(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 September 2020 (17.09.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/182615 A1

(51) International Patent Classification:
C12N 9/88 (2006.01) *C30B 29/58* (2006.01)

(21) International Application Number:
PCT/EP2020/055859

(22) International Filing Date:
05 March 2020 (05.03.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
19161520.2 08 March 2019 (08.03.2019) EP

(71) Applicants: **VIB VZW** [BE/BE]; Rijvisschestraat 120, 9052 Gent (BE). **UNIVERSITEIT GENT** [BE/BE]; Sint-Pietersnieuwstraat 25, 9000 Gent (BE).

(72) Inventors: **SAVVIDES, Savvas**; Morinnegoed 9, 8500 Kortrijk (BE). **VERSTRAETE, Kenneth**; Monterreystraat 58, 9000 Gent (BE).

(74) Common Representative: **VIB VZW**; Rijvisschestraat 120, 9052 Gent (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2020/182615 A1

(54) Title: CRYSTAL STRUCTURE OF ATP CITRATE LYASE AND USES THEREOF

(57) Abstract: The present invention provides crystals of ATP citrate lyase (ACL) and the citryl-CoA lyase (CCL) module of ACL and the corresponding three-dimensional crystal structures. The invention also provides computer-assisted and other methods for selecting molecules able to modulate the activity of ACL.

CRYSTAL STRUCTURE OF ATP CITRATE LYASE AND USES THEREOF

Field of the invention

The present invention provides crystals of ATP citrate lyase (ACL) and the citryl-CoA lyase (CCL) module of ACL and the corresponding three-dimensional crystal structures. The invention also provides computer-assisted and other methods for selecting molecules able to modulate the activity of ACL.

Introduction to the invention

ATP citrate lyase (ACL) catalyzes the ATP and coenzyme A (CoA) dependent conversion of citrate, a metabolic product of the Krebs cycle, to oxaloacetate and the high-energy biosynthetic precursor acetyl-CoA[1,2]. The latter enables a plethora of essential biochemical reactions such as fatty acid, cholesterol, and acetylcholine synthesis[3], and histone or protein acetylation[4,5]. Intriguingly, ACL in autotrophic prokaryotes is a hallmark enzyme of the reductive tricarboxylic acid (rTCA) pathway or reverse Krebs cycle, which fixates two molecules of carbon dioxide in acetyl-CoA[6,7]. In humans, ACL is strongly expressed at sites for *de novo* lipogenesis, such as the liver and adipose tissue, and in cholinergic neurons[8], and its post-translational stability[9] and allosteric activitation[10] are well regulated. The central role of ACL in human metabolism inspired its possible therapeutic relevance in breast and lung cancer[11-14], and for lowering low-density lipoprotein cholesterol in atherosclerotic cardiovascular disease[15,16]. In prokaryotes, fungi, and plants, ACL is composed of two separate subunits, ACL-A and ACL-B, whereas mammalian ACL is a contiguous polypeptide featuring a citryl-CoA synthetase module (CCS), comprised of CCSβ and CCSα regions, and a C-terminal citryl-CoA lyase (CCL) domain[17,18]. The CCS module initiates catalysis by the ATP-driven auto-phosphorylation of an active-site histidine (His760 in human ACL), resulting in the formation of citryl-phosphate, which then reacts with CoA to generate the high-energy citryl-CoA intermediate[19] (Fig. 5a). In the last step of the reaction, citryl-CoA is cleaved into the reaction products oxaloacetate and acetyl-CoA by the CCL domain, which displays homology to citrate synthase[17]. Unfortunately, the structural basis of ACL's function is not known. Initial structural analysis on crystallization-recalcitrant full-length human ACL (hACL) by negative-stain electron microscopy revealed that the ACL enzyme displays flexible arms around a compact core but also suggested substantial conformational heterogeneity under such experimental conditions (Fig. 5b).

In view of the importance of ACL in glucose and lipid metabolism, investigators have done much work for designing and developing ACL inhibitors for hypolipidemic treatment of metabolic and malignant diseases. In particular, mechanism-based, active-site directed, or tight-binding small chemicals have been developed, having a specific inhibition on ACL. In order to be able to identify more specific, and more potent inhibitors for ACL there is a need to have a better understanding of the catalytic mechanism

of this enzyme and to be able to guide the optimization of existing (and selection of future) compounds by the three-dimensional structure of the enzyme. The present invention satisfies this need.

Summary of the invention

Therefore, the present invention provides in a first embodiment a composition in crystalline form comprising a citryl-CoA lyase (CCL) and its natural ligand selected from citrate, or citrate and coenzyme A (CoA), wherein CCL is a module of ATP citrate lyase, and which module has an amino acid sequence with at least 90% identity to amino acid residues 836 to 1101 depicted in SEQ ID NO: 1, characterized in that the crystal is:

i) a crystal of CCL bound by citrate in the space group P212121, with the following crystal lattice constants: a=67.34 Å ± 5%, b=123.89 Å ± 5%, c=139.80 Å ± 5%, α=β=γ=90°, or

ii) a crystal of CCL bound by citrate and CoA in the space group P21, with the following crystal lattice constants: a=67.71 Å ± 5%, b=147.90 Å ± 5%, c=113.69 Å ± 5%, α=90°, β=105.28° ± 5%, γ=90°, or

iii) a crystal of CCL bound by citrate and CoA in the space group C2221, with the following crystal lattice constants: a=82.83 Å ± 5%, b=111.19 Å ± 5%, c=143.37 Å ± 5%, α=β=γ=90°.

In a further embodiment the invention provides a composition which has a three-dimensional structure wherein the crystal i) comprises an atomic structure characterized by the crystallographic (or atomic coordinates which is an equivalent word for crystallographic coordinates) coordinates 6HXK or a subset of crystallographic coordinates of 6HXK or wherein the crystal ii) comprises an atomic structure characterized by the crystallographic coordinates of 6HXL or a subset of crystallographic coordinates of 6HXL or wherein the crystal iii) comprises an atomic structure characterized by the crystallographic coordinates of 6HXM or a subset of crystallographic coordinates of 6HXM.

In yet another embodiment the invention provides a catalytic pocket, consisting of a subset of atomic coordinates, present in the crystals i), ii) or iii) as defined herein before wherein said catalytic pocket consists of the amino acid residues His_{900}, Val_{904}, Phe_{935}, Gly_{936}, Lys_{964}, Leu_{969}, Ile_{970}, Met_{971}, Gly_{972}, Ile_{973}, Gly_{974}, His_{975}, Arg_{986}, Lys_{1018}, Leu_{1021}, Asn_{1020}, Asn_{1024}, Val_{1025}, Asp_{1026}, Arg_{1065} and Arg_{1085} as depicted in SEQ ID NO : 1 and wherein said amino acid residues represent the citrate and CoA binding sites.

In yet another embodiment the invention provides a crystallizable composition comprising a mixture of

i. a solubilized CCL and its natural ligand citrate or its natural ligands citrate and CoA, of which the amino acid sequence of CCL has at least 90% identity to amino acid residues 836 to 1101 depicted in SEQ ID NO: 1, and

ii. a mother liquor solution, wherein the mother liquor solution comprises about 0.2M Na_2SO_4 pH6.7 and about 20% w/v PEG_{3350} to obtain a crystal i) as defined in claims 1 or 2, or wherein the mother liquor solution comprises about 21% w/v PEG_{8000}, about 0.1M HEPES pH7.0 and about 0.15M ammonium sulphate to obtain a crystal ii) as defined in claims 1 or 2, or wherein the mother liquor solution comprises about 0.1M Na/K phosphate pH5.5, about 0.1 M RbCl, about 0.1 M Na-citrate pH 5.5 and about 25% v/v PEG Smear medium to obtain a crystal iii) as herein before.

In yet another embodiment the invention provides a method for producing a composition of as defined herein before comprising the steps of:

i. producing and purifying a citryl-CoA lyase (CCL) domain or module which has an amino acid sequence with at least 90% identity to amino acid residues 836 to 1101 depicted in SEQ ID NO: 1 and,

ii. mixing with a mother liquor solution and citrate or citrate and CoA, wherein the mother liquor solution comprises about 0.2M Na_2SO_4 pH6.7 and about 20% w/v PEG_{3350} and citrate to obtain a crystal i) as defined herein before, or wherein the mother liquor solution comprises about 21% w/v PEG_{8000}, about 0.1M HEPES pH7.0 and about 0.15M ammonium sulphate and citrate and CoA to obtain a crystal ii) as defined herein before, or wherein the mother liquor solution comprises about 0.1M Na/K phosphate pH5.5, about 0.1 M RbCl, about 0.1 M Na-citrate pH5.5 and about 25% v/v PEG Smear medium and citrate and CoA to obtain a crystal iii) as defined herein before.

In yet another embodiment the invention provides a computer-assisted method of identifying, designing or screening for a compound that can potentially interact with a crystal of CCL selected from a crystal i), ii) or iii) as defined herein before, comprising performing structure-based identification, design or screening of a compound based on the compound's interactions with a structure defined by the atomic coordinates as defined in herein before.

In yet another embodiment the invention provides a method for identifying a compound that can bind to CCL, comprising dipping candidate small molecule compounds with CCL, or with CCL and citrate, or with CCL, citrate and CoA and allowing co-crystallization, and screening candidate ligands or antagonists

by using a method for measuring intermolecular interaction and comparing, designing and docking the 3D structures i), ii) or iii) as defined in before and a candidate ligand by computer modeling.

In yet another embodiment the invention provides a method of identifying, designing or screening for a compound that can interact with CCL, including performing structure-based identification, design, or screening of a compound based on the compound's interactions with a CCL structure or a subset thereof.

In yet another embodiment the invention provides a method for identifying an agonist or antagonist compound of ATP citrate lyase comprising an entity selected from the group consisting of an antibody, a peptide, a non-peptide molecule and a chemical compound, wherein said compound is capable of enhancing, eliciting or blocking biological activity of ATP citrate lyase resulting from an interaction with CCL wherein said process includes:

i. introducing into suitable computer program parameters defining an interacting surface based on the conformation of CCL corresponding to the atomic coordinates of 6HXK, 6HXL or 6HXM or a subset of said atomic coordinates, wherein said program displays a three-dimensional model of the interacting surface,
ii. creating a three-dimensional structure of a test compound in said computer program;
iii. displaying a superimposing model of said test compound on the three-dimensional model of the interacting surface;
iv. and assessing whether said test compound model fits spatially into a binding site.

The atomic coordinates (or crystallographic coordinates) of the structures are available in the Protein Data Bank (PDB), which can be found on the internet on the following link https://www.rcsb.org/. Please note that in the priority document Appendix I corresponds with 6HXK, Appendix II corresponds with 6HXL and Appendix III corresponds with 6HXM.

In yet another embodiment the invention provides a screening method for identifying molecules interacting with CCL comprising:

i. soaking a multiplicity of molecules into the crystals i), ii) or iii) as defined herein before,
ii. determining the co-crystal structure with the at least one molecule binding to the crystal i), ii) or iii).

Figure legends

Figure 1: Structure and mechanism of human ATP citrate lyase

(a) Domain organization of hACL. The dotted line represents the linker that was removed to create hACL-A/B used in crystallographic studies. (b) Cartoon representation of a crystal structure for hACL-A/B in

space group *P*1 colored according to the scheme in panel a. The three perpendicular 2-fold axes of the D2 symmetric CCL module are shown. Bound substrates (citrate, CoASH, and ADP) are shown as colored spheres. (c) Overlay of a crystal structure for the isolated human CCL module (pink CoA-binding domains) with the human CCL module extracted from the hACL-A/B crystal structure (white CoA-binding domains). Citrate and CoASH are shown as colored spheres. (d) Superposition of open and closed CCL protomers (as in panel c) with the active sites of the CCS and CCL modules highlighted. Citrate and CoASH are shown as sticks. (e) Catalytic itinerary in the ACL enzyme.

Figure 2: Evolutionary origin of ACL and its distinct metabolic functions

Simplified evolutionary tree illustrating the distribution of ACL, CCS and CCL enzymes. Determined crystal structures of *H. sapiens*, *M. concilii* and *C. limicola* ACL, and *H. thermophilus* CLL and CCS are shown as cartoons using the same coloring scheme as in Fig. 1. Distinct metabolic roles of ACL, CCS and CCL in the prokaryotic (two-step) rTCA cycle and eukaryotic citrate shuttle (and lipid metabolism) are indicated.

Figure 3: ACL enzymes undergo conformational switching during catalysis

(a-b) Views of the crystal structures for hACL-A/B (panel a) and *C. limicola* ACL-A/B (panel b) with the CCL modules in surface mode. The CCSα β-hairpin (orange) and two-helical stalk regions (green) extending from the CCL modules are highlighted. The curved arrows indicate the proposed conformational switching of human ACL. (c-d) SAXS profiles for hACL-A/B (panel c) and *C. limicola* ACL-A/B (panel d) plotted as the scattered intensity in function of momentum transfer $s = 4\pi\sin\vartheta/\lambda$, where λ is the beam wavelength and 2ϑ is the scattering angle: (*i*) comparison between the experimental scattering profiles recorded in HBS buffer (green) and in HBS buffer supplemented with citrate (brown), (*ii*) comparison between the experimental scattering profiles recorded in HBS buffer (green) and in HBS buffer supplemented with both citrate and CoASH (purple), (*iii*) experimental scattering profiles recorded in HBS buffer overlayed with scattering profiles (red) calculated from substrate-bound crystal structures for hACL-A/B and *C. limicola* ACL-A/B, and (*iv*) scattering profiles recorded in HBS buffer supplemented with both citrate and CoASH overlayed with scattering profiles (red) calculated from substrate-bound crystal structures for hACL-A/B and *C. limicola* ACL-A/B. Fits were calculated with Crysol 3.0. In panel c, the reported Chi2-values are the average and standard deviation for the fits calculated from the different hACL-A/B crystal structures (n=4) extracted from the *P*1 and *C*2 crystals forms, and the best fitting theoretical curve is shown.

Figure 4: Citrate synthase evolved from an ancestral citryl-CoA lyase module.

(a) Cartoon representation of *H. thermophilus* CCL and *P. furiosis* CS (pdb 1aj8) with CCL and CS colored by chain. Only helical secondary structure elements are shown. Bound substrates are shown as colored spheres. (b) View on the 20-helix cores and wrapping C-terminal tails of CCL and CS. In this view the CoA-binding domains are omitted for clarity. (c) Proposed evolutionary history of CCL and CS. (d) Side view

on CCL and CS highlighting the pseudo-twofold symmetry in the CS protomer. CoA is shown as sticks. (e) Proposed evolutionary relationship between the reverse Krebs (or rTCA) cycle and the oxidative Krebs cycle based on the relation between CCL and CS. (f) Detail of the CCL active site of hACL in complex with citrate and CoA (pdb 6hxl). The low-barrier hydrogen bond (2.4 Å) between Asp1026 and bound citrate is indicated. (g) Detail of the CCL active site of *C. limicola* ACL in complex with L-malate and acetyl-CoA (pdb 6qcl). (h) Detail of the active site of chicken CS in complex with D-malate and acetyl-CoA (pdb 4csc). Dashed lines indicate polar interactions.

Figure 5: Structure of the human ACL

(a) Reaction scheme for ATP citrate lyase. In the first step, ACL undergoes autophosphorylation at His760. Citryl-phosphate (citryl-P) and citryl-CoA form non-covalent enzyme-intermediate complexes.

(b) Left: representative class averages for hACL as obtained by negative stain electron microscopy. The size of the box is 40 x 40 nm. Right: flowchart of the 3D-reconstruction in C2 symmetry. SDG: stochastic gradient descent.

(c) Coomassie-stained SDS-PAGE gel for recombinantly produced ACL enzymes. Lane 1: hACL-A/B, lane 2: hACL, lane 3: *M. concilii* ACL-A/B, lane 4: *C. limicola* ACL-A/B and lane 5: hACL-His760Ala.

