cn

M.J. jmj@ucas.ac.cn

X.M. maxiaofeng@ucas.ac.cn

Corresponding authors:

Mingjuan Ji

Tel.: +86 10 88256326; Fax: +86 10 88256093

Xiaofeng Ma

Tel/Fax: +86 10 88256353

Abstract

Fatty acid synthase (FAS) is a multifunctional homodimeric protein, and is the key enzyme required for the anabolic conversion of dietary carbohydrates to fatty acids. FAS synthesizes long-chain fatty acids from three substrates: acetyl-CoA as a primer, malonyl-CoA as a 2 carbon donor, and NADPH for reduction. The entire reaction is composed of numerous sequential steps, each catalyzed by a specific functional domain of the enzyme. FAS comprises seven different functional domains, among which the β-ketoacyl synthase (KS) domain carries out the key condensation reaction to elongate the length of fatty acid chain. Acyl tail length controlled fatty acid synthesis in eukaryotes is a classic example of how a chain building multienzyme works. Different hypotheses have been put forward to explain how those sub-units of FAS are orchestrated to produce fatty acids with proper molecular weight. In the present study, molecular dynamics simulation based binding free energy calculation and access tunnels analysis showed that the C16 acyl tail fatty acid, the major product of FAS, fits to the active site on KS domain better than any other substrates. These simulations supported a new hypothesis about the mechanism of fatty acid production ratio: the geometric shape of active site on KS domain might play a determinate role.

Keywords

Fatty acid synthase (FAS); β-ketoacyl synthase (KS); binding free energy calculation; fatty acid production ratio; chain length.

1. Introduction

As well known, lipids is one of the major elements in biological molecules. Lipids play an important role in living systems and diverse intracellular functions, from energy storage and plasma membranes structure to signal transduction cascades and protein acylation [1]. Fatty acids and fatty acid residues are key constituents of lipids [2]. The membrane lipids and other lipids with fatty acid residues possess hydrophobicity because of their hydrocarbon chain. The structure of fatty acids is a complex of long-chain hydrocarbon tail of various lengths terminated with carboxylic acid groups, and the length of acyl tail determined the property of fatty acid residues in lipid molecular. Fatty acid biosynthesis exists in all plants and animals. *In vivo*, fatty acid synthase (EC 2.3.1.85, FAS) catalyzes the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent condensation of acetyl-CoA and malonyl-CoA to produce the saturated 16-carbon fatty acid, palmitate [3, 4].

As one of the most complex cellular multi-homodimeric mammalian enzymes, FAS comprises two identical subunits, each 270-kilodalton polypeptide chain carries a set of seven discrete functional domains, which are malonylacetyl transferase (MAT), β-ketoacyl reductase (KR), β-ketoacyl synthase (KS), β-hydroxyacyl dehydratase (DH), enoyl reductase (ER), thioesterase (TE) and acyl carrier protein (ACP). Stepwise elongation of precursors is achieved by cyclic decarboxylative condensation of acyl-CoA with the elongation substrate malonyl-CoA, initiated by the starter substrate acetyl-CoA [5, 6]. In the priming step, the acetyl transferase loads acetyl-CoA onto the terminal thiol of the phosphopantheteine cofactor of the ACP,

which passes the acetyl moiety over to the cysteine in the active site of KS. MAT transfers the malonyl group from malonyl-CoA to ACP, and KS catalyzes the decarboxylative condensation of the acetyl and malonylmoieties to an ACP-bound β-ketoacyl intermediate. The b-carbon position is then modified by sequential action of the NADPH–dependent KR, DH, and ER to yield a saturated acyl product elongated by two carbon units. This acyl group functions as a starter substrate for the next round of elongation, until the growing fatty acid chain reaches to a length of 16 to 18 carbon atoms and is released from ACP as free fatty acids by TE domain [7].

This entire reaction is composed of a series of sequential reactions where acyl intermediates are each catalyzed by a specific enzyme domain. Interestingly, the main product of FAS catalytic reaction is C16 fatty acid. In fact, a significant amount of the fatty acids produced by FAS, as well as fatty acids taken up from the diet, are further elongated into long chain fatty acids containing 18 carbon atoms or longer, i.e., very long chain fatty acids (VLCFA) [8-10]. Obviously the catalytic function of FAS has strong product specificity.

Among the seven moieties of FAS, the KS domain catalyzes the condensation reaction, which determines the length of the carbon chain. The condensation reaction is a two-step process, first the acyl component of an activated acyl primer is transferred to a cysteine residue of the enzyme and then condensed with an activated malonyl donor with the concomitant release of carbon dioxide [11-13].

Although TE was believed to be involved in the chain-terminating step of fatty acid synthesis [14], the precise mechanisms of the specificity of FAS catalyzed reaction remain unclear. In the present study, we offered a new possible mechanism of regulating the fatty acyl chain length of FAS based on our calculating.