(d) Initial reaction rates for hACL-A/B, hACL and hACL-His760Ala plotted in function of ATP concentration. The data was fitted by a Michaelis-Menten equation and the obtained K_m and k_{cat} values, and standard deviation are shown. Data points represent averages of technical replicates (n=3).

(e) SEC elution profile of engineered hACL-A/B plotted as the light scattering intensity at 90° in function of the elution volume. The molecular weight determined by MALLS is indicated. The theoretical weight for hACL-A/B is 462 kDa.

(f) Representative crystal structure for hACL-A/B extracted from the *P*1 crystal form and colored by chain. Bound substrates are shown as colored spheres.

(g) View on the helical bundle core of the CCL module with the protruding two-helical stalk regions indicated. CoA-binding domains are omitted for clarity.

(h) Overlay of the four hACL-A/B crystal structures extracted from the *P*1 and C2 crystal forms. The overlay is based on the superposition of the CCL modules. Structures are colored according to the scheme shown in figure 1a. A zoom-in view shows the structural plasticity around the two-helical stalk region.

(i) View on the helical bundle core of CCL from *H. thermophilus* colored by chain. The N- and C- termini of a single chain are indicated.

Figure 6: CoA-binding modes in the CCS-module of ACL and related CCS

(a) View on the CoA-binding mode in hACL-A/B crystal structures. (b) Detail of the CoA-binding mode at the interface between the CCL and CCS modules. The so-called power helices in the CCS module are

indicated. Dashed lines represent polar interactions. (c) View on the CoA-binding mode in the crystal structure for ACL-A/B from *M. concilii*. (d) View on the CoA-binding modes in the crystal structure for ACL-A/B from *C. limicola*. In this structure, the phosphopantheine tails of the CoA-molecules were partly disordered. (e) View on the CoA-binding mode in CCS from *H. thermophilus*. (f) Cartoon representation of CCS from *H. thermophilus* and succinyl-CoA synthetase (SCS) from *E. coli*, both in complex with CoASH. α-subunits are colored in blue and β-subunits in grey. The C-terminal tail extending from CCSα to CCSβ is in orange.

Figure 7: Structural plasticity in the citryl-CoA lyase modules of human and *C. limicola* ACL

(a) Crystal structure of the CCL module of hACL in space group $P2_12_12_1$, in complex with citrate. (b) Crystal structure of the CCL module of hACL in space group $C222_1$, in complex with citrate and CoA. (c) Crystal structure of the CCL module of hACL in space group $P2_1$, in complex with citrate and CoA. This crystal form contained two tetramers in the crystallographic asymmetric unit (asu). In panels a-c, CoA-binding domains are colored according to the structural state of the CCL active site: open (white), intermediate (blue) and closed (magenta), and substrates are shown as colored spheres.

(d) Binding mode of CoA as seen in the hACL-A/B crystal structure (left) compared to CoA binding in a closed CCL module protomer (right). Substrates are shown as colored sticks and dashed lines indicate polar interactions.

(e) A CCL module protomer in the open state as seen in the hACL-A/B structure (white CoA-binding domain) overlayed with a protomer in the closed state (magenta CoA-binding domain). The latter was extracted from a crystal structure for the isolated CCL module of hACL (panel c). Arrows indicate structural transitions.

(f) Crystal structure for the CCL module of *C. limicola* ACL in space group $P2_1$, in complex with citrate. This crystal form contained two tetramers in the asu. In the second tetramer (right), one of the CoA-binding domains was not modelled due to disorder. (g) Crystal structure for the CCL module of *C. limicola* ACL in space group $P3_121$, in complex with CoA. (h) CCL module of *C. limicola* ACL as observed in the *C. limicola* ACL crystal structure. (i) Overlay of *C. limicola* CCL module protomers colored according to the structural state of their active site: open (white), intermediate (blue) and closed (magenta).

Figure 8: ACL structures across different domains of life

Cartoon representations of the crystal structures of ACL from *H. sapiens* (left), *M. concilii* (middle) and *C. limicola* (right). The CCS modules are shown in surface mode. Distinct structural regions are colored according to the coloring scheme in figure 1a.

Figure 9: Conformational switching of ACL during catalysis

(a) View on the interaction between the CCS and CCL modules in a representative hACL-A/B crystal structure (space group *P*1), with the CCSα β-hairpin (orange) and CoA-binding domain (pink) in cartoon

mode. Bound CoASH is shown as colored spheres. (b) Overlay of a human CCL protomer in the closed state (pink CoA-binding domain) with the crystal structure of hACL-A/B as in panel a (white CoA-binding domain). The resulting clash between the CoA-binding domain (with bound CoA-molecule) and the β-hairpin is indicated by a red box. (c) View on the interaction between the CCS and CCL module in the crystal structure of *C. limicola* ACL-A/B with the CCSα β-hairpin (orange) and CoA-binding domain (pink) in cartoon mode. (d) Zoom-in view on the stalk region in the crystal structures for hACL-A/B and *C. limicola* ACL-A/B based on the superposition of the helical core of the CCL module. (e,f) Interactions at the stalk region and β-hairpin as observed in the crystal structures of hACL-A/B and *C. limicola* ACL-A/B. (g) Two-state rigid-body SAXS model for apo-hACL-A/B (MultiFoXS, χ^2=2.8) overlaid with the hACL-A/B crystal structures in space groups *P*1 and *C*2 (grey). (h) Single-state rigid-body SAXS model for hACL-A/B (MultiFoXS, χ^2=2.8) in the presence of both citrate and CoASH overlaid with the hACL-A/B crystal structures in space groups *P*1 and *C*2 (grey). (i) Comparison between in-solution SAXS scattering profiles measured from linker-deleted hACL-A/B and full-length hACL: (*i*) profiles recorded from hACL-A/B (green) and hACL (grey) in HBS-buffer, (ii) profiles recorded from hACL-A/B (purple) and hACL (black) in HBS-buffer supplemented with citrate and CoASH, (*iii*) profiles recorded from hACL in HBS-buffer (grey) and from hACL in HBS-buffer supplemented with both citrate and CoASH (black), (*iv*) profiles recorded from hACL-A/B in HBS-buffer (green) and from hACL-A/B in HBS-buffer supplemented with both citrate and CoASH (purple), and (*v*) fit of the theoretical scattering profile (red) calculated from an AllosMod-FoXS model for hACL (as shown in panel j) to the experimental scattering profile recorded in the presence of citrate and CoASH (black). (j) AllosMod-FoXS SAXS-model for hACL in HBS-buffer supplemented with citrate and CoASH, overlaid with the hACL-A/B crystal structures in space groups *P*1 and *C*2 (grey).

In panels g, h and i, the bottom numeric table presents an all-residue (Cα) rmsd matrix for the hACL-A/B crystal structures and presented SAXS-models and for each crystal structure and model the calculated fit (χ^2-value) against the recorded SAXS-data is shown as calculated by FoXS, Crysol and Crysol 3.0. P1-hACL-A/B_1 and P1-hACL-A/B_2: structures for hACL-A/B extracted from the *P*1 crystal form, C2-hACL-A/B_1 and C2-hACL-A/B_2: structures for hACL-A/B extracted from the *C*2 crystal form.

Figure 10: Citrate synthase evolved from an ancestral CCL module

(a) Side-by-side comparison and overlay of the helical bundle cores of *H. thermophilus* CCL and *P. furiosis* CS (pdb 1aj8). (b) Two adjacent CCL protomers (CCL and CCL') extracted from the *H. thermophilus* CCL tetramer. (c) A CS protomer extracted from *P. furiosis* CS. (d) CCL without its CoA binding domain (residues 2-100 and 204-231) aligned with the N-terminal half of CS (residues 6 - 143) (e) CCL' (residues 30-256) aligned with the C-terminal half of CS (residues 154 - 376) (f) CCL/CCL' aligned with the CS protomer. (g) Sequence alignment between CCL/CCL' and CS protomer sequences. Top secondary structure elements correspond to *H. thermophilus* CCL and CCL', bottom secondary structures

correspond to *P. furiosis* CS. CS active site residues are indicated with a purple arrow. (h-i) Details of the overlay between CCL/CCL' and the CS protomer.

Figure 11: Homology between ACL and citrate synthase

(a) Sequence alignment between the C-terminal regions of ACL, CCL and CS. Active site residues are highlighted according to the chicken CS numbering scheme. His274, highlighted in yellow, is not conserved in ACL sequences. (b) Comparison between crystal structures for the CCL module of hACL and chicken CS in open (pdb 5csc) and closed (pdb 5cts) states. For clarity only the helical secondary structure elements are shown as cylinders. Bound substrates are shown as colored spheres. (c) Overlay of the CCL active site of hACL (blue), in complex with citrate and CoASH, with the active site of chicken CS (orange), in complex with oxaloacetate and carboxymethyl-CoA (pdb 5cts). The interaction of Asp1026 with the carboxylate group of citrate in hACL, and the interaction of Asp375 with the carboxylate group of carboxymethyl-CoA in chicken CS is indicated by a dashed line. (d) By analogy to the aldol condensation of acetyl-CoA and oxaloacetate to citryl-CoA as catalyzed by CS (top), citryl-CoA may undergo retro-aldol cleavage catalyzed by ACL as indicated by the chemical reaction arrows (bottom). Dashed lines indicate polar interactions.

Detailed description of the invention

The present invention will be described with respect to particular embodiments and with reference to certain drawings but the invention is not limited thereto but only by the claims. Any reference signs in the claims shall not be construed as limiting the scope. The drawings described are only schematic and are non-limiting. In the drawings, the size of some of the elements may be exaggerated and not drawn on scale for illustrative purposes. Where the term "comprising" is used in the present description and claims, it does not exclude other elements or steps. Where an indefinite or definite article is used when referring to a singular noun e.g. "a" or "an", "the", this includes a plural of that noun unless something else is specifically stated. Furthermore, the terms first, second, third and the like in the description and in the claims, are used for distinguishing between similar elements and not necessarily for describing a sequential or chronological order. It is to be understood that the terms so used are interchangeable under appropriate circumstances and that the embodiments of the invention described herein are capable of operation in other sequences than described or illustrated herein. The following terms or definitions are provided solely to aid in the understanding of the invention. Unless specifically defined herein, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook et al., Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Press, Plainsview, New York (2012); and Ausubel et al., current Protocols in Molecular Biology (Supplement 100), John Wiley & Sons, New York (2012), for definitions

and terms of the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g. in molecular biology, biochemistry, structural biology, and/or computational biology).

The human ATP citrate lyase enzyme is a central component of the citrate shuttle which transports carbohydrate-derived acetyl-CoA from mitochondria to the cytoplasm for *de novo* lipogenesis and cholesterol synthesis (see Fig. 2). The structural and mechanistic insights presented in the current invention collectively provide the structural framework for targeting human ACL in cancer[11] and hyperlipidemia[16].

As used herein the term "homologue" means a protein having at least 80% amino acid sequence identity with human ACL, in particular at least 80% amino acid sequence identity with the CCL module of ACL. Preferably, the percentage identity is 85, 90%, 95% or higher.

The present invention provides crystals of ACL and crystals of the CCL module of ACL.

The crystal structure of the human ACL is presented in Figures 1, 5 and 8.

The crystal structure of the human CCL module of ACL is presented in Figure 4 and Figure 7.

Details about the crystal structures are depicted in Table 1.

Table 1: Crystallographic data and refinement statistics[§,*]

	ACL-A/B			
Source	*H. sapiens*	*H. sapiens*	*M. concilii*	*C. limicola*
Crystallization conditions	13.5% PEG_{3350} 0.2 M Na_2HPO_4 pH 8.15	17% PEG_{3350} 0.2 M Na_2HPO_4 pH 8.15	14% PEG_{3350} 4% tacsimate pH 8.0	24% PEG Smear High 0.1 M Tris pH 8.0
Protein concentration	8 mg/ml	14 mg/ml	7 mg/ml	12 mg/ml
Cryoprotectant	25% glycerol	25% glycerol	22% PEG_{400}	22% PEG_{400}
Ligands	Citrate, CoASH, Mg.ADP, Phosphate ion	Citrate, CoASH, Mg, ADP, Phosphate ion	Citrate, CoASH	Citrate, CoASH
Data collection				
X-ray source (beamline)	SLS (PXI)	PETRAIII (P14)	SLS (PXIII)	ESRF (ID30-B)
Wavelength (λ)	0.999977	0.9763	1.00003	0.97625
Space group	*P*1	*C2*	*P*21212	*P*21
Cell dimensions				
a, *b*, *c* (Å)	150.36, 154.01, 154.09	214.28, 215.40, 158.42	118.45, 275.02, 72.72	101.67, 134.93, 177.71
α,β,γ (°)	91.53, 110.04, 107.46	90.00, 117.27, 90.00	90.00, 90.00, 90.00	90.00, 101.65, 90.00
Resolution (Å)	50.00-3.30 (3.38-3.30)	50.00-3.25 (3.45-3.25)	50.00-2.10 (2.23-2.10)	50.00-2.58 (2.74-2.58)
R_{meas} (%)	20.1 (280.0)	12.4 (199.1)	11.2 (141.5)	11.1 (176.5)
$< I/\sigma >$	8.38 (0.77)	11.11 (0.91)	13.07 (1.09)	9.15 (0.86)
CC ½ (%)	99.7 (28.3)	99.9 (34.1)	99.9 (38.7)	99.7 (36.1)
Completeness (%)	95.4 (95.7)	99.2 (98.2)	99.4 (97.1)	98.5 (93.5)
Redundancy	7.5 (7.5)	4.5 (4.5)	6.1 (5.1)	3.3 (3.2)
Wilson B factor ($Å^2$)	106.72	97.57	43.11	68.91
Refinement				
Resolution (Å)	49.00-3.30	47.83-3.25	39.00-2.10	47.50-2.58
No. reflections	176 285	99 634	138 531	144 937
R_{work} / R_{free} (%)	15.58 / 18.71	16.16 / 19.25	17.94/ 20.79	18.33 / 21.63
No. atoms				
Protein	64 032	31 725	15 704	27 964
Ligand/ion	760	320	177	497
Water	16	0	724	254
B-factors ($Å^2$)				
Protein	127.39	137.05	64.03	93.32
Ligand/ion	128.41	138.21	52.79	111.17

ACL-A/B

Source	*H. sapiens*	*H. sapiens*	*M. concilii*	*C. limicola*
Water	98.04	-	56.89	70.81
R.m.s. deviations				
Bond lengths (Å)	0.010	0.010	0.010	0.010
Bond angles (°)	1.16	1.13	1.07	1.17
Pdb id	6hxh	6qfb	6hxi	6hxj

CCL module of ACL

Source organism	*H. sapiens*	*H. sapiens*	*H. sapiens*	*C. limicola*	*C. limicola*	*C. limicola*
Crystallization conditions	0.2 M Na_2SO4 pH 6.7 20% PEG_{3350}	21% PEG_{8000} 0.1 M HEPES pH 7.0 0.15 M ammonium sulphate	0.1 M Na/K phosphate pH 5.5 0.1 M RbCl 0.1 M Na-citrate pH 5.5 25% PEG Smear Medium	1.6 M Lithium sulphate 0.1 M Tris pH 8.0	22% PEG Smear Broad 0.1 M MES pH 6.5	0.2 M $CaCl_2$ pH 5.1 20% PEG_{3350}
Protein concentration	18 mg/ml	7 mg/ml	7 mg/ml	13 mg/ml	12 mg/ml	12 mg/ml
Cryoprotectant	25% PEG_{400}	25% PEG_{400}	25% PEG_{400}	20% glycerol	25% PEG_{400}	30% PEG_{400}
Ligands	Citrate	Citrate, CoASH	Citrate, CoASH	CoASH	Citrate	L-malate, Acetyl-CoA
Data collection						
X-ray source (beamline)	PETRAIII (P14)	SLS (PXIII)	SLS (PXIII)	SOLEIL (Proxima 2A)	ESRF (ID23-1)	PETRAIII (P13)
Wavelength (λ)	0.9762	1.00003	1.00003	0.980042	0.97242	0.9762
Space group	*P*212121	*P*21	*C*2221	*P*3121	*P*21	*P21*
Cell dimensions						
a, *b*, *c* (Å)	67.34, 123.89, 139.80	67.71, 147.90, 113.69	82.83, 111.19, 143.37	110.03, 110.03, 92.75	99.26, 106.41, 105.49	102.73, 100.82, 105.01
α,β, γ (°)	90.00, 90.00, 90.00	90.00, 105.28, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 120.00	90.00, 102.11, 90.00	90.00, 102.99, 90.00
Resolution (Å)	50.00-1.85 (1.96-1.85)	50.00-1.35 (1.43-1.35)	50.00-1.30 (1.38-1.30)	50.00-1.70 (1.80-1.70)	50.00-1.50 (1.59-1.50)	50.00-1.60 (1.70-1.60)
R_{meas} (%)	20.5 (160.6)	12.8 (164.5)	13.5 (203.2)	16.7 (154.2)	9.5 (164.1)	12.6 (146.2)
$< I/\sigma >$	8.69 (1.26)	8.96 (1.05)	12.10 (1.12)	14.87 (1.64)	9.07 (0.78)	11.86 (1.31)
CC ½ (%)	99.6 (49.9)	99.7 (42.1)	99.9 (50.8)	99.9 (64.8)	99.8 (32.4)	99.8 (55.2)
Completeness (%)	99.8 (99.5)	97.4 (94.2)	99.5 (97.6)	99.9 (99.5)	97.4 (95.6)	98.8 (97.6)

CCL module of ACL						
Source organism	*H. sapiens*	*H. sapiens*	*H. sapiens*	*C. limicola*	*C. limicola*	*C. limicola*
Redundancy	9.2 (9.2)	6.9 (6.7)	10.3 (10.3)	15.3 (15.4)	3.5 (3.5)	6.8 (6.9)
Wilson B factor (Å^2)	18.31	13.00	11.56	25.82	27.85	14.32
Refinement						
Resolution (Å)	61.94-1.85	47.85-1.35	48.72-1.30	34.00-1.70	48.53-1.50	45.62-1.60
No. reflections	100 130	458 802	160 728	71 480	332 071	270 868
R_{work} / R_{free} (%)	16.50 / 19.32	15.62 / 17.43	16.67 / 17.85	17.28 / 19.21	19.39 / 21.12	15.67 / 17.73
No. atoms						
Protein	8 104	16 312	4106	3960	15 342	15 543
Ligand/ion	107	478	138	59	114	288
Water	593	1713	438	476	1281	1066
B-factors (Å^2)						
Protein	29.01	19.60	17.59	20.48	28.21	27.53
Ligand/ion	51.14	24.54	19.86	37.11	30.50	29.40
Water	36.24	30.49	27.22	32.98	34.11	32.52
R.m.s. deviations						
Bond lengths (Å)	0.010	0.010	0.010	0.010	0.010	0.010
Bond angles (°)	1.01	1.11	1.06	0.92	0.95	1.01
Pdb id	6hxk	6hxl	6hxm	6hxn	6hxo	6qcl

	CCL	CCS
Source organism	*H. thermophilus*	*H. thermophilus*
Crystallization conditions	45% MPD 0.1 M HEPES pH 7.5 0.2 M ammonium acetate	18% PEG_{4000} 20% propanol 0.1 M sodium citrate pH 5.6
Protein concentration	7 mg/ml	11 mg/ml
Cryoprotectant	49% MPD	22% PEG_{400}
Ligands	CoASH	CoASH
Data collection		
X-ray source (beamline)	SLS (PXIII)	SLS (PXIII)
Wavelength (λ)	1.00003	1.00003
Space group	*I*4122	*P*212121
Cell dimensions		
a, *b*, *c* (Å)	130.60, 130.60, 83.95	118.50, 127.03, 143.80
α,β,γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00-1.45 (1.54-1.45)	50.00-2.91 (3.08-2.91)
R_{meas} (%)	5.9 (202.4)	15.3 (217.2)
$< I / \sigma >$	16.29 (0.88)	12.83 (0.83)
CC ½ (%)	100 (46.7)	99.8 (48.4)
Completeness (%)	99.9 (99.5)	99.5 (97.8)
Redundancy	6.7 (6.0)	6.5 (6.4)
Wilson B factor (Å^2)	20.88	80.84
Refinement		
Resolution (Å)	47.94-1.45	47.60-2.91
No. reflections	63 952	48 218

	CCL	CCS
Source organism	*H. thermophilus*	*H. thermophilus*
R_{work} / R_{free} (%)	14.48 / 16.22	18.93 / 22.29
No. atoms		
Protein	2006	11 552
Ligand/ion	58	149
Water	211	0
B-factors (Å²)		
Protein	34.78	99.24
Ligand/ion	75.82	95.78
Water	47.12	-
R.m.s. deviations		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.03	1.21
Pdb id	6hxp	6hxq

§Each dataset was collected from a single crystal. *Values in parentheses are for highest-resolution shell.

In the present invention the abbreviations CoA and CoASH are used interchangeably and refer to coenzyme A.

As used herein, the term "crystal" means a structure (such as a three-dimensional (3D) solid aggregate) in which the plane faces intersect at definite angles and in which there is a regular structure (such as an internal structure) of the constituent chemical species. The term "crystal" refers in particular to a solid physical crystal form such as an experimentally prepared crystal.

Crystals may be constructed with the wild-type CCL polypeptide sequences or variants thereof, including allelic variants and naturally occurring mutations as well as genetically engineered variants. Typically, variants have at least 90%, at least 95% sequence identity with a corresponding wild-type CCL polypeptide. In a preferred embodiment the CCL polypeptide is human CCL. The CCL module of ACL corresponds to amino acid residues 836-1101 in SEQ ID NO: 1. SEQ ID NO: 1 depicts the full-length amino acid sequence of human ACL.

The amino acid sequence of the human CCL module is depicted in SEQ ID NO: 2.

Optionally, the crystals of the ACL, particularly crystals from the CCL module, may comprise one or more (natural) ligands (such as citrate and/or coenzyme A) or compounds which interact with these proteins, or otherwise are soaked into the crystal or co-crystallized with ACL, particularly the CCL module. Such compounds may be candidate pharmaceutical agents intended to modulate the activity of ACL, particularly the CCL module of ACL.

In a preferred embodiment, crystals of the CCL module in complex with its natural ligands have atomic coordinates set forth in the PDB databank as 6HXK, 6HXL and 6HXM.

6HXK depicts the atomic coordinates of the co-crystal of the CCL module of human ACL bound by its natural ligand citrate, the atomic coordinates represent the open state of CCL (active site is open), Figure 7a shows the structural, active site open state of CCL.

6HXL depicts the atomic coordinates of the co-crystal of the CCL module of human ACL bound by its natural ligands citrate and CoASH, the atomic coordinates represent the closed state of CCL (active site is closed), Figure 7c shows the closed state of CCL.

6HXM depicts the atomic coordinates of the co-crystal of the CCL module of human ACL bound by its natural ligands citrate and CoASH, the atomic coordinates represent the intermediate state of CCL (active site is between open and closed – intermediate), Figure 7b shows the structural intermediate state of CCL.

As used herein, the term "atomic coordinates" or "set of coordinates" refers to a set of values which define the position of one or more atoms with reference to a system of axes. It will be understood by those skilled in the art that the atomic coordinates may be varied, without affecting significantly the accuracy of models derived therefrom. Thus, although the invention provides a very precise definition of a preferred atomic structure, it will be understood that minor variations are envisaged and the claims are intended to encompass such variations.

It will be understood that any reference herein to the atomic coordinates or subset of the atomic coordinates shown in the PDB entries 6HXK, 6HXL and 6HXM shall include, unless specified otherwise, atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5 Å, preferably not more than 1 Å, when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in the PDB entries 6HXK, 6HXL and 6HXM.

The following defines what is intended by the term "root mean square deviation (`RMSD`)" between two data sets. For each element in the first data set, its deviation from the corresponding item in the second data set is computed. The squared deviation is the square of that deviation, and the mean squared deviation is the mean of all these squared deviations. The root mean square deviation is the square root of the mean squared deviation.

In a preferred embodiment, the crystals have the atomic coordinates as shown in the PDB entries 6HXK, 6HXL and 6HXM.

Further, it will be appreciated that a set of atomic coordinates for one or more polypeptides is a relative set of points that define a shape in three dimensions. Thus, it is possible that an entirely different set of coordinates could define a similar or identical shape. Moreover, slight variations in the individual coordinates will have little effect on overall shape. The variations in coordinates may be generated due to mathematical manipulations of the atomic coordinates. For example, the atomic coordinates set forth in the PDB entries 6HXK, 6HXL and 6HXM could be manipulated by crystallographic permutations of the atomic coordinates, fractionalization of the atomic coordinates, integer additions or subtractions to sets of the structure coordinates, inversion of the atomic coordinates, special labeling of amino acids, polypeptide chains, heteroatoms, ligands, solvent molecules, or combinations thereof.
Alternatively, modification in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids, or other changes in any of the components that make up the crystal could also account for variations in atomic coordinates.

Various computational analyses are used to determine whether a molecular complex or a portion thereof is sufficiently similar to all or parts of the structure of the CCL module. Such analyses may be carried out in available software applications which are known to the skilled person. For example, a molecular similarity program permits comparisons between different structures, different conformations of the same structure, and different parts of the same structure. Comparisons typically involve calculation of the optimum translations and rotations required such that the root mean square deviation of the fit over the specified pairs of equivalent atoms is an absolute minimum. This number is given in Angstroms (Å). Accordingly, atomic coordinates of a CCL module of the present invention include atomic coordinates related to the atomic coordinates listed in the PDB entries 6HXK, 6HXL and 6HXM by whole body translations and/or rotations. Accordingly, RMSD values listed above assume that at least the backbone atoms of the structures are optimally superimposed which may require translation and/or rotation to achieve the required optimal fit from which to calculate the RMSD value. A three-dimensional structure of a CCL module or a region thereof which substantially conforms to a specified set of atomic coordinates can be modelled by a suitable modelling computer program, using information, for example, derived from the following data: (1) the amino acid sequence of the CCL polypeptide; (2) the amino acid sequence of the related portion(s) of the protein represented by the specified set of atomic coordinates having a three-dimensional configuration; and (3) the atomic coordinates of the

specified three-dimensional configuration. A three-dimensional structure of a CCL polypeptide which substantially conforms to a specified set of atomic coordinates can also be calculated by a method such as molecular replacement, which is described in detail below.

Atomic coordinates are typically loaded onto a machine-readable medium for subsequent computational manipulation. Thus, models and/or atomic coordinates are advantageously stored on machine-readable media, such as magnetic or optical media and random-access or read-only memory, including tapes, diskettes, hard disks, CD-ROMs and DVDs, flash memory cards or chips and servers. Typically, the machine is a computer. The atomic coordinates may be used in a computer to generate a representation, e.g. an image of the three-dimensional structure of the CCL polypeptide which can be displayed by the computer and/or represented in an electronic file. The atomic coordinates and models derived therefrom may be used for a variety of purposes such as drug discovery, biological reagent (binding protein) selection and X-ray crystallographic analysis of other protein crystals.

Molecular Replacement

The structure coordinates of the CCL polypeptide such as those set forth in the PDB entries 6HXK, 6HXL and 6HXM or a subset thereof, can also be used for determining the three-dimensional structure of a distant crystallized CCL polypeptide (e.g. derived from another species). This may be achieved by any of a number of well-known techniques, including molecular replacement. Methods of molecular replacement are generally known by those skilled in the art. Generally, molecular replacement involves the following steps: i) X-ray diffraction data are collected from the crystal of a crystallized target structure, then ii) the X-ray diffraction data are transformed to calculate a Patterson function, then iii) the Patterson function of the crystallized target structure is compared with a Patterson function calculated from a known structure (referred to herein as a search structure or search model) , iv) the Patterson function of the search structure is rotated on the target structure Patterson function to determine the correct orientation of the search structure in the crystal to obtain a rotation function, v) a translation function is then calculated to determine the location of the search structure with respect to the crystal axes. Alternatively, likelihood-based molecular replacement methods can be used to determine the location of the search structure. Once the search structure has been correctly positioned in the unit cell, initial phases for the experimental data can be calculated. These phases are necessary for calculation of an electron density map from which structural features and differences can be observed to allow construction of a molecular model and refinement of the structure. Preferably, the structural features (e.g., amino acid sequence, conserved disulphide bonds, and beta-strands or beta-sheets) of the search molecule are related to the crystallized target structure. The electron density map

can, in turn, be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown (i.e. target) crystallized molecular structure. Obtaining accurate values for the phases, by methods other than molecular replacement, is often a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a homologous portion has been solved, the phases from the known structure provide a satisfactory starting estimate of the phases for the unknown structure. By using molecular replacement, all or part of the structure coordinates of the CCL polypeptide (and set forth in the PDB entries 6HXK, 6HXL and 6HXM) can be used to determine the structure of another crystallized CCL molecule whose structure is unknown, more rapidly and more efficiently than attempting to determine such information *ab initio*.

The structure of any portion of any crystallized molecular CCL molecule that is sufficiently homologous to any portion of the human CCL molecule can be solved by this method.

Such structure coordinates are also particularly useful to solve the structure of crystals of CCL co-complexed with a variety of molecules, such as chemical entities. For example, this approach enables the determination of the optimal sites for the interaction between chemical entities, and the interaction of candidate ACL, in particular CCL agonists or antagonists.

Target Sites for Compound Identification, Design or Screening

The three-dimensional structure of the CCL module provided by the present invention (see the PDB entries 6HXK, 6HXL and 6HXM) can be used to identify potential target binding sites as well as in methods for identifying or designing compounds which interact with identified target sites in CCL.

A particularly preferred target site, identified in the present invention, is the catalytic pocket of CCL which pocket consists of the amino acid residues His_{900}, Val_{904}, Phe_{935}, Gly_{936}, Lys_{964}, Leu_{969}, Ile_{970}, Met_{971}, Gly_{972}, Ile_{973}, Gly_{974}, His_{975}, Arg_{986}, Lys_{1018}, Leu_{1021}, Asn_{1020}, Asn_{1024}, Val_{1025}, Asp_{1026}, Arg_{1065} and Arg_{1085} as depicted in SEQ ID NO : 1 and wherein said amino acid residues represent the citrate and CoASH binding sites of CCL.

Next to the identification of compounds binding to specific target sites on CCL, it is for example also feasible to specifically target the “open” or “closed” conformation of CCL. Thus, it is possible to use the “open” conformation of CCL which atomic coordinates are depicted in the PDB entry 6HXK, or it is possible to use the “closed” conformation of CCL which atomic coordinates are depicted in the PDB entry 6HXM.

Design, Selection, Fitting and Assessment of Chemical Entities that bind ACL, in particular the CCL module of ACL

Using a variety of known modelling techniques, the crystal structures of the present invention can be used to produce models for evaluating the interaction of compounds with ACL, in particular the CCL module of ACL. As used herein, the term "modelling" includes the quantitative and qualitative analysis of molecular structure and/or function based on atomic structural information and interaction models. The term "modelling" includes conventional numeric-based molecular dynamic and energy minimization models, interactive computer graphic models, modified molecular mechanics models, distance geometry and other structure-based constraint models. Molecular modelling techniques can be applied to the atomic coordinates of CCL or parts thereof to derive a range of 3D models and to investigate the structure of binding sites, such as the binding sites with compounds. These techniques may also be used to screen for or design small and large chemical entities which are capable of binding CCL and modulate the activity of ACL. Such a screen may employ a solid 3D screening system or a computational screening system. Such modelling methods are to design or select chemical entities that possess stereochemical complementary to identified binding sites in CCL. By "stereochemical complementarity" it is meant that the compound makes a sufficient number of energetically favourable contacts with CCL as to have a net reduction of free energy on binding to CCL. Modelling methods may also be used to design or select chemical entities that possess stereochemical similarity to substrates binding to CCL, such as CoASH and citrate. By "stereochemical similarity" we mean that the compound makes about the same number of energetically favourable contacts with CCL set out by the coordinates shown in the PDB entries 6HXK, 6HXL and 6HXM. In addition, modelling methods may also be used to design or select chemical entities that possess stereochemical complementarity to CCL. By stereochemical complementarity it is meant that the compound makes energetically favourable contacts with CCL as defined by coordinates shown in the PDB entries 6HXK, 6HXL and 6HXM. By "match" we mean that the identified portions interact with the surface residues, for example, via hydrogen bonding or by non-covalent van der Waals and Coulomb interactions (with surface or residue) which promote dissolvation of the molecule within the site, in such a way that retention of the molecule at the binding site is favored energetically. It is preferred that the stereochemical complementarity is such that the compound has a K_d for the binding site of less than 10^{-4}M, more preferably less than 10^{-5}M and more preferably 10^{-6}M. In a most preferred embodiment, the K_d value is less than 10^{-8}M and more preferably less than 10^{-9}M.