2. Materials and Methods

2.1 Preparation of the starting conformation

The coordinates of the crystal structure of KS domain were retrieved from Protein Data Bank (PDB entry: 3HHD [15]). The substrate structure was shown in Scheme 1 and the coordinates were optimized by Gaussian 03 [16] at the HF/6-311G* level, and then the atomic partial charges were obtained by fitting the electrostatic potentials using the RESP fitting technique [17] in AMBERTools12. The generations of the partial charges and the force field parameters for the substrates were accomplished using the antechamber program in AMBERTools12 [18]. Then, the molecular docking software AutoDock 4.2 [19] was performed to tackle the binding models of the growing saturated fatty acyl-ACP inside the ketoacyl synthesis cavity of KS domain. The Lamarckian genetic algorithm [20] (LGA) was applied to optimize the binding conformations of acyl-fatty acid in the active site. The docking calculation was performed in a three dimensional grid with 60*60*60 points and a spacing of 0.375 Å. The docking grid located in the ketoacyl synthesis cavity. There were 256 runs of LGA optimization for fatty acid. For each run the numbers of generations and the energy evaluations were set to 27000 and 500000, respectively. After 256 runs of docking calculation, 256 conformations were generated, and then these conformations

were clustered by the root mean squared deviations (RMSDs) of heavy atoms. The most popular cluster with the lowest energy conformation was selected as the initial structure to perform the molecular mechanics (MM) optimization and molecular dynamics (MD) simulation.

In the MM optimizations and MD simulations, AMBER03 (parm03) force field [21] and general amber force field(gaff) [22] were used to establish the potentials of protein and substrates in binding model, respectively. As the simulations were performed at pH 7.0, titratable residues, including lysine, arginine, aspartic acid, and glutamic acid, were typically charged, and the histidine residues were protonated at the epsilon position. The whole system was immersed with TIP3P water molecules [23] in a box of 12 Å from any solute atoms, and was neutralized with the counterions of Na^+.

2.2 Molecular dynamics simulation

Prior to molecular dynamic sampling, the initial structures generated by AutoDock were minimized by the following three-stage protocol using the sander program in AMBER12 [24]. First, the protein was restrained, and solvent molecules and ions were relaxed (2000 cycles of steepest descent and 2000 cycles of conjugate gradient minimizations). Second, the protein backbone was restrained, and the side chains and substrate were minimized (2500 cycles of steepest descent and 2500 cycles of conjugate gradient minimizations). Finally, the whole system was minimized without

any restrain (2500 cycles of steepest descent and 2500 cycles of conjugate gradient minimizations).

In the following MD simulations, the relaxed structures were gradually heated from 0 to 310 K over 60 ps. Then, the systems were simulated by 20 ns MD in the canonical ensemble in a constant temperature of 310 K by the weak coupling algorithm using the pmemd program in AMBER12 [25]. SHAKE [26] was used to constrain bonds involving hydrogen atoms, and the time step was 2.0 fs. The particle mesh Ewald (PME) algorithm [27] was employed to calculate the long-range electrostatics. The nonbonded cutoff was set to 10 Å. During the whole sampling process, the coordinates were saved every 0.2 ps.

2.3 Binding free energy prediction

The binding free energies were calculated by the MM/PBSA technique according to the following equation [28-37].

$$\begin{aligned}\Delta G_{binding} &= G_{complex} - G_{protein} - G_{ligand} \\ &= \Delta H + \Delta G_{solvation} - T\Delta S \\ &= \Delta E_{MM} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S\end{aligned}$$

Where ΔE_{MM} is the molecular mechanics interaction energy between the protein and the inhibitor; ΔG_{PB} and ΔG_{SA} are the electrostatic and nonpolar contributions to desolvation upon inhibitor binding, respectively; and $-T\Delta S$ is the conformational entropy change, which is not take into consideration for the high computational cost and low prediction accuracy.

Here, the polar part of desolvation was calculated by solving the Poisson-Boltzmann (PB) equations. The grid size used to solve the PB equation was 0.5 Å, and the values of interior dielectric constant and exterior dielectric constant were set to 1 and 80, respectively [30]. The nonpolar desolvation term ΔG_{SA} was estimated from the solvent accessible surface area (SASA): G_{np} = 0.0072 × SASA + 0.00.area [38]. The protein-substrate binding free energy was calculated from the 200 snapshots taken from the last 2.0 ns MD simulation trajectories of the complex. The calculations for binding free energies were accomplished by using the mm pbsa program in AMBER12 [39-44].

2.4 Analysis of access tunnels

CAVER is a tool widely used for the identification and characterization of transport cavities in macromolecular structures [45]. Herein, the CAVER 3.0 [45-47] was applied for the analysis of molecular dynamics trajectory of the most stable complex to find the possible access tunnels, which thread from the surface to the active site of KS domain. 200 snapshots extracted from the last 2.0 ns trajectory of the most stable complex were analyzed. The snapshots for access tunnel analysis were totally the same as MM/PBSA calculation. The SG atom of Cys-161 was chosen as the starting atom of the tunnel searching. The probe radius was set to 1.0 Å to identify the tunnels with the bottleneck (minimum) radius over 1.0 Å, and the clustering threshold were set to 5, which meant the tree hierarchy of tunnel clusters were cut at the value of 5. Default settings for other parameters were used throughout the calculations. The

maximum number of tunnel clusters reported was set to 999. Then the tunnels were visualized using PyMOL [48].

3. Results and Discussions

3.1 Binding free energy profile of KS-substrate complex

AutoDock 4.2 [19] was performed to tackle the binding models of the growing saturated fatty acyl-ACP inside the ketoacyl synthesis cavity of KS domain. The results of AutoDock were shown in Figure S1 and Figure S2, and the binding models with the lowest predicted free energy in each most populated cluster were set as initial structures to perform molecular dynamics simulation by using AMBER12 [24]. The time evolution of potential energy and root mean square deviation of backbone atomic positions in MD simulation trajectory was shown in Figure S3 and Figure S4 respectively. Then, in order to obtain the binding free energy profile of the ACP carried fatty acid during the elongation of aliphatic tail, MM/PBSA was used to calculate the binding free energy with the snapshots extracted from the MD trajectory [28]. The results of MM/PBSA binding free energy prediction are shown in Figure 1, for the short tail carboxylic acids, the binding free energy of ACP carried fatty acid and FAS-KS domain complexes decreased with the elongation of aliphatic tail of fatty acid. The C16 group, the precursor compound of octadecanoic acid was the most stable complex, with a binding free energy of -50.34 kcal·mol^{-1}. On the other hand, the complexes with higher molecular weight showed a reverse trend, and the binding free energies of C18 and C20 groups were -44.30 kcal·mol^{-1} and -42.93 kcal·mol^{-1}

respectively, both higher than the most stable one. This simulation result tallied with experiment. The C16 and C18 fatty acid are the main products of FAS. The lower the binding free energy, the more stable the complex, which composed of ACP carried growing fatty acid chain and KS domain. And it also meant a higher yield of fatty acid with the corresponding length of aliphatic tail. Consequently, it is reasonable to explain the product distribution of FAS by the stability assay of ACP carried fatty acid and KS domain complex.

The correlations between the predicted binding free energies and each energy term were compared to confirm the key energy term that determines the difference of the binding affinities between different fatty acid and KS domain. We found that both the nonpolar contribution ($\Delta E_{vdw} + \Delta G_{SA}$) and the polar contribution ($\Delta E_{ele} + \Delta G_{PB}$) had linear correlation (r=0.69) with the binding free energy, and the correlation coefficient of nonpolar contribution and total energy was significant higher than that of the polar contribution and total energy, which were 0.93 and 0.69 respectively. This result suggested that the nonpolar contribution, especially the van der Waals term, determined the difference of the binding affinities of different acyl tail length fatty acids and KS domain. Therefore, according to the correlation analysis, it inferred that the nonpolar interactions ($\Delta E_{vdw} + \Delta G_{SA}$) played important roles in determing the molecular weight of FAS products.

3.2 Analysis of access tunnels of KS domain

Binding free energy calculation uncovered that the nonpolar energy term was a determinant of fatty acids acyl tail length distribution. As well known, the van der Waals interaction reflects the distance between substrate atoms and enzyme atoms, thus, it is helpful to analyze the route of fatty acid through protein surface to reach the active site for reaction. CAVER [45], a widely adopted software, was applied to analyze the macro-molecular transport cavities. 200 snapshots were extracted from the last 4ns MD simulation to analyze the characteristics of possible fatty acid transport tunnels in KS domain active site. The tunnels identified from each snapshot were clustered based on the pairwise distances of the tunnels, and 9 tunnel clusters were got finally. Herein, the occurrence frequency of a tunnel cluster was defined as occupancy of snapshots in which at least one cavity with bottleneck radius over 1.0 Å accounted entire snapshot, and six tunnel clusters with occurrence frequencies higher than 3% were listed in Figure 2 and Table 1. As shown in Table 1 the occurrence frequencies and mean throughputs of tunnel green and tunnel blue were significantly higher than those of the other 4 tunnel clusters, which implicated that tunnel green and tunnel blue had were much more possible to be KS domain active site tunnels. Moreover, the data in Table 1 indicated that the active pocket of KS domain was very narrow and deep. Once the substrate complex with lost almost all of its translational degree of freedom and rotational degree of freedom, and most of the vibration degrees of freedom would lost either, which meant that the complex between substrate and KS domain was a stable and rigid structure.