Chemical entities which are complementary to the shape and electrostatics or chemistry of CCL characterized by amino acids positioned at atomic coordinates set out in the PDB entries 6HXK, 6HXL

and 6HXM will be able to bind to CCL, and when the binding is sufficiently strong, substantially inhibit the catalytic activity of ACL, particularly the CCL module.

A number of methods may be used to identify chemical entities possessing stereochemical and structural complementarity to the structure or substructures of CCL. For instance, the process may begin by visual inspection of a selected binding site in CCL on the computer screen based on the coordinates in the PDB entries 6HXK, 6HXL and 6HXM generated from the machine-readable storage medium. Alternatively, selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within the selected binding site. Modelling software is well known and available in the art. This modelling step may be followed by energy minimization with standard available molecular mechanics force fields. Once suitable chemical entities or fragments have been selected, they can be assembled into a single compound. In one embodiment, assembly may proceed by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of selected binding site in CCL. This is followed by manual model building, typically using available software. Alternatively, fragments may be joined to additional atoms using standard chemical geometry. The above-described evaluation process for chemical entities may be performed in a similar fashion for chemical compounds.

Databases of chemical structures are available from a number of sources including Cambridge Crystallographic Data Centre (Cambridge, U.K.), Molecular Design, Ltd., (San Leandro, Calif.), Tripos Associates, Inc. (St. Louis, Mo.), Chemical Abstracts Service (Columbus, Ohio), the Available Chemical Directory (Symyx Technologies, Inc.), the Derwent World Drug Index (WDI), BioByteMasterFile, the National Cancer Institute database (NCI), Medchem Database (BioByte Corp.), and the Maybridge catalogue. Once an entity or compound has been designed or selected by the above methods, the efficiency with which that entity or compound may bind to CCL can be tested and optimised by computational evaluation. For example, a compound that has been designed or selected to function as a CCL binding compound must also preferably traverse a volume not overlapping that occupied by the binding site when it is bound to the native CCL. An effective CCL binding compound must preferably demonstrate a relatively small difference in energy between its bound and free states (i.e. a small deformation energy of binding). Thus, the most efficient CCL binding compound should preferably be designed with a deformation energy of binding of not greater than about 10 kcal/mole, preferably, not greater than 7 kcal/mole. CCL binding compounds may interact with CCL in more than one conformation that are similar in overall binding energy. In those cases, the deformation energy of binding is taken to be the difference between the energy of the free compound and the average energy of the

conformations observed when the compound binds to the protein. Further, a compound designed or selected as binding to CCL may be further computationally optimised so that in its bound state it would preferably lack repulsive electrostatic interaction with the target protein. Once a CCL-binding compound has been optimally selected or designed, as described above, substitutions may then be made in some of its atoms or side groups to improve or modify its binding properties. Generally, initial substitutions are conservative, i.e. the replacement group will have approximately the same size, shape, hydrophobicity and charge as the original group. It should, of course, be understood that components known in the art to alter conformation should be avoided. Such substituted chemical compounds may then be analysed for efficiency of fit to CCL by the same computer methods described in detail above.

Naturally, specific computer software is available in the art to evaluate compound deformation energy and electrostatic interaction. The screening/design methods may be implemented in hardware or software, or a combination of both. However, preferably, the methods are implemented in computer programs executing on programmable computers each comprising a processor, a data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. Program code is applied to input data to perform the functions described above and generate output information. The output information is applied to one or more output devices, in known fashion. The computer may be, for example, a personal computer, microcomputer, or workstation of conventional design. Each program is preferably implemented in a high level procedural or object-oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. In any case, the language may be compiled or interpreted language. Each such computer program is preferably stored on a storage medium or device (e.g., ROM or magnetic diskette) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

Compounds

Compounds of the present invention include both those designed or identified using a screening method of the invention and those which are capable of recognising and binding to CCL as defined above. Compounds capable of recognising and binding to CCL may be produced using a screening method based

on use of the atomic coordinates corresponding to the 3D structure of CCL. Compounds capable of recognising and binding to CCL may be produced using a screening method based on the use of the atomic coordinates corresponding to the 3D structure of CCL. The candidate compounds and/or compounds identified or designed using a method of the present invention may be any suitable compound, synthetic or naturally occurring, preferably synthetic. In one embodiment, a synthetic compound selected or designed by the methods of the invention preferably has a molecular weight equal to or less than about 5000, 4000, 3000, 2000, 1000 or 500 daltons. A compound of the present invention is preferably soluble under physiological conditions. The compounds may encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons, preferably less than 1,500, more preferably less than 1,000 and yet more preferably less than 500. Such compounds can comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The compound may comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Compounds can also comprise biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogues, or combinations thereof. Compounds may include, for example: (1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids; (2) phosphopeptides (e.g. members of random and partially degenerate, directed phosphopeptide libraries, (3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies, nanobodies as well as Fab, $(Fab)_2$, Fab expression library and epitope-binding fragments of antibodies); (4) non-immunoglobulin binding proteins such as but not restricted to avimers, DARPins and lipocalins; (5) nucleic acid-based aptamers; and (6) small organic and inorganic molecules.

Synthetic compound libraries are commercially available from, for example, Maybridge Chemical Co. (Tintagel, Cornwall, UK), AMRI (Budapest, Hungary) and ChemDiv (San Diego, Calif.), Specs (Delft, The Netherlands). In addition, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts can be readily produced. In addition, natural or synthetic compound libraries and compounds can be readily modified through conventional chemical, physical and biochemical means and may be used to produce combinatorial libraries. In another approach, previously identified pharmacological agents can be subjected to directed or random chemical modifications, such as

acylation, alkylation, esterification, amidification, and the analogues can be screened for ACL, in particular CCL activity. In addition, numerous methods of producing combinatorial libraries are known in the art, including those involving biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide or peptide libraries, while the other four approaches are applicable to polypeptide, peptide, nonpeptide oligomer, or small molecule libraries of compounds. Compounds also include those that may be synthesized from leads generated by fragment-based drug design, wherein the binding of such chemical fragments is assessed by soaking or co-crystallizing such screen fragments into crystals provided by the invention and then subjecting these to an X-ray beam and obtaining diffraction data. Difference Fourier techniques are readily applied by those skilled in the art to determine the location within the CCL structure at which these fragments bind, and such fragments can then be assembled by synthetic chemistry into larger compounds with increased affinity for CCL. Further, compounds identified or designed using the methods of the invention can be a peptide or a mimetic thereof. The isolated peptides or mimetics of the invention may be conformationally constrained molecules or alternatively molecules which are not conformationally constrained such as, for example, non-constrained peptide sequences. The term "conformationally constrained molecules" means conformationally constrained peptides and conformationally constrained peptide analogues and derivatives. In addition, the amino acids may be replaced with a variety of uncoded or modified amino acids such as the corresponding D-amino acid or N-methyl amino acid. Other modifications include substitution of hydroxyl, thiol, amino and carboxyl functional groups with chemically similar groups. With regard to peptides and mimetics thereof, still other examples of other unnatural amino acids or chemical amino acid analogues/derivatives can be introduced as a substitution or addition. Also, a peptidomimetic may be used. A peptidomimetic is a molecule that mimics the biological activity of a peptide but is no longer peptidic in chemical nature. By strict definition, a peptidomimetic is a molecule that no longer contains any peptide bonds (that is, amide bonds between amino acids). However, the term peptide mimetic is sometimes used to describe molecules that are no longer completely peptidic in nature, such as pseudo-peptides, semi-peptides and peptoids. Whether completely or partially non-peptide, peptidomimetics for use in the methods of the invention, and/or of the invention, provide a spatial arrangement of reactive chemical moieties that closely resembles the three-dimensional arrangement of active groups in the peptide on which the peptidomimetic is based. As a result of this similar active-site geometry, the peptidomimetic has effects on biological systems which are similar to the biological activity of the peptide. There are sometimes advantages for using a mimetic of a given peptide rather than the peptide itself, because peptides commonly exhibit two undesirable properties: (1) poor

bioavailability; and (2) short duration of action. Peptide mimetics offer an obvious route around these two major obstacles, since the molecules concerned are small enough to be both orally active and have a long duration of action. There are also considerable cost savings and improved patient compliance associated with peptide mimetics, since they can be administered orally compared with parenteral administration for peptides. Furthermore, peptide mimetics are generally cheaper to produce than peptides. Naturally, those skilled in the art will recognize that the design of a peptidomimetic may require slight structural alteration or adjustment of a chemical structure designed or identified using the methods of the invention. In general, chemical compounds identified or designed using the methods of the invention can be synthesized chemically and then tested for ability to modulate ACL activity, in particular CCL activity, using any of the methods described herein. The peptides or peptidomimetics of the present invention can be used in assays for screening for candidate compounds which bind to selected regions or selected conformations of CCL. Binding can be either by covalent or non-covalent interactions, or both. Examples of non-covalent interactions include electrostatic interactions, van der Waals interactions, hydrophobic interactions and hydrophilic interactions.

When a compound of the invention interacts with ACL, in particular interacts with the CCL module of ACL, it preferably "modulates" ACL activity. By "modulate" it is meant that the compound changes an activity of ACL by at least 50%. Suitably, a compound modulates ACL activity by increasing or decreasing the enzymatic activity of ACL, in particular CCL. The ability of a candidate compound to increase or decrease the activity of ACL, in particular CCL can be assessed by any one of the ACL, in particular CCL, enzymatic or cell-based assays known in the art (e.g. Granchi C (2018) *Eur. J. Med. Chem.* 5, 157, 1276 and Pinkosky SL *et al* (2017) *Trends Mol. Med.* 23(11): 1047), and in the materials and methods section below. Compounds of the present invention preferably have an affinity for ACL, in particular CCL, sufficient to provide adequate binding for the intended purpose. Suitably, such compounds have an affinity (K_d) of from 10^{-5} to 10^{-15} M. For use as a therapeutic, the compound suitably has an affinity (K_d) of from 10^{-7} to 10^{-15} M, preferably from 10^{-8} to 10^{-12} M and more preferably from 10^{-10} to 10^{-12} M. As will be evident to the skilled person, the crystal structure presented herein has enabled, for the first time, direct visualisation of the regions binding the natural ligands of ACL in the structure.

Screening Assays and Confirmation of Binding

Compounds of the invention may be subjected to further confirmation of binding to ACL, in particular CCL, by co-crystallization of the compound with ACL, in particular CCL and structural determination, as described herein. Compounds designed or selected according to the methods of the present invention are preferably assessed by a number of *in vitro* and *in vivo* assays of ACL, in particular CCL function to

confirm their ability to interact with and modulate ACL, in particular CCL activity. Libraries may be screened in solution by methods generally known in the art for determining whether ligands competitively bind at a common binding site. Such methods may include screening libraries in solution, or on beads or chips. Where the screening assay is a binding assay, ACL, in particular CCL, may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescent molecules, chemiluminescent molecules, enzymes, specific binding molecules, particles, e.g., magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin, etc. For the specific binding members, the complementary member would normally be labelled with a molecule that provides for detection, in accordance with known procedures. A variety of other reagents may be included in the screening assay. These include reagents like salts, neutral proteins, e.g., albumin, detergents, etc., which are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, antimicrobial agents, etc., may be used. The components are added in any order that produces the requisite binding. Incubations are performed at any temperature that facilitates optimal activity, typically between 4°C and 40°C. Direct binding of compounds to ACL, in particular CCL can also be done for example by Surface Plasmon Resonance (BIAcore).

It is to be understood that although particular embodiments, specific configurations as well as materials and/or molecules, have been discussed herein for cells and methods according to the present invention, various changes or modifications in form and detail may be made without departing from the scope and spirit of this invention. The following examples are provided to better illustrate particular embodiments, and they should not be considered limiting the application. The application is limited only by the claims.

Examples

Example 1

To facilitate high-resolution structural studies by X-ray crystallography, in the present invention an enzymatically active variant of hACL was produced, termed hACL-A/B, that lacked the linker region connecting the ancestral ACL-A and ACL-B parts (Fig. 1a, Fig. 5c-e). Ensuing crystal structures in space groups *P1* and *C2* for hACL-A/B, complexed with ADP, citrate, and CoA show that the CCL domains of four hACL chains form an intertwined, D2-symmetric 120 kDa CCL module that serves as the oligomerization platform of the ACL enzyme (Fig. 1b, Fig. 5 f,g and Table 1). Four CCSα/β modules are connected to this CCL module via a long linker that tethers the C-terminal end of CCSα to a two-helical stalk region protruding from the CCL module (Fig. 1a,b and Fig. 5g). The CCL module contains four citrate

synthase-like CoA-binding domains[20] that each grasp the adenosine head of a CoA molecule with the 3' phosphoryl group of CoA accommodated at the interface between the CoA-binding domains and a juxtaposed CCSα/β module. The phosphopantothenic arms of these CoA molecules bind across the nucleotide-binding motif of a juxtaposed CCSα-region and protrude into the CCS active site (Fig. 6a,b). At this site, the reactive thiol group of CoA reacts with ATP-derived citryl-phosphate to form citryl-CoA[21,22]. The observed arrangement of hACL-A/B is further stabilized by few contacts between the CCS module and the CCL CoA-binding domains, and by a β-hairpin motif in CCSα (residues 531-555) contacting the CCL stalk region (Fig. 1b). Superposition of the different hACL-A/B crystal structures displays a medley of CCS module conformations with respect to the tetrameric CCL core platform, providing evidence for the extensive structural plasticity of ACL tetramers around the CCL stalk region (Fig. 5h).

Example 2

In a next step we pursued crystallographic studies of the isolated CCL module of hACL at high resolution in complex with citrate and CoA and showed that the CCL protomers can adopt a closed state, whereby the CoA-binding domains undergo a ~10° rotation compared to the open state observed in the hACL-A/B crystal structures (Fig. 1c, Fig. 7a-c, Table 1). Furthermore, in the isolated CCL module the bound CoA molecules adopt a compact conformation wherein the phosphopantheine tail folds back across its adenosine head to reach the CCL citrate binding site at the junction between the CoA-binding domains and the helical bundle core of the CCL module (Fig. 7d). The rotation of the CoA-binding domain is necessary to fully form and shield the CCL active site, as this domain contributes residues His975, Arg986, and catalytic Asp1026 which bind citrate (Fig. 7 d, e). These structural analyses establish a set of large structural transitions underlying the catalytic itinerary of ACL. Following the ATP-driven formation of citryl-CoA at the CCS module, the citryl-thioester head of citryl-CoA swings back over a distance of ~35 Å to reach the citrate binding site of the CCL module (Fig. 1d). Furthermore, shuttling of citryl-CoA is associated with the rotation of the CoA-binding domain and closure of the CCL active site to allow cleavage of citryl-CoA at the heart of the ACL enzyme to yield the reaction products acetyl-CoA and oxaloacetate (Fig. 1e).