3.3 Align analysis of fatty acid and possible fatty acid cavity

Aligned the tunnel analysis result and molecular docking result could further ensured that which fatty acid cavity was the most possible one. The molecular align result was shown in Figure 3. It was clear that the fatty acid was located in the tunnel green accurately, thus the tunnel green was the major cavity for fatty acid.

As shown in Figure 3A, the short-tail fatty acid (C4 group) and long-tail fatty acid (C16 group) complex were aligned in tunnel green. It informed that the C16 fatty acid occupied almost all the space inside tunnel green, while the C4 fatty acid only filled the entrance of the cavity. Such difference led to diverse atom connections between fatty acid and KS domain, and eventually made different the Lennard-Jones potential contribution of fatty acid binding free energy. On the contrary, as shown in Figure 3B, both the C16 and C20 fatty acid tails were packing the whole cavity, but C20 left a longer acyl tail outside the active cave than that of C16 group. As known, the acyl chain of fatty acid is a hydrophobic structure, therefore, the C16 complex is more stable than C20 complex.

3.4 Tunnel-lining residues analysis

According to Figure 4A, and 4B, FAS carried seven discrete functional domains, and the KS domain had only one active site in 161, a cysteine, which could covalent binding with carboxyl of fatty acid [49, 50]. In Figure 4B, the binding pocket around Cys-161 was deep and narrow, so the substrates complexed with KS domain showed much less conformations than in solvent. Such characteristical structure reduced the difficulty of KS domain-substrate complex structure prediction, as there were much

less candidate complex conformation, which meant higher credibility of molecular simulation study. On the other hand, the deep and narrow binding shape meant that the structure and properties of binding pocket should determined by the bottleneck residues. So it was very important to identify the bottleneck residues. CAVER 3.0 could statistically show the number of snapshots in which the residues lined each tunnel [46]. According to these results, a detailed picture of fatty acid access channels could be drawn, and the key residues of KS domain active site could be identified. The tunnel-lining residues analysis of the most possible fatty acid cavity was shown in Figure 5. There are totally 43 residues lined in tunnel green, the most possible fatty acid access channel. Among those 43 residues, the highest occupation rate was 55%, the residues with the highest occupation rate were shown in Figure 5.

The bottleneck residues of tunnel green in each snapshot during MD simulation were analyzed by CAVER3.0 too, and the result was shown in Table 2. There were 9 residues with bottleneck occupation rate higher than 20%, and all were the highest occupation rate residues. According to Figure 4C and Table 2, most of the highest occupation rate bottle neck residues were hydrophobic amino acid or aromatic amino acids, such as F263, P264, H292, F393, and F395. Those hydrophobic bottle residues gathering in one side of wall of tunnel green, became a hydrophobic bottleneck in the narrow and deep cave. It was reasonable to deduce that such a hydrophobic bottleneck was friendly to the fatty acid substrate entering the active site of KS domain. Moreover, as mentioned above, those residues determined the narrowest part of fatty acid cavity. Any mutation of those residues might lead to quite different van der

Walls interactions between substrate and FAS, and finally change the molecular weight of the main product.

Among seven functional domains of FAS, TE carries out the chain-terminating step of fatty acid synthesis. TE was believed to play a key role in regulating the carbon chain length of the major product of FAS. Using various acyl-CoA derivatives as model substrates, the specificity of rat TE (obtained from rat liver and mammary gland) was studied and the results showed that TE domain hydrolyzes long chain, in preference to short chain, thioesters of CoA, in fatty acid synthesis [51-53]. The structure analysis of a 2.6-Å resolution crystal structure of human FAS TE domain also provided evidence that TE may contribute to the specificity of FAS toward carbon chain length [7, 13]. However, the substrate selectivity of KS domain has never been studied before. In the present study, based on above docking and calculating results, we offered a new possible mechanism of how FAS regulating the fatty acyl chain length. FAS is a multifunctional enzyme complex. In order to complete the catalytic reaction more quickly and efficiently, it is reasonable that there are multiple domains which having substrate selectivity take part in the whole synthesis process.

4. Conclusions

Acyl tail length controlled fatty acid synthesis in eukaryotes is a classic model of how a chain building multienzyme works. Biochemical studies have demonstrated the way in which the sub-units of FAS are orchestrated to produce fatty acids with proper molecular weight. Based on these analyses and our molecular docking, the hypothesis

of fatty acid production ratio is offered: the geometric shape of active site on KS domain might be one of the determinants for production ratios of FAS.

To prove the hypothesis, the binding free energy between KS domain and fatty acid substrates with different molecular weight were estimated by MD simulation and MM/PBSA calculation. The results of molecular simulation compared with other substrates revealed that the C16 acyl tail was more stable with a higher correlation coefficient to van der Walls interaction than polar energy contribution with total binding free energy. Compared with the Coulomb potential, the Lennard-Jones potential was more sensitive to the shape of active site and substrate.

In conclusion, it is reasonable to explain the product ratio of FAS by shape of active site of KS domain. The length and bottleneck radius of fatty acid cavity in KS domain makes the C16 acyl tail perfectly fit to active site, so that it could form the atomic connections between substrate and KS domain as much as possible, which make the complex the most stable one. For the substrates shorter than C16 groups, the cavity could not be filled by the fatty acid, so the higher Lennard-Jones potential leads to an unstable complex. On the other hand, for the substrates over C16 group, the chain is longer than the tunnel, resulting in a part of the hydrophobic tail exposure in the polar solvent, which will lead to a more unstable complex and eventually decrease the yield of longer fatty acids. To sum up, it is reasonable to assume that the active site of KS domain plays a key role in length determination of fatty acyl tail. Only the fatty acid

with a proper tail length could form the most stable complex with the enzyme and become the main product of FAS.

Competing interests

All authors declare that there is no conflict of interest relevant to this study.

Acknowledgments

This work was supported by NNSF of China (Projects 21173264, 21403229); Fusion of Science and Education Special Fund, College of Life Sciences, University of Chinese Academy of Sciences (KJRH2015-012); Application Basic Research Project of Qinghai Province (2015-ZJ-728); Youth Innovation Promotion Association, CAS; 2014 Youth National Natural Science Foundation of China (No. 31300292); The Key Program of "The Dawn of West China" Talent Foundation of CAS (2012); and the molecular simulation study was supported by Scientific Computing Center of University of Chinese Academy of Sciences.

References

[1] S. Subramaniam, E. Fahy, S. Gupta, M. Sud, R.W. Byrnes, D. Cotter, A.R. Dinasarapu, M.R. Maurya, Bioinformatics and systems biology of the lipidome, Chem. Rev., 111 (2011) 6452-6490.

[2] E. Fahy, S. Subramaniam, R.C. Murphy, M. Nishijima, C.R. Raetz, T. Shimizu, F. Spener, G. van Meer, M.J. Wakelam, E.A. Dennis, Update of the LIPID MAPS comprehensive classification system for lipids, J. Lipid Res., 50 (2009) S9-S14.

[3] S.J. Wakil, Fatty-Acid Synthase, a Proficient Multifunctional Enzyme, Biochemistry-Us, 28 (1989) 4523-4530.

[4] S. Smith, The animal fatty acid synthase: one gene, one polypeptide, seven enzymes, FASEB J., 8 (1994) 1248-1259.

[5] S.J. Wakil, Mechanism of fatty acid synthesis, J. Lipid Res., 2 (1961) 1-24.

[6] S.J. Wakil, J.K. Stoops, V.C. Joshi, Fatty acid synthesis and its regulation, Annu. Rev. Biochem., 52 (1983) 537-579.

[7] B. Chakravarty, Z.W. Gu, S.S. Chirala, S.J. Wakil, F.A. Quiocho, Human fatty acid synthase: Structure and substrate selectivity of the thioesterase domain, P. Natl. Acad. Sci. USA, 101 (2004) 15567-15572.

[8] A.E. Leonard, S.L. Pereira, H. Sprecher, Y.-S. Huang, Elongation of long-chain fatty acids, Prog. Lipid Res., 43 (2004) 36-54.

[9] A. Jakobsson, R. Westerberg, A. Jacobsson, Fatty acid elongases in mammals: Their regulation and roles in metabolism, Prog. Lipid Res., 45 (2006) 237-249.

[10] M. Ikeda, Y. Kanao, M. Yamanaka, H. Sakuraba, Y. Mizutani, Y. Igarashi, A. Kihara, Characterization of four mammalian 3-hydroxyacyl-CoA dehydratases involved in very long-chain fatty acid synthesis, FEBS Lett., 582 (2008) 2435-2440.

[11] T. Maier, S. Jenni, N. Ban, Architecture of mammalian fatty acid synthase at 4.5 angstrom resolution, Science, 311 (2006) 1258-1262.

[12] T. Maier, M. Leibundgut, N. Ban, The crystal structure of a mammalian fatty acid synthase, Science, 321 (2008) 1315-1322.

[13] T. Maier, M. Leibundgut, D. Boehringer, N. Ban, Structure and function of eukaryotic fatty acid synthases, Q. Rev. Biophys., 43 (2010) 373-422.

[14] S. Smith, E. Agradi, L. Libertini, K. Dileepan, Specific release of the thioesterase component of the fatty acid synthetase multienzyme complex by limited trypsinization, Proc. Natl. Acad. Sci. USA, 73 (1976) 1184-1188.