Example 3

Given the structural diversity of homo– and heteromeric ACL assemblies across different kingdoms of life we sought to determine representative crystal structures for heteromeric ACL-A/B[18]. Structures of heteromeric ACL-A/B from the methanogenic archaeon *Methanosaeta concilii*[23] and the green sulphur

bacteria *Chlorobium limicola*[7] reveal that prokaryotic ACL enzymes form heteromeric ACL-A/B assemblies similar to hACL-A/B and also feature CCSα β-hairpins interacting with two-helical stalks extending from the CCL module (Fig. 2, Fig. 6c,d, and Fig. 8, and Table 1). Moreover, similar to its counterpart in hACL the CCL module of *C. limicola* ACL can also adopt open and closed states (Fig. 7f-i and Table 1).

Example 4

In the ancestral variant of the rTCA cycle found in members of the early branching bacterial phylum *Aquificae,* the citrate-cleavage reaction is catalyzed by the tandem action of two distinct enzymes related to ACL: α/β heteromeric CCS and CCL[17,24] (Fig. 2). The structure of CCL from *H. thermophilus*[17] shows that stand-alone CCL adopts an intertwined tetrameric structure superimposing with the CCL module of hACL (rmsd = 2.2 Å for 658 aligned Cα-atoms) (Fig. 5i and Table 1). CCSα/β from *H. thermophilus*[25] is structurally similar to the CCS module of hACL (rmsd = 2.3 Å for 330 aligned Cα-atoms), although in *H. thermophilus* two CCS α/β modules associate to form an $\alpha_2\beta_2$-heterotetramer as seen for *E. coli* succinyl-CoA synthetase[26] (Fig. 6e,f and Table 1). Thus, the ACL enzyme arose from the fusion of a tetrameric CCL module with CCS, and by the acquisition of specific structural elements such as the two-helical stalk and interacting β-hairpins that underlie the conserved ACL architecture.

Example 5

Corroborating our structural findings, previous binding studies[22] also show that hACL displays four CoA binding sites. In contrast, the crystal structure of *C. limicola* ACL establishes that this bacterial homologue can bind eight CoA molecules: four in the CCL module and four in the CCS modules (Fig. 6d). Based on this fundamental difference in substrate binding, we propose that ACL from *C. limicola* may represent an ancestral form of ACL wherein the nucleotide binding motif of the CCS module is still capable of binding the adenosine moiety of CoA, whereas human and archaeal ACL display *bona fide* hybrid CoA-binding sites.

Example 6

Superposition of the human CCL module in the closed state with the hACL-A/B crystal structure showed that the 3'-phosphoryl moiety of CoA would clash with the nucleotide binding domain and β-hairpin motif of the juxtaposed CCSα region (Fig. 9a,b). This implies that the CCS modules need to reorient upon rotation of the CoA-binding domains and transfer of citryl-CoA. While the structure of the archaeal ACL-A/B enzyme is similar to the structure for hACL-A/B with its CCL module in an open state and four equivalent CoA molecules bridging the CCL and CCS modules (Fig. 6c), ACL-A/B from *C. limicola* adopts a

distinct structural state (Fig. 8, Fig. 9c). In the latter structure all four CCL active sites are closed, with the CCS modules reoriented when compared to the human and archaeal ACL structures (Fig. 3a,b). A comparison between these different functional states offers a plausible view of the large-scale structural transitions that the ACL enzyme might undergo upon transfer of the citryl-CoA intermediate. Thus, we infer that upon closure of the CCL module the CCS module reorients around the CCL module mediated by a pivoting axis defined by the interaction between the CCSα β hairpin and the CCL stalk region (Fig. 9d-f).

Example 7

In order to obtain additional insights into the structural plasticity of ACL we conducted in-solution small-angle X-ray scattering (SAXS) experiments on hACL-A/B and *C. limicola* ACL-A/B (Table 2). Whereas supplementing the SAXS measuring buffer with citrate alone did not result in a significant change in the scattering profile recorded from apo-ACL-A/B, addition of both citrate and CoA induces a structural transition towards the conformational states observed in the crystal structures (Fig. 3c,d and Fig. 9g,h). This shows that ACL enzymes have the capacity to oscillate between distinct structural states in a ligand-dependent manner, albeit with differing conformational amplitudes. Ensuing rigid-body modelling of apo-hACL-A/B suggests an increased structural plasticity around the two-helical stalk region in the absence of CoA (Fig. 9g). Additional SAXS data recorded for full-length hACL indicates a similar in-solution structure and ligand-induced rearrangements as compared to linker-deleted hACL-A/B (Fig. 9i,j).

Table 2: SAXS data-collection and scattering-derived parameters*

Data collection parameters	
Instrument	P12 beamline (Petra III synchrotron)
Experimental setup	SEC–SAXS
SEC column	Agilent Bio SEC-3 with 300 Å pore size
Loading concentration	17 mg/ml for *C. limicola* ACL 24 mg/mL for hACL-A/B 20 mg/mL for hACL
Injection volume (µL)	50
SEC flow rate (mL/min)	0.4
Wavelength	$\lambda = 0.124$ nm
Beam size	0.2×0.12 mm^2
Detector	Pilatus 6M
Detector distance (m)	3.1
s measurement range (nm^{-1})	0 – 5
Normalization	The data were normalized to the intensity of the transmitted beam
Exposure time per frame	1 s
Acquisition rate / exposure time	Continuous 1s data-frame measurements
Temperature (K)	293

Structural parameters

				Estimated molecular weight (kDa)				
Sample	***SEC buffer***	**R_g (nm) [from Guinier]**	**D_{max} (nm) [from P(r)]**	**Porod** **DAM**	**Mow**	**VC**		Theoretical molecular weight (kDa)
C. limicola ACL-A/B	HBS	6,12	20	415	401	419	452	430
	HBS + Citrate	6,03	20	419	419	415	449	
	HBS + Citrate + CoA	5,7	17,5	468	378	422	450	
hACL-A/B	HBS	6,04	17,5	461	445	489	477	462
	HBS + Citrate	6,05	17,5	467	434	497	508	
	HBS + Citrate + CoA	5,75	16,5	441	430	477	484	
hACL	HBS	6,08	19	478	463	504	492	493
	HBS + Citrate	6,16	19	492	554	522	515	
	HBS + Citrate + CoA	5,87	17	484	496	510	517	

Software employed	
Primary data reduction	CHROMIXS
Guinier analysis	AutoRg
P(r) calculation	GNOM 5
Particle exclude volume	DATPOROD
Molecular weight estimation	DATMW
Atomic structure modelling	Crysol, Crysol 3.0, FoXS, MultiFoXS, AllosMOD-FoXS

*where $s = 4\pi\sin\theta/\lambda$, and 2θ is the scattering angle

Example 8

CCL – as a stand-alone enzyme or as the core module of ACL - shows significant sequence identity to citrate synthase (CS), the first enzyme in the oxidative Krebs cycle. Prototypical CS is homodimeric and also cycles between open and closed states during catalysis[20,27]. Structural comparison of CCL with CS shows that the 20-helix core in tetrameric CCL superimposes well with the 20-helix core in dimeric CS (Fig. 4a,b and Fig. 10a). Ensuing molecular dissection of CS unravels that both the N-terminal and C-terminal half of the CS protomer show structural and sequence homology to a CCL protomer (Fig. 10b-g). This provides strong evidence that the CS protomer evolved from the fusion of two ancestral CCL protomers with the overall structure of the dimeric CS assembly remaining conserved with respect to tetrameric CCL. The N-terminal half of CS corresponds to a CCL protomer in which the CoA binding domain (Fig. 10d,g) is replaced by a short loop (Fig. 10h). The C-terminal CCL-homology region in CS encompasses the CoA-binding domain, citrate binding site and C-terminal tail wrapping around the adjacent protomer (Fig. 10e,g). This fusion event (Fig. 4c, Fig. 10i) and the introduced internal structural repeat in CS is also apparent from the pseudo-twofold symmetry axis present in the CS protomer which corresponds to one of the three twofold symmetry axes in D2-symmetrical CCL (Fig. 4d). Importantly, the molecular transition of CCL to CS indicates that the rTCA cycle, supporting autotrophy, predates the oxidative Krebs cycle or the recently described reversed oxidative Krebs cycle[28,29] (Fig. 4e).

Example 9

By analogy to the CS-catalyzed reversible condensation of oxaloacetate and acetyl-CoA to citryl-CoA, citryl-CoA cleavage by ACL is thought to proceed through a retro-aldol cleavage mechanism with the acetyl-CoA enolate as an intermediate[19,30]. Comparison of crystal structures at high-resolution for the substrate-bound CCL module of human and *C. limicola* ACL with CS shows an equivalent active site configuration indicating a similar catalytic mechanism (Fig. 4f-h, Fig. 9a-c and Table 1). In CS, the catalytic aspartate residue (Asp375 in chicken CS) initiates enolization of acetyl-CoA by abstracting a Cα methyl proton[20]. We propose that during the analogous retro-aldol cleavage of citryl-CoA by hACL, the corresponding Asp1026 protonates the acetyl-CoA enolate intermediate, while the main chain nitrogen of Gly936 is poised to polarize the carbonyl oxygen of the citryl-CoA thioester (Fig. 11d). Consistent with its catalytic role, we note that hACL-Asp1026 engages in a low-barrier hydrogen-bond[31] (distance = 2.4 Å) with the pro-S carboxylate of the bound citrate molecule (Fig. 4f). Additional studies, e.g. by neutron diffraction, will be needed to identify the base abstracting the hydroxyl proton of citryl-CoA to initiate retro-aldol cleavage (Fig. 11d). Proposals for the acid responsible for protonating the carbonyl oxygen of

oxaloacetate in the reverse reaction by CS have suggested His320 (ref. [20]), Arg329 (ref. [32]), as well as an alternative mechanism[33].

Materials and methods

Recombinant protein production and purification

Protein expression constructs generated in this study - as described below - are available via the GeneCorner Plasmid Collection at http://bccm.belspo.be through the associated LMBP accession number. cDNA encoding full-length human ACL (hACL, Uniprot ID P53396-2) was obtained from a liver cDNA databank and cloned into the pTrcHis2 vector, in frame with a C-terminal Myc- and His-tag, resulting in pTrcHis2-hACL (LMBP 11127). To produce a heteromeric form of human ACL (hACL-A/B) without the long linker region (residues 426-487) codon-optimized cDNA fragments encoding residues 1-425 (hACL-A) and residues 488-1101 (hACL-B) of human ACL (Uniprot ID P53396-1) were cloned into the pET-Duet vector, resulting in pET-DUET-hACL-A/B (LMBP 11131). The hACL-B fragment carried a C-terminal His-tag. Codon optimized cDNA fragments encoding the *M. concilii* ACL-A (NCBI ID WP_048131686.1) and ACL-B (NCBI ID WP_048131683.1) subunits, and the *H. thermophilus* CCSα (Uniprot ID Q75VW6) and CCSβ (Uniprot ID D3DK29) subunits were cloned in the pET-DUET vector, resulting in vectors pET-Duet-Mco-ACL-A/B (LMBP 11132) and pET-Duet-Hth-CCS (LMBP 11134), respectively, with the ACL-B and CCSβ fragments carrying a C-terminal His-tag. Codon optimized cDNA fragments encoding the *C. limicola* ACL-A (Uniprot ID Q9AQH6) and ACL-B (Uniprot ID Q9AJC4) subunits were cloned in the pET11a vector interspersed with an *E. coli* ribosomal binding sequence[34], resulting in the bicistronic construct, pET11a-Cli-ACL-A/B (LMBP 11125), with the ACL-B subunit carrying a C-terminal His-tag. A codon optimized cDNA fragment encoding *H. thermophilus* CCL (Uniprot ID Q75VX1) was cloned into the pET11a plasmid in frame with a C-terminal His-tag, resulting in pET11a-Hth-CCL (LMBP 11133). Codon-optimized cDNA fragments encoding the CCL core modules of human ACL (Uniprot ID P53396-1, residues 836-1101) and *C. limicola* ACL (Uniprot ID Q9AQH6, residues 351 - 608) were cloned into the pET15b vector, in frame with a thrombin-cleavable N-terminal His-tag, resulting in expression plasmids, pET15b-hCCL (LMBP 11128) and pET15b-Cli-CCL (LMBP 11129), respectively. pTrcHis2-hACL was expressed in the *E. coli* C43 (DE3) strain grown at 28° C. All other constructs were expressed in the *E. coli* BL21(DE3) strain. pET11a-Cli-ACL-A/B was expressed at 28° C and other constructs at 20° C. Cultures were grown in Luria-Bertani broth and expression was induced by addition of 1 mM IPTG at an optical density at 600 nm of 0.6 – 0.7. Following overnight expression cultures were harvested by centrifugation and bacterial cells resuspended in IMAC binding buffer: 50 mM sodium phosphate, pH 7.4 and 150 mM NaCl, supplemented with 1 mM DTT. To prevent proteolysis in the long linker region in full-

length hACL cOmplete Protease Inhibitor Cocktail without EDTA (Roche) was added. Bacterial cells were lysed by sonication and insoluble material was removed by centrifugation. The resulting supernatant was clarified using a 0.22 μm filter and loaded onto an cOmplete His-tag (Roche) or Ni Sepharose column equilibrated with IMAC binding buffer. The IMAC column was washed and recombinant His-tagged proteins were eluted with binding buffer supplemented with increasing concentrations of imidazole. Elution fractions containing the His-tagged protein of interest were pooled and concentrated using ultracentrifugation. Proteins were further purified by size-exclusion chromatography (SEC) using HiLoad 16/600 Superdex 200 and Superose 6 (Increase) columns. As a SEC running buffer 20 mM HEPES (pH 7.4), 150 mM NaCl supplemented with 1 mM DTT or, for the purification of ACL enzymes, 20 mM citrate (pH 6,0), 150 mM NaCl supplemented with 1 mM DTT was used. ACL enzymes were purified at 4° C. The N-terminal His-tag of the purified CCL core modules of human ACL and *C. limicola* ACL was removed using thrombin and following overnight incubation the digestion mixture was injected on a SEC column. Elution fractions corresponding to the protein of interest were either used immediately or stored at -80°C until further use.

Enzymatic activity assays

Initial reaction rates for hACL-A/B, hACL and the hACL-His760Ala mutant in function of ATP concentration were measured using the malate dehydrogenase coupled assay[35]. Assays were set up in a transparent Nunc 96-well plate using a total reaction volume of 250 μL. Oxidation of NADH was followed by measuring the absorbance at 340 nm in function of time at 25° C using a FLUOstar Omega microplate reader (BMG Labtech). The reaction buffer contained 20 mM Tris buffer pH 8.5, 20 mM citrate, 10 mM $MgCl_2$, 0.5 mM CoASH, 4 mM DTT, 0.2 mM NADH and 4 units/mL of malate dehydrogenase (Roche). The reaction was started by adding ATP. Initial velocity rates, as the average of 3 replicate measurements, were plotted in GraphPad Prism and fitted to a Michaelis-Menten equation to obtain the parameters Km and kcat.

Crystallization and structure determination

Purified protein samples were concentrated to 5 – 15 mg/mL. For co-crystallization experiments with ligands the protein samples were supplemented with 50 mM citrate and/or 10 mM CoASH and 50 mM Mg.ADP. Nanoliter-scale vapour diffusion crystallisation experiments were set up at 293 K or 277 K using commercially available sparse matrix crystals screens (Molecular Dimensions, Hampton Research) and using a Mosquito crystallisation robot (TTP Labtech). Promising hits were further optimized using gradient optimization in 24-well crystallization plates with drop sizes of 1 μL. These optimized conditions are listed in Table 1. Crystals were cryoprotected in a quick soak in mother liquid supplemented with the

respective cryo solution (Table 1) and ligands when appropriate. Crystals were cryo-cooled by direct plunging into liquid nitrogen. X-ray diffraction measurements were conducted in synchrotron radiation facilities PETRA III (beamlines P13, P14), SOLEIL (Proxima 2A), ESRF (ID23-1, ID23-2, ID30-B) and SLS (PXI, PXIII). All data were integrated and scaled using the XDS suite[36] and AIMLESS[37] and data quality was analyzed by Phenix.xtriage[38]. To solve the initial structure of ACL-A/B from *C. limicola* molecular replacement (MR) was performed with Phaser[39] using the structure of the N-terminal truncated human ACL as a search model[40]. After placement of four search CCS modules, and initial refinement the CCL core module was built manually in the remaining difference density using the homology with citrate synthase as a structural guide. Parrot[41] density modification and NCS averaging was used to improve the quality of initial electron density maps. Subsequent structures for the human and *M. concilii* ACL-A/B, human and bacterial CCL modules and *H. thermophilus* CCL were solved by MR using derived search models. For *H. thermophilus* CCSα/β, search models for the CCSα and CCSβ subunits were prepared from homologous succinyl-CoA synthetase structures (pdb 2yv2 for the CCSα subunit and pdb 3ufx for the CCSβ subunit). Model (re)building was performed in COOT[42] and individual coordinate and ADP refinement was performed in PHENIX[43] and autoBuster[44]. Model and map validation tools in COOT and the PHENIX suite, the CCP4 package[45] and the PDB_REDO server[46] were used throughout the work flow to guide improvement and to validate the quality of crystallographic models.