[15] G. Pappenberger, J. Benz, B. Gsell, M. Hennig, A. Ruf, M. Stihle, R. Thoma, M.G. Rudolph, Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design, J. Mol. Biol., 397 (2010) 508-519.

[16] E. Frisch, M. Frisch, G.W. Trucks, Gaussian 03, Gaussian, 2003.

[17] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, J. Phys. Chem., 97 (1993) 10269-10280.

[18] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J. Mol. Graphics Modell., 25 (2006) 247-260.

[19] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem., 30 (2009) 2785-2791.

[20] J. Fuhrmann, A. Rurainski, H.P. Lenhof, D. Neumann, A new Lamarckian genetic algorithm for flexible ligand-receptor docking, J. Comput. Chem., 31 (2010) 1911-1918.

[21] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, A point-harge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem., 24 (2003) 1999-2012.

[22] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem., 25 (2004) 1157-1174.

[23] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, J. Chem. Phys., 79 (1983) 926-935.

[24] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem., 26 (2005) 1668-1688.

[25] S.C. Harvey, R.K.Z. Tan, T.E. Cheatham III, The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition, J. Comput. Chem., 19 (1998) 726-740.

[26] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys., 23 (1977) 327-341.

[27] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, J. Chem. Phys., 98 (1993) 10089-10092.

[28] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S.H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E.

Cheatham, Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, Accounts Chem. Res., 33 (2000) 889-897.

[29] J.M. Wang, T.J. Hou, X.J. Xu, Recent advances in free energy calculations with a combination of molecular mechanics and continuum models, Curr. ComputAid. Drug., 2 (2006) 287-306.

[30] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, J. Chem. Inf. Model., 51 (2011) 69-82.

[31] T.J. Hou, J.M. Wang, Y.Y. Li, W. Wang, Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking, J. Comput. Chem., 32 (2011) 866-877.

[32] T. Hou, R. Yu, Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance, J. Med. Chem., 50 (2007) 1177-1188.

[33] H. Gohlke, D.A. Case, Converging free energy estimates: MM-PB (GB) SA studies on the protein-protein complex Ras-Raf, J. Comput. Chem., 25 (2004) 238-250.

[34] S. Genheden, U. Ryde, Comparison of the Efficiency of the LIE and MM/GBSA Methods to Calculate Ligand-Binding Energies, J. Chem. Theory Comput., 7 (2011) 3768-3778.

[35] N. Homeyer, H. Gohlke, Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method, Mol. Inf., 31 (2012) 114-122.

[36] L. Xu, H. Sun, Y. Li, J. Wang, T. Hou, Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models, J. Phys. Chem.. B, 117 (2013) 8408-8421.

[37] H. Sun, Y. Li, S. Tian, L. Xu, T. Hou, Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set, Phys. Chem. Chem. Phys., 16 (2014) 16719-16729.

[38] J. Weiser, P.S. Shenkin, W.C. Still, Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO), J. Comput. Chem., 20 (1999) 217-230.

[39] H. Gohlke, C. Kiel, D.A. Case, Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes, J. Mol. Biol., 330 (2003) 891-913.

[40] T.J. Hou, W. Zhang, D.A. Case, W. Wang, Characterization of domain-peptide interaction interface: A case study on the amphiphysin-1 SH3 domain, J. Mol. Biol., 376 (2008) 1201-1214.

[41] T.J. Hou, Z. Xu, W. Zhang, W.A. McLaughlin, D.A. Case, Y. Xu, W. Wang, Characterization of Domain-Peptide Interaction Interface: a generic structure-based

model to decipher the binding specificity of SH3 domains, Mol. Cell Proteomics, 8 (2009) 639-649.

[42] T.J. Hou, N. Li, Y.Y. Li, W. Wang, Characterization of Domain-Peptide Interaction Interface: Prediction of SH3 Domain-Mediated Protein-Protein Interaction Network in Yeast by Generic Structure-Based Models, J. Proteome Res., 11 (2012) 2982-2995.

[43] Z. Xu, T. Hou, N. Li, Y. Xu, W. Wang, Proteome-wide Detection of Abl1 SH3-binding Peptides by Integrating Computational Prediction and Peptide Microarray, Mol. Cell Proteomics, 11 (2012) O111.010389.

[44] H. Sun, Y. Li, S. Tian, J. Wang, T. Hou, P-loop Conformation Governed Crizotinib Resistance in G2032R-Mutated ROS1 Tyrosine Kinase: Clues from Free Energy Landscape, PLoS Comput. Biol., 10 (2014) e1003729.

[45] M. Petrek, M. Otyepka, P. Banas, P. Kosinova, J. Koca, J. Damborsky, CAVER: a new tool to explore routes from protein clefts, pockets and cavities, BMC Bioinformatics, 7 (2006) 316.

[46] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, J. Damborsky, CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures, PLoS Comput. Biol., 8 (2012) e1002708.

[47] B. Kozlikova, E. Sebestova, V. Sustr, J. Brezovsky, O. Strnad, L. Daniel, D. Bednar, A. Pavelka, M. Manak, M. Bezdeka, P. Benes, M. Kotry, A. Gora, J.

Damborsky, J. Sochor, CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures, Bioinformatics, (2014).

[48] W.L. DeLano, The PyMOL molecular graphics system, (2002).

[49] A. Witkowski, A.K. Joshi, Y. Lindqvist, S. Smith, Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine, Biochemistry, 38 (1999): 11643-11650.

[50] A. Witkowski, A.K. Joshi, V.S. Rangan, A.M. Falick, H.E. Witkowska, S. Smith, Dibromopropanone cross-linking of the phosphopantetheine and active-site cysteine thiols of the animal fatty acid synthase can occur both inter- and intrasubunit. Reevaluation of the side-by-side, antiparallel subunit model, J Biol Chem., 274 (1999): 11557-11563.

[51] M. Pazirandeh, S.S. Chirala, W.Y. Huang, S.J. Wakil, Characterization of recombinant thioesterase and acyl carrier protein domains of chicken fatty acid synthase expressed in Escherichia coli, J Biol Chem, 264 (1989) 18195-18201.

[52] C.Y. Lin, S. Smith, Properties of the thioesterase component obtained by limited trypsinization of the fatty acid synthetase multienzyme complex, J Biol Chem, 253 (1978) 1954-1962.

[53] L.J. Libertini, S. Smith, Purification and properties of a thioesterase from lactating rat mammary gland which modifies the product specificity of fatty acid synthetase, J Biol Chem, 253 (1978) 1393-1401.

Figure captions

Figure 1. Change of nonpolar contribution, polar contribution and total binding free energy of different length aliphatic tail fatty acid and KS domain complex. According to the total binding free energy data, the C16 group is the most stable complex compared with other fatty acids. Compared with polar energy term, the nonpolar contribution is better correlated with total binding free energy, implicating that the nonpolar contribution of binding free energy plays an important role in the fatty acid aliphatic tail length fitting of KS domain.

Figure 2. Six top ranked collective potential fatty acid cavities identified throughout the molecular dynamics simulations with occurrence over 4% are all depicted in one frame as cavity surfaces. Cavities following the ranking order are shown as green, blue, yellow, orange, red and violet, respectively. The structures are shown in cartoon representation and covered by the molecular surface. According to the results of CAVER 3.0 program calculation, the tunnel green and tunnel blue are the most possible fatty acid cavities as the occurrence of those two tunnels are much higher than any other cavities.

Figure 3. The aligned fatty acids and tunnel green. (A) the C4 group (the red stick model) and C16 group (the blue stick model) are aligned in tunnel green surface (the green mesh model). As shown in the figure, the green cavity is filled by the C16 group fatty acid, while the C4 group only filled half of the active cave; (B) the C16

group (the blue stick model) and C20 group (the red stick model) are aligned in tunnel green surface (the green mesh model). As shown in the figure, both the C16 and C20 group fatty acids fill the tunnel green, but some part of aliphatic tail of C20 fatty acid is outside the tunnel green surface, which implicated that the length of C16 fatty acid tail is closest to the mean cavity length of tunnel green. Those comparisons show that the C16 fatty acid best fits to the shape of tunnel green, and the atomic connections between C16 fatty acid and KS domain will lead to the best nonpolar binding free energy contribution.

Figure 4. The active site of KS domain. According to the molecular dynamics simulation and binding free energy calculation, the active site of KS domain may regulate the molecular weight distribution of production in *de novo* fatty acid synthesis. (A) The FAS carries seven discrete functional domains. Here in, KS domain was covered by pale green surface, and the active site surface of KS domain was in more bright green than other residues of KS domain. (B) The crystal structure simulated contains two functional domain, KS domain and MAT domain. Shape of the crystal structure looks like a Dumbbell, KS domain and MAT domain are in two side of the dumbbell. Here in the active site of KS domain, which is a deep and narrow cave on KS domain surface, was covered by a bright green molecular surface. (C) There are 9 highest occupation rate bottle neck residues in KS domain active site, those key residues were shown in stick model while other residues were shown in cartoon model.

Figure 5. The residues lined tunnel green in KS domain. CAVER 3.0 statistically showed the occupation rate of residues lined tunnel green in MD trajectory of C16 group. Those analyses were helpful for the identification of the key residues of KS domain active site. There were totally 43 residues lined tunnel green, the most possible fatty acid access channel. Among those 43 residues, the highest occupation rate was 55%, which meant in 55% of MD trajectory snapshots, those residues lined the most possible tunnel of fatty acid.

Availability of supporting data

Part S1. The detail of partial charges calculation for the substrates.

Scheme S1. The 2D structure and partial charges of C4 substrate analog.