Small-angle X-ray scattering data collection and analysis

In-solution SAXS data for *C. limicola* ACL-A/B, linker-deleted hACL-A/B and full-length hACL were measured on the P12 beamline of EMBL at the Petra III storage ring (DESY, Hamburg). 50 µl of protein sample (~20 mg/mL) in HBS (20 mM HEPES, 150 mM NaCl, pH 7.2) was injected onto an Agilent 4.6 x 300 mm Bio SEC-3 column with 300 Å pore size, and with HBS as running buffer at a flow speed of 0.4 ml min^{-1} at 20 °C. The scattering data were collected in continuous flow mode with 1s exposure time per frame. To collect SAXS data in the presence of citrate, or both citrate and CoASH, the SEC-SAXS buffer was supplemented with 50 mM citrate pH 7.2, or both 50 mM citrate pH 7.2 and 2 mM CoASH, and prior to injection ACL samples were incubated with citrate alone, or both citrate and CoASH. The program CHROMIXS[47] was used to select the buffer and sample frames from the collected SEC-SAXS data. Overall parameters were calculated using the ATSAS suite version 2.8 (ref. [48]). Calculation of theoretical scattering curves and fitting to experimental scattering data was performed with Crysol, Crysol 3.0 and FoXS. Rigid-body modelling of in-solution scattering data for linker-deleted hACL-A/B was performed using MultiFoXS[49] using a crystal structure for hACL-A/B in space group *C*2 as starting model. The presented two-state SAXS-model for apo-hACL-A/B in HBS (Chi^2= 2.8, and with w_1 = 0.58 and w_2 = 0.42) was obtained by defining the CCS modules and central CCL module as rigid bodies, and with residues 808

to 810 and residues 821 to 824 in each CCS-CCL linker region defined as flexible. The presented single-state SAXS-model for hACL-A/B in HBS supplemented both citrate and CoASH (Chi^2 = 2.8) was obtained by defining the CCS modules and central CCL module as rigid bodies with residues 808 to 811 in each CCS-CCL linker region defined as flexible. This SAXS-model for hACL-A/B was then used to model full-length hACL in HBS supplemented with both citrate and CoASH using AllosMOD-FoXS[50], resulting in a Chi^2-value of 5.5. Initial SAXS data on *C. limicola* ACL-A/B in HBS buffer without substrates was measured in SEC-SAXS mode on the SWING beamline at the SOLEIL Synchrotron (Gif-sur-Yvette, France) and analyzed in Foxtrot (developed at Synchrotron SOLEIL and provided by Xenocs, Sassenage, France).

Negative stain EM

3 µl of purified full-length hACL supplemented with 2 mM coenzyme A was applied to the clean side of carbon on a carbon–mica interface and stained with 2% w/v sodium silicotungstate (SST). Images were recorded on a FEI Tecnai T12 microscope operated at 120 kV with a Gatan Orius 1000 camera, at a nominal magnification of 29,000x, corresponding to a pixel size of 2.0 Å at the object scale. Semi-automatic particle selection on a few micrographs using BOXER[51] with a box size of 180 pixels resulted in an initial dataset containing 9881 particles. Subsequent image analysis was performed in RELION2.1[69]. After CTF estimation followed by 2D reference- free classification, a set of classes representing different views of hACL was used as an input for automated particle picking, resulting in a dataset of 25832 particles. Following 2D reference-free classification, particles from the best 6 classes were used to make an initial model with applied C2 symmetry. A subsequent 3D refinement was performed by using the initial model as an input. The resulting refined model was used to re-pick particles employing the fast projection matching (FPM) method[70,71], resulting in an extended dataset of 44817 particles that was further processed in RELION2.1. After two rounds of 2D reference-free classification and cleaning, a final dataset of 27293 particles was subjected to 3D classification using C2 symmetry. Out of the four resulting classes, 2 showed highly similar features and together accounted for 57.1 % of all particles (15617). A final 3D refinement was performed on the particles within these two 3D classes, using an averaged 3D model generated from their corresponding maps in Chimera[72]. This resulted in a final 3D map with a resolution of 26.6 Å according to the FSC = 0.143 criterion.

Multi-angle laser light scattering

hACL-A/B (100 µL) was injected onto a Superdex 200 Increase 10/300 GL column (GE Healthcare), with 20 mM citrate pH 7.4 and 150 mM NaCl as running buffer at 0.5 ml min^{-1}, coupled to an online UV-detector (Shimadzu), a multi-angle light scattering miniDAWN TREOS instrument (Wyatt) and a Optilab T-rEX refractometer (Wyatt) at 25 °C. A refractive index increment (dn/dc) value of 0.185 ml g^{-1} was used

for protein concentration and molecular mass determination. Data were analyzed using the ASTRA6 software (Wyatt). Correction for band broadening was applied using parameters derived from BSA (2 mg/mL, Pierce) injected under identical running conditions.

Structure and sequence analysis

Structures were superimposed with Chimera[52]. Sequence alignments were created using Clustal Omega[53] and formatted with ESPript[54]. Secondary structures elements of crystallographic structures were assigned with DSSP[55,56]. Figures containing structural models were prepared in PyMOL[57].

References

1 Srere, P. A. The citrate cleavage enzyme. I. Distribution and purification. *J Biol Chem* 234, 2544-2547 (1959).

2 Chypre, M., Zaidi, N. & Smans, K. ATP-citrate lyase: a mini-review. *Biochemical and biophysical research communications* 422, 1-4 (2012).

3 Sun, J. *et al.* BNIP-H recruits the cholinergic machinery to neurite terminals to promote acetylcholine signaling and neuritogenesis. *Developmental cell* 34, 555-568 (2015).

4 Wellen, K. E. *et al.* ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* 324, 1076-1080 (2009).

5 Sivanand, S. *et al.* Nuclear Acetyl-CoA Production by ACLY Promotes Homologous Recombination. *Molecular Cell* 67, 252-265. e256 (2017).

6 Hügler, M. & Sievert, S. M. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annual Review of Marine Science* 3, 261-289 (2011).

7 Kanao, T., Fukui, T., Atomi, H. & Imanaka, T. ATP-citrate lyase from the green sulfur bacterium Chlorobium limicola is a heteromeric enzyme composed of two distinct gene products. *The FEBS Journal* 268, 1670-1678 (2001).

8 Beigneux, A. P. *et al.* ATP-citrate lyase deficiency in the mouse. *Journal of Biological Chemistry* 279, 9557-9564 (2004).

9 Lin, R. *et al.* Acetylation stabilizes ATP-citrate lyase to promote lipid biosynthesis and tumor growth. *Molecular cell* 51, 506-518 (2013).

10 Potapova, I. A., El-Maghrabi, M. R., Doronin, S. V. & Benjamin, W. B. Phosphorylation of recombinant human ATP:citrate lyase by cAMP-dependent protein kinase abolishes homotropic allosteric regulation of the enzyme by citrate and increases the enzyme activity. Allosteric activation of ATP:citrate lyase by phosphorylated sugars. *Biochemistry* 39, 1169-1179 (2000).

11 Granchi, C. ATP citrate lyase (ACLY) inhibitors: An anti-cancer strategy at the crossroads of glucose and lipid metabolism. *Eur J Med Chem* 157, 1276-1291 (2018).

12 Hatzivassiliou, G. *et al.* ATP citrate lyase inhibition can suppress tumor cell growth. *Cancer cell* 8, 311-321 (2005).

13 Migita, T. *et al.* ATP citrate lyase: Activation and therapeutic implications in non–small cell lung cancer. *Cancer research* 68, 8547-8554 (2008).

14 Zaidi, N., Swinnen, J. V. & Smans, K. ATP-citrate lyase: a key player in cancer metabolism. *Cancer research* 72, 3709-3714 (2012).

15 Pinkosky, S. L. *et al.* Liver-specific ATP-citrate lyase inhibition by bempedoic acid decreases LDL-C and attenuates atherosclerosis. *Nature communications* 7, 13457 (2016).

16 Pinkosky, S. L., Groot, P. H. E., Lalwani, N. D. & Steinberg, G. R. Targeting ATP-Citrate Lyase in Hyperlipidemia and Metabolic Disorders. *Trends Mol Med* 23, 1047-1063 (2017).

17 Aoshima, M., Ishii, M. & Igarashi, Y. A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. *Molecular Microbiology* 52, 763-770 (2004).

18 Fatland, B. L. *et al.* Molecular characterization of a heteromeric ATP-citrate lyase that generates cytosolic acetyl-coenzyme A in Arabidopsis. *Plant Physiology* 130, 740-756 (2002).

19 Fan, F. *et al.* On the catalytic mechanism of human ATP citrate lyase. *Biochemistry* 51, 5198-5211 (2012).

20 Remington, S. J. Structure and mechanism of citrate synthase. *Curr Top Cell Regul* 33, 209-229 (1992).

21 Hu, J., Komakula, A. & Fraser, M. E. Binding of hydroxycitrate to human ATP-citrate lyase. *Acta Crystallographica Section D: Structural Biology* 73, 660-671 (2017).

22 Sun, T., Hayakawa, K., Bateman, K. S. & Fraser, M. E. Identification of the citrate-binding site of human ATP-citrate lyase using X-ray crystallography. *Journal of Biological Chemistry* 285, 27418-27428 (2010).

23 Barber, R. D. *et al.* Complete genome sequence of Methanosaeta concilii, a specialist in aceticlastic methanogenesis. *J Bacteriol* 193, 3668-3669, (2011).

24 Hügler, M., Huber, H., Molyneaux, S. J., Vetriani, C. & Sievert, S. M. Autotrophic CO2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. *Environmental Microbiology* 9, 81-92 (2007).

25 Aoshima, M., Ishii, M. & Igarashi, Y. A novel enzyme, citryl-CoA synthetase, catalysing the first step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. *Molecular microbiology* 52, 751-761 (2004).

26 Bailey, D. L., Fraser, M. E., Bridger, W. A., James, M. N. & Wolodko, W. T. A dimeric form of Escherichia coli succinyl-CoA synthetase produced by site-directed mutagenesis1. *Journal of molecular biology* 285, 1655-1666 (1999).

27 Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiosus at 1.9 Å resolution. *Biochemistry* 36, 9983-9994 (1997).

28 Nunoura, T. *et al.* A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* 359, 559-563 (2018).

29 Mall, A. *et al.* Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* 359, 563-567 (2018).

30 Das, N. & Srere, P. A. Exchange of methyl protons of acetyl coenzyme A catalyzed by adenosine triphosphate citrate lyase. *Biochemistry* 11, 1534-1537 (1972).

31 Usher, K. C., Remington, S. J., Martin, D. P. & Drueckhammer, D. G. A very short hydrogen bond provides only moderate stabilization of an enzyme-inhibitor complex of citrate synthase. *Biochemistry* 33, 7753-7759 (1994).

32 van der Kamp, M. W., Perruccio, F. & Mulholland, A. J. High-level QM/MM modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. *Chemical communications*, 1874-1876 (2008).

33 Aleksandrov, A., Zvereva, E. & Field, M. The mechanism of citryl-coenzyme A formation catalyzed by citrate synthase. *The journal of physical chemistry B* 118, 4505-4513 (2014).

34 Kanao, T., Fukui, T., Atomi, H. & Imanaka, T. ATP-citrate lyase from the green sulfur bacterium Chlorobium limicola is a heteromeric enzyme composed of two distinct gene products. *The FEBS Journal* 268, 1670-1678 (2001).

35 Linn, T. C. & Srere, P. A. Identification of ATP citrate lyase as a phosphoprotein. *J Biol Chem* 254, 1691-1698 (1979).

36 Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* 66, 125-132, (2010).

37 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography* 67, 282-292 (2011).

38 Zwart, P., Grosse-Kunstleve, R. & Adams, P. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsl* 43, 27-35 (2005).

39 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* 40, 658-674 (2007).

40 Sun, T., Hayakawa, K., Bateman, K. S. & Fraser, M. E. Identification of the citrate-binding site of human ATP-citrate lyase using X-ray crystallography. *Journal of Biological Chemistry* 285, 27418-27428 (2010).

41 Zhang, K. Y., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods in enzymology* 277 53-64 (1997).

42 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 66, 486-501 (2010).

43 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography* 68, 352-367 (2012).

44 Bricogne G., B. E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, S. A., Smart O.S., Vonrhein C., Womack T.O. . BUSTER version 2.10.3 *Cambridge, United Kingdom: Global Phasing Ltd.* (2017).

45 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* 67, 235-242 (2011).

46 Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* 1, 213-220 (2014).

47 Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* 34, 1944-1946 (2017).

48 Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography* 50, 1212-1225 (2017).

49 Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* 44, W424-429, (2016).

50 Guttman, M., Weinkam, P., Sali, A. & Lee, K. K. All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. *Structure* 21, 321-331, (2013).

51 Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128, 82-97, (1999).

52 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612, (2004).

53 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7, 539, (2011).

54 Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31, 3320-3323 (2003).

55 Kabsch, W. & Sander, C. Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577-2637, (1983).

56 Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* 43, D364-D368, (2015).

57 Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.7.

Claims

1. A composition in crystalline form comprising a citryl-CoA lyase (CCL) and its natural ligand selected from citrate or citrate and coenzyme A (CoA), wherein CCL is a module of ATP citrate lyase, and which module has an amino acid sequence with at least 90% identity to amino acid residues 836 to 1101 depicted in SEQ ID NO: 1, characterized in that the crystal is:
   i) a crystal between CCL and citrate in the space group P212121, with the following crystal lattice constants: a=67.34 Å ± 5%, b=123.89 Å ± 5%, c=139.80 Å ± 5%, α=β=γ=90°, or
   ii) a crystal between CCL, citrate and CoA in the space group P21, with the following crystal lattice constants: a=67.71 Å ± 5%, b=147.90 Å ± 5%, c=113.69 Å ± 5%, α=90°, β=105.28° ± 5%, γ=90°, or
   iii) a crystal between CCL, citrate and CoA in the space group C2221, with the following crystal lattice constants: a=82.83 Å ± 5%, b=111.19 Å ± 5%, c=143.37 Å ± 5%, α=β=γ=90°.

2. The composition of claim 1 which has a three-dimensional structure wherein the crystal i) comprises an atomic structure characterized by the coordinates of the PDB entry 6HXK or a subset of atomic coordinates of the PDB entry 6HXK or wherein the crystal ii) comprises an atomic structure characterized by the coordinates of the PDB entry 6HXL or a subset of atomic coordinates of the PDB entry 6HXL or wherein the crystal iii) comprises an atomic structure characterized by the coordinates of the PDB entry 6HXM or a subset of atomic coordinates of the PDB entry 6HXM.

3. A catalytic pocket, consisting of a subset of atomic coordinates, present in the crystals i), ii) or iii) as defined in claims 1 and 2 wherein said catalytic pocket consists of the amino acid residues His_{900}, Val_{904}, Phe_{935}, Gly_{936}, Lys_{964}, Leu_{969}, Ile_{970}, Met_{971}, Gly_{972}, Ile_{973}, Gly_{974}, His_{975}, Arg_{986}, Lys_{1018}, Leu_{1021}, Asn_{1020}, Asn_{1024}, Val_{1025}, Asp_{1026}, Arg_{1065} and Arg_{1085} as depicted in SEQ ID NO : 1 and wherein said amino acid residues represent the citrate and CoA binding sites.