Figure S1. The Result of Clustering analysis of the Predicted Binding Model Generated by Autodock.

Figure S2. The Binding Model with Lowest Binding Energy in the Most Populated Cluster.

Figure S3. The Total Potential Energy of each MD simulation trajectory.

Figure S4. Root Mean Square Deviation of Backbone Atomic Positions of each MD simulation trajectory.

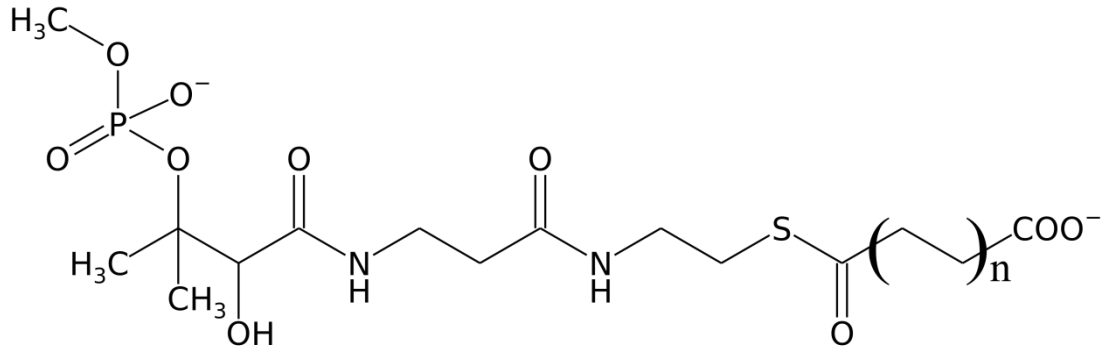
Table 1. The Characteristics of the Top Ranked Fatty acid tunnel of KS domain

	occurrence	max. bottleneck radius [Å]	mean bottleneck radiusb [Å]	mean cavity length [Å]	mean throughput
Tunnel Green	55%	1.42	1.25	14.07	0.75
Tunnel Blue	51%	1.41	1.17	17.60	0.60
Tunnel Yellow	7%	1.14	1.05	14.98	0.47
Tunnel Orange	5%	1.13	1.04	16.97	0.42
Tunnel Red	4.5%	1.15	1.07	16.56	0.43
Tunnel Violet	4%	1.10	1.04	20.11	0.39

Table 2. Residues lined tunnel green with high bottleneck occupation rate

Residues	Lined snapshots	Lined occupation rate	Bottleneck snapshots	bottleneck occupation rate
294GLY	108	54.0%	110	55.0%
304GLU	108	54.0%	110	55.0%
393PHE	108	54.0%	110	55.0%
391ASN	107	53.5%	109	54.5%
301ASP	106	53.0%	110	55.0%
292ALA	92	46.0%	104	52.0%
300GLY	82	41.0%	110	55.0%
394GLY	53	26.5%	110	55.0%
295THR	46	23.0%	110	55.0%

Scheme 1. The 2D structure of substrate analog of KS domain active site



Supplementary Materials

Part S1. The detail of partial charges calculation for the substrates

1.1 Method of partial charges

Here in, we used RESP method to calculate the partial charge for the substrate analogues. RESP is abbreviation of Restrained Electrostatic Potential [1].

First, a serial of first-principles calculation were performed to optimize the conformation of fatty acids, and the electrostatic potential grid around the molecules was also generated by the same calculation. Here in the Ab initio calculations for the lipids consisted of geometry optimisations of the substrates structures at the HF/6-31G(d) level of theory and basis set. All the quantum chemistry calculations were carried out using the Gaussian 03 software [2].

Then, the electrostatic potential grid around the molecule was restrained to the nearest atom, and the intermolecular interaction properties of molecules could be replaced by a group of atom centered point charges. The restrain calculations were carried out using the Antechamber software which was the force field related program in AMBER12 [3].

1.2 Results

Here in the C4 substrate analog was selected to show the result of RESP partial charge. To simplify the scheme, the atom centered point charges of nonpolar hydrogens, such as hydrogens in CH_3 and CH_2, were restrained to the center carbon atom. In Scheme S1, all the CH_2 groups in acyl tail in fatty acid substrate are neutral

groups, and the carboxyl tail and phosphate ester head groups are carried much more charge than other groups in substrate analog. According to the result of RESP partial charges calculation, all of the 9 fatty acid substrate analogs showed the same charge distribution model as C4 substrate analog.

References and Notes

[1] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, The Journal of Physical Chemistry, 97 (1993) 10269-10280.

[2] E. Frisch, M. Frisch, G.W. Trucks, Gaussian 03, Gaussian, 2003.

[3] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, Journal of molecular Graphics and Modelling, 25 (2006) 247-260.

Scheme S1. The 2D structure and partial charges of C4 substrate analog.