4. A crystallizable composition comprising a mixture of
   i. a solubilized CCL and its natural ligand citrate or its natural ligands citrate and CoA, of which the amino acid sequence of CCL has an amino acid sequence with at least 90% identity to amino acid residues 836 to 1101 depicted in SEQ ID NO: 1, and
   ii. a mother liquor solution, wherein the mother liquor solution comprises about 0.2M Na_2SO_4 pH6.7 and about 20% w/v PEG_{3350} to obtain a crystal i) as defined in claims 1 or 2, or wherein the mother liquor solution comprises about 21% w/v PEG_{8000}, about 0.1M HEPES pH7.0 and

about 0.15M ammonium sulphate to obtain a crystal ii) as defined in claims 1 or 2, or wherein the mother liquor solution comprises about 0.1M Na/K phosphate pH5.5, about 0.1 M RbCl, about 0.1 M Na-citrate pH 5.5 and about 25% v/v PEG Smear medium to obtain a crystal iii) as defined in claims 1 or 2.

5. A method of obtaining a composition of any one of claims 1 to 4 comprising the steps of:
   i. producing and purifying a citryl-CoA lyase which has an amino acid sequence with at least 90 % identity to SEQ ID NO: 1 and,
   ii. mixing with a mother liquor solution and citrate or citrate and CoA, wherein the mother liquor solution comprises about 0.2M Na_2SO_4 pH6.7 and about 20% w/v PEG_{3350} and citrate to obtain a crystal i) as defined in claims 1 or 2, or wherein the mother liquor solution comprises about 21% w/v PEG_{8000}, about 0.1M HEPES pH7.0 and about 0.15M ammonium sulphate and citrate and CoA to obtain a crystal ii) as defined in claims 1 or 2, or wherein the mother liquor solution comprises about 0.1M Na/K phosphate pH5.5, about 0.1 M RbCl, about 0.1 M Na-citrate pH5.5 and about 25% v/v PEG Smear medium and citrate and CoA to obtain a crystal iii) as defined in claims 1 or 2.

6. A computer-assisted method of identifying, designing or screening for a compound that can potentially interact with a crystal of CCL selected from a crystal i), ii) or iii) as defined in claims 1 or 2, comprising performing structure-based identification, design or screening of a compound based on the compound's interactions with a structure defined by the atomic coordinates as defined in claim 2.

7. A method for identifying a compound that can bind to CCL, comprising dipping candidate small molecule compounds with CCL, or with CCL and citrate, or with CCL, citrate and CoA and allowing co-crystallization, and screening candidate ligands or antagonists by using a method for measuring intermolecular interaction and comparing, designing and docking the 3D structures i), ii) or iii) as defined in claim 2 and a candidate ligand by computer modeling.

8. A method of identifying, designing or screening for a compound that can interact with CCL, including performing structure-based identification, design, or screening of a compound based on the compound's interactions with a CCL structure or a subset thereof.

9. A method for identifying an agonist or antagonist compound of ATP citrate lyase comprising an entity selected from the group consisting of an antibody, a peptide, a non-peptide molecule and a chemical compound, wherein said compound is capable of enhancing, eliciting or blocking biological activity of ATP citrate lyase resulting from an interaction with CCL wherein said process includes:
    i. introducing into suitable computer program parameters defining an interacting surface based on the conformation of CCL corresponding to the atomic coordinates of the PDB entries 6HXK, 6HXL and 6HXM or a subset of said atomic coordinates, wherein said program displays a three-dimensional model of the interacting surface,
    ii. creating a three-dimensional structure of a test compound in said computer program;
    iii. displaying a superimposing model of said test compound on the three-dimensional model of the interacting surface;
    iv. and assessing whether said test compound model fits spatially into a binding site.

10. A screening method for identifying molecules interacting with CCL comprising:
    i. soaking a multiplicity of molecules with the crystals i), ii) or iii) as defined in claims 1 or 2,
    ii. determining the co-crystal structure with at least one molecule binding to the crystal i), ii) or iii).

11. A screening method for identifying molecules interacting with CCL comprising:
    i. soaking a multiplicity of molecules with crystals containing CCL or a CCL fragment in the crystallographic asymmetric unit, where CCL or the CCL fragment has an RMSD-value smaller than 1.5 Å as compared to the set of atomic coordinates, or part of the set of atomic coordinates, defined in appendices I, II, and III, and whereby CCL, or the CCL fragment, has a level sequence identity with SEQ 2 of 90% or higher,
    ii. determining the co-crystal structure with at least one molecule binding to CCL or a CCL fragment, where CCL or the CCL fragment has RMSD-value smaller than 1.5 Å as compared to the set of atomic coordinates, or part of the set of atomic coordinates, defined in appendices I, II, and III, and whereby CCL, or the CCL fragment, has a level sequence identity with SEQ 2 of 90% or higher.

Figure 1

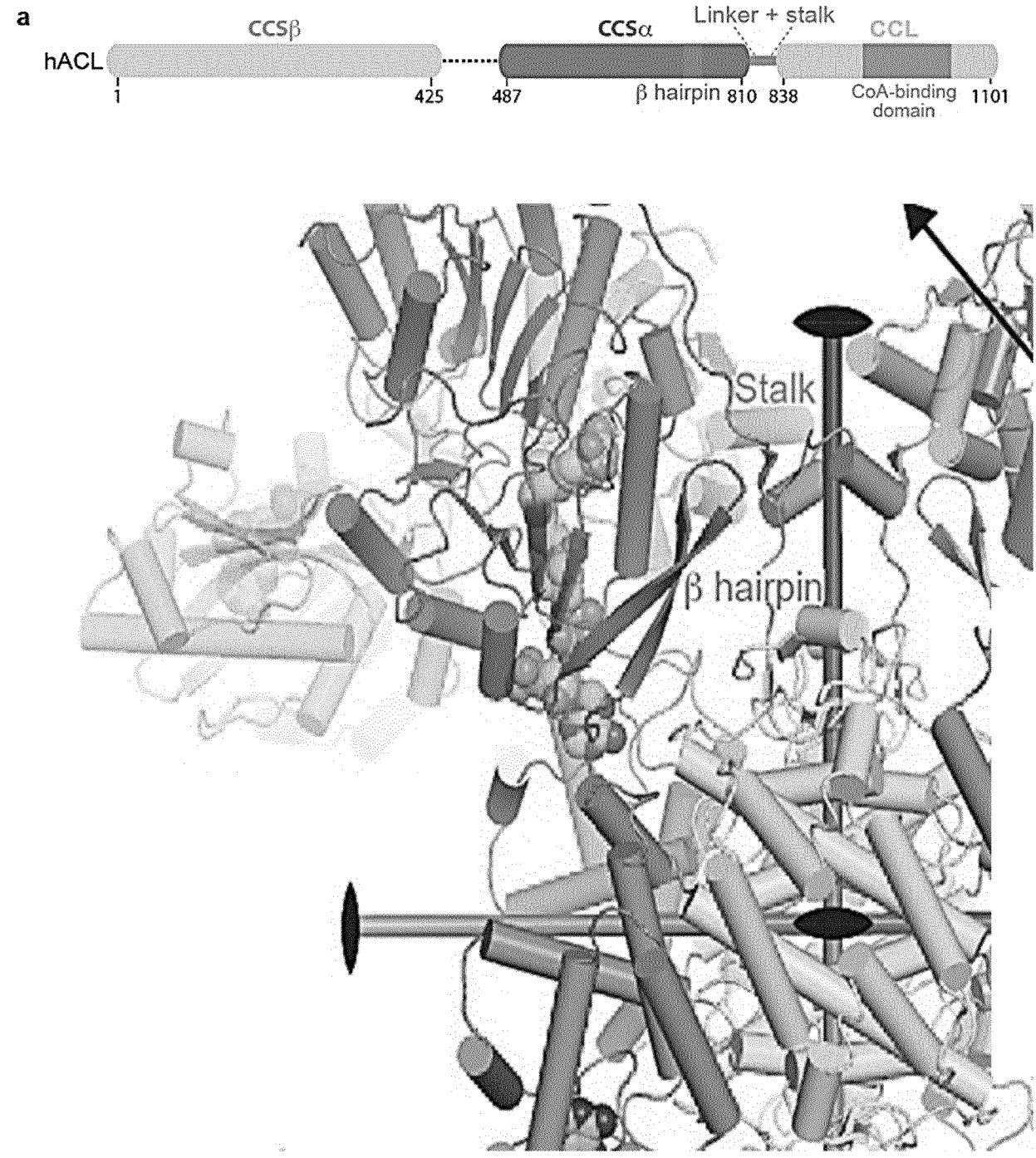
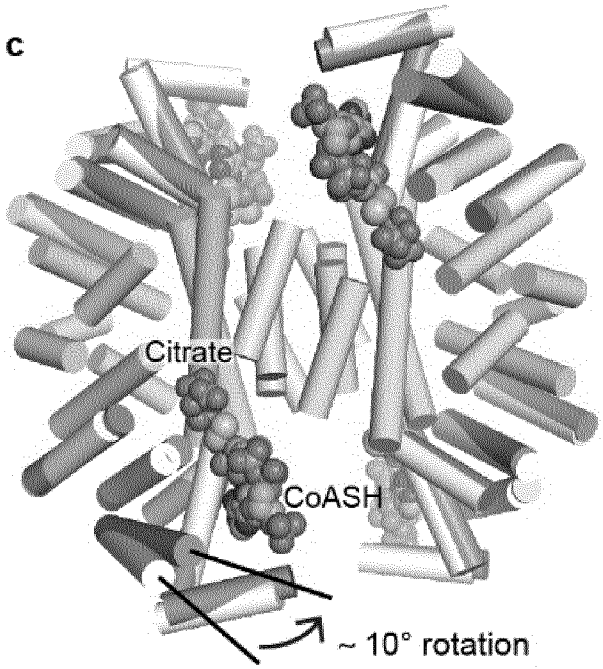

a
CCSβ
CCSα
Linker + stalk
CCL
hACL
1
425
487
β hairpin
810
838
CoA-binding domain
1101
Stalk
β hairpin

Figure 1 continued



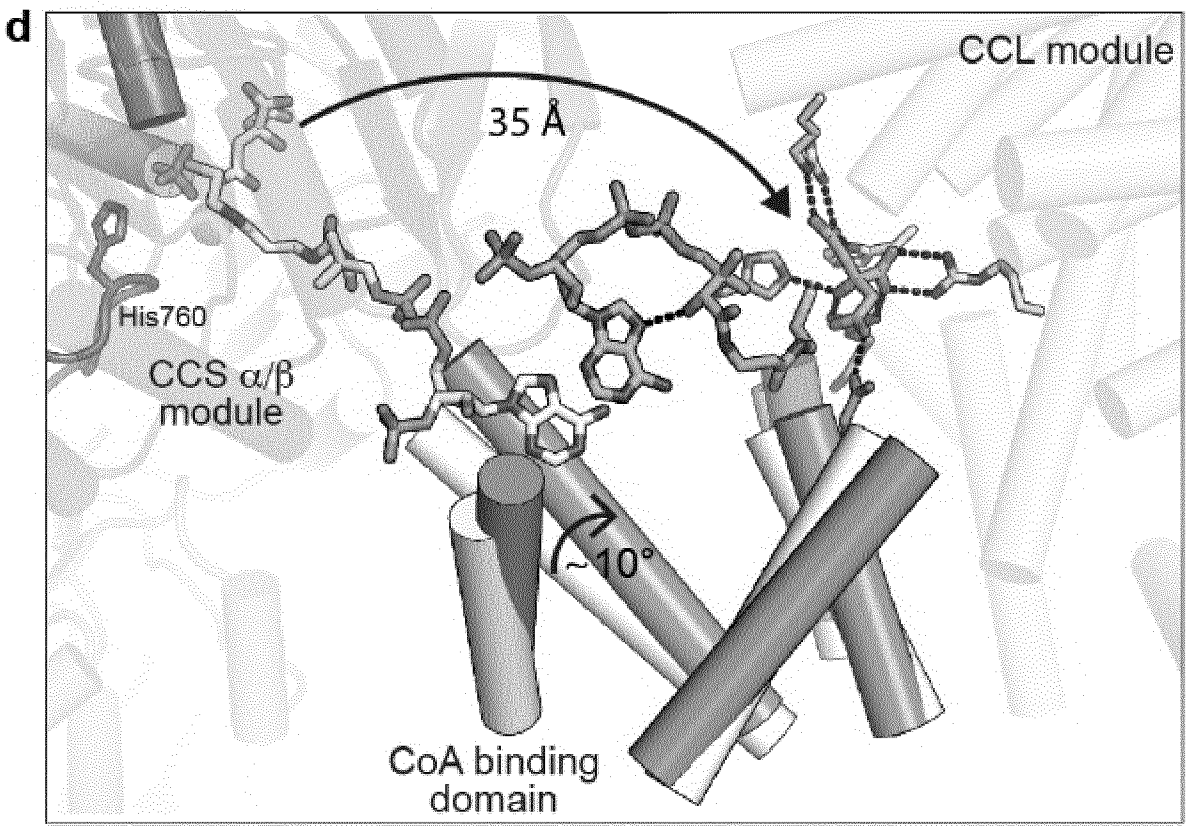
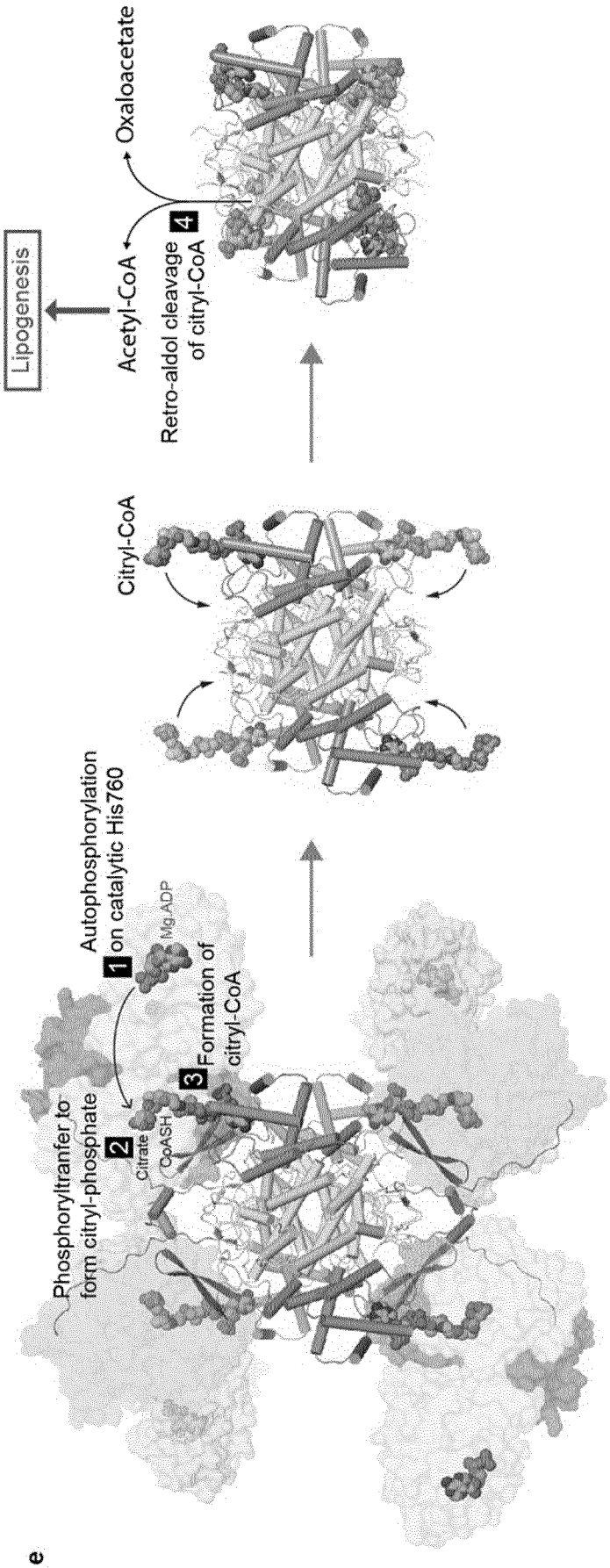


Figure 1 continued



e

Figure 2

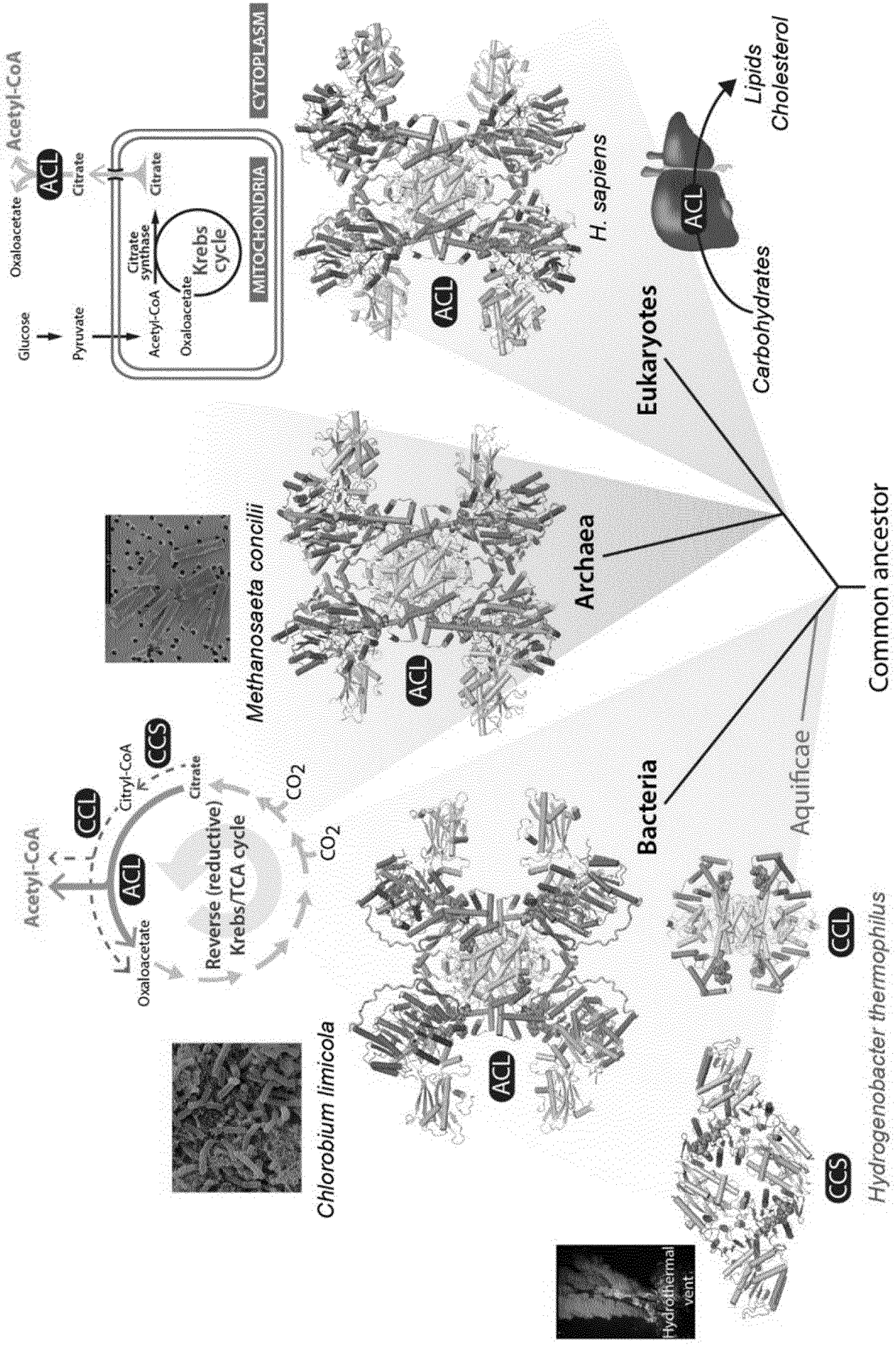

Glucose
Pyruvate
Acetyl-CoA
Oxaloacetate
ACL
Citrate
Citrate synthase
Krebs cycle
CYTOPLASM
MITOCHONDRIA
H. sapiens
Eukaryotes
Lipids
Cholesterol
Carbohydrates
Methanosaeta concilii
Archaea
Bacteria
Common ancestor
Aquificae
Chlorobium limicola
Hydrogenobacter thermophilus
Reverse (reductive) Krebs/TCA cycle
Citryl-CoA
CCL
CCS
CO_2
Hydrothermal vent

Figure 3

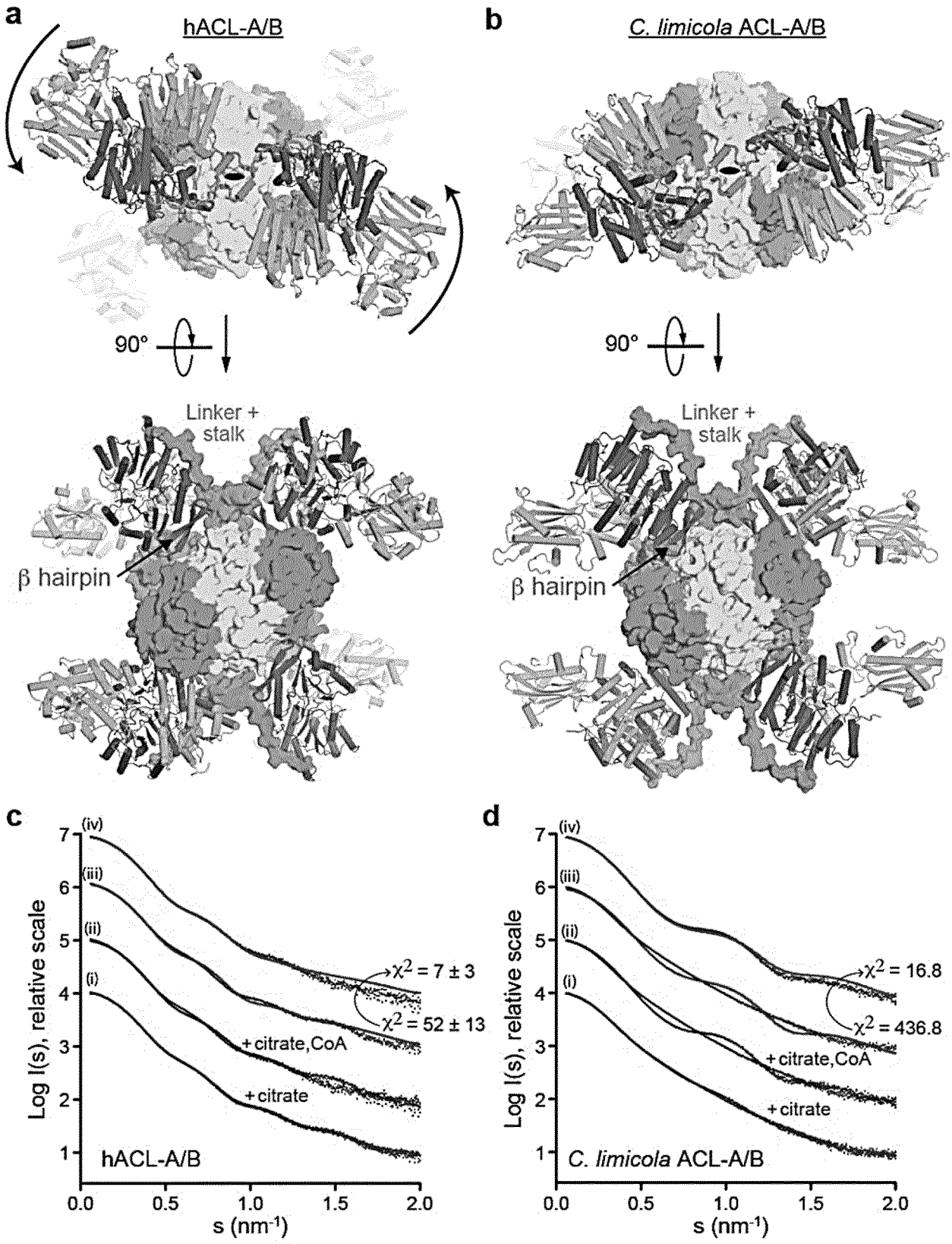

a
hACL-A/B
90°
Linker + stalk
β hairpin
b
C. limicola ACL-A/B
90°
Linker + stalk
β hairpin
c
Log I(s), relative scale
(iv)
(iii)
(ii)
(i)
χ2 = 7 ± 3
χ2 = 52 ± 13
+citrate,CoA
+citrate
hACL-A/B
s (nm-1)
d
Log I(s), relative scale
(iv)
(iii)
(ii)
(i)
χ2 = 16.8
χ2 = 436.8
+citrate,CoA
+citrate
C. limicola ACL-A/B
s (nm-1)

Figure 4

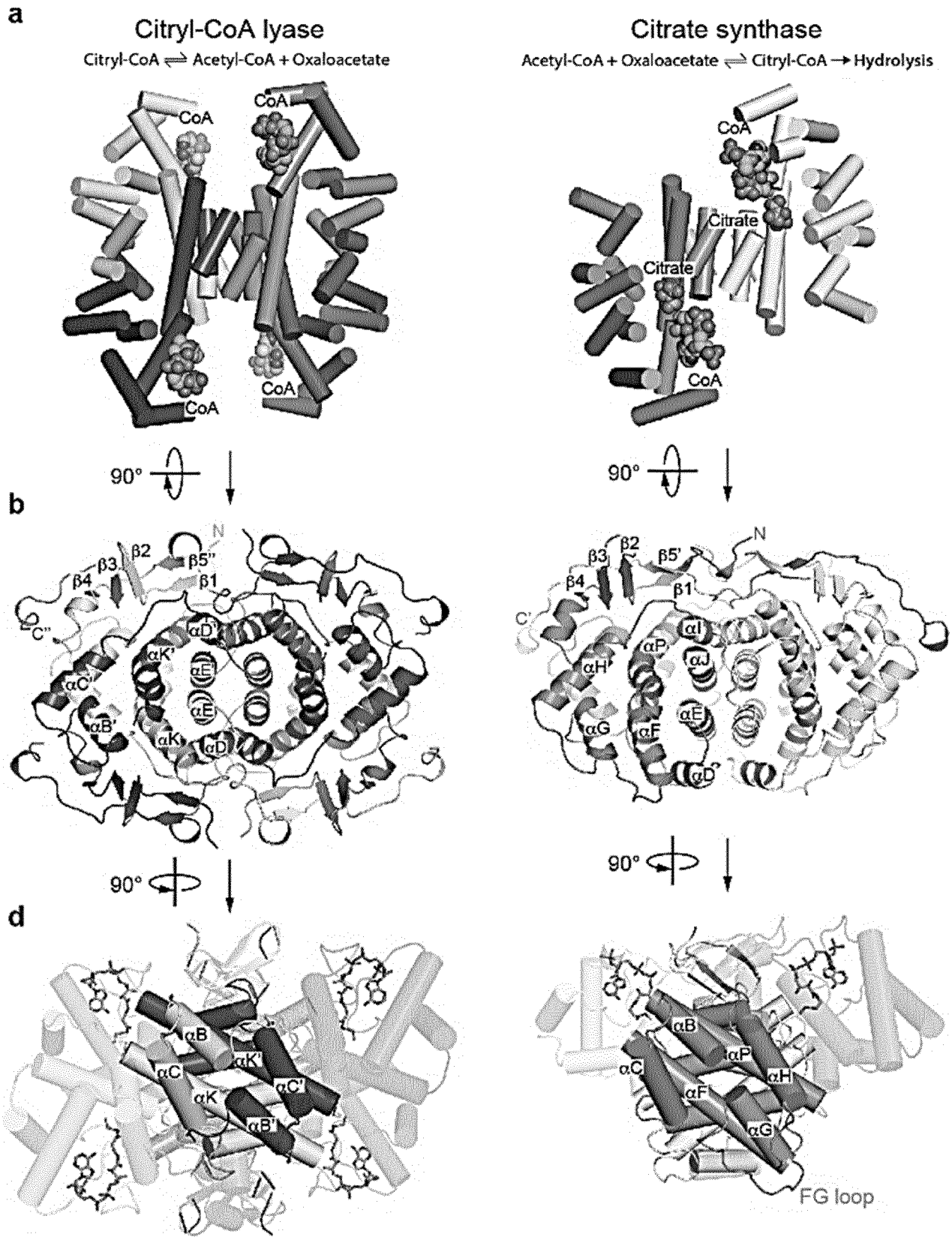
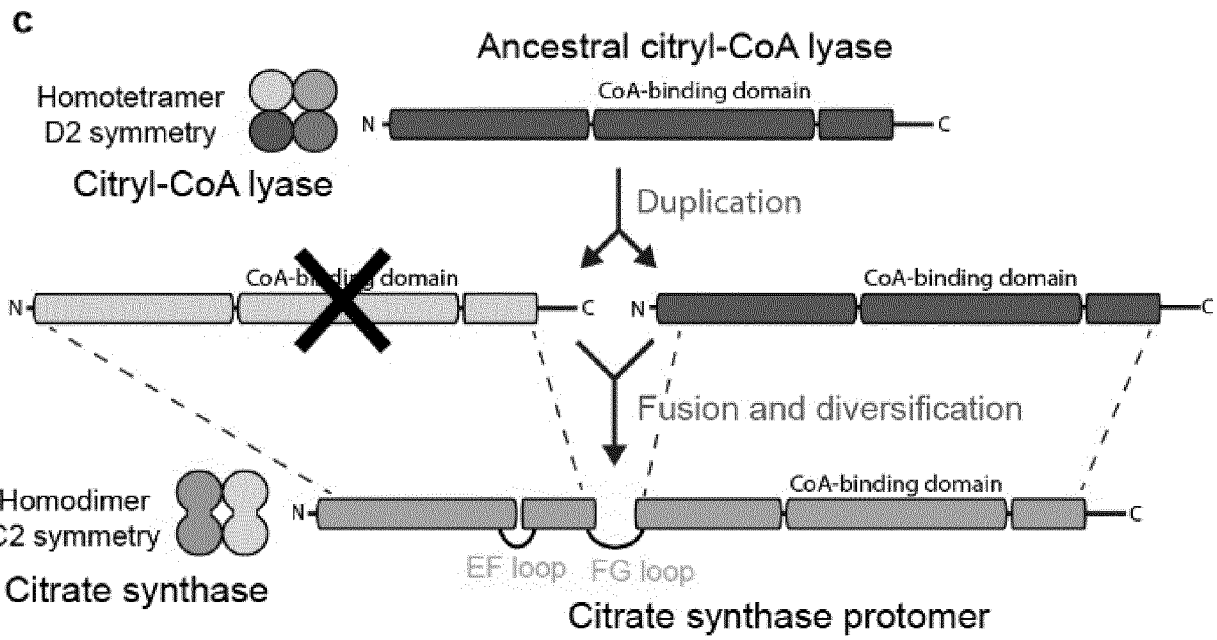

a
Citryl-CoA lyase
Citryl-CoA ⇌ Acetyl-CoA + Oxaloacetate
Citrate synthase
Acetyl-CoA + Oxaloacetate ⇌ Citryl-CoA → Hydrolysis
CoA
Citrate
90°
b
N
C"
C'
β1
β2
β3
β4
β5"
β5'
αB'
αC'
αD'
αE'
αK'
αB
αC
αD
αE
αF
αG
αH
αI
αJ
αK
αP
d
FG loop

Figure 4 continued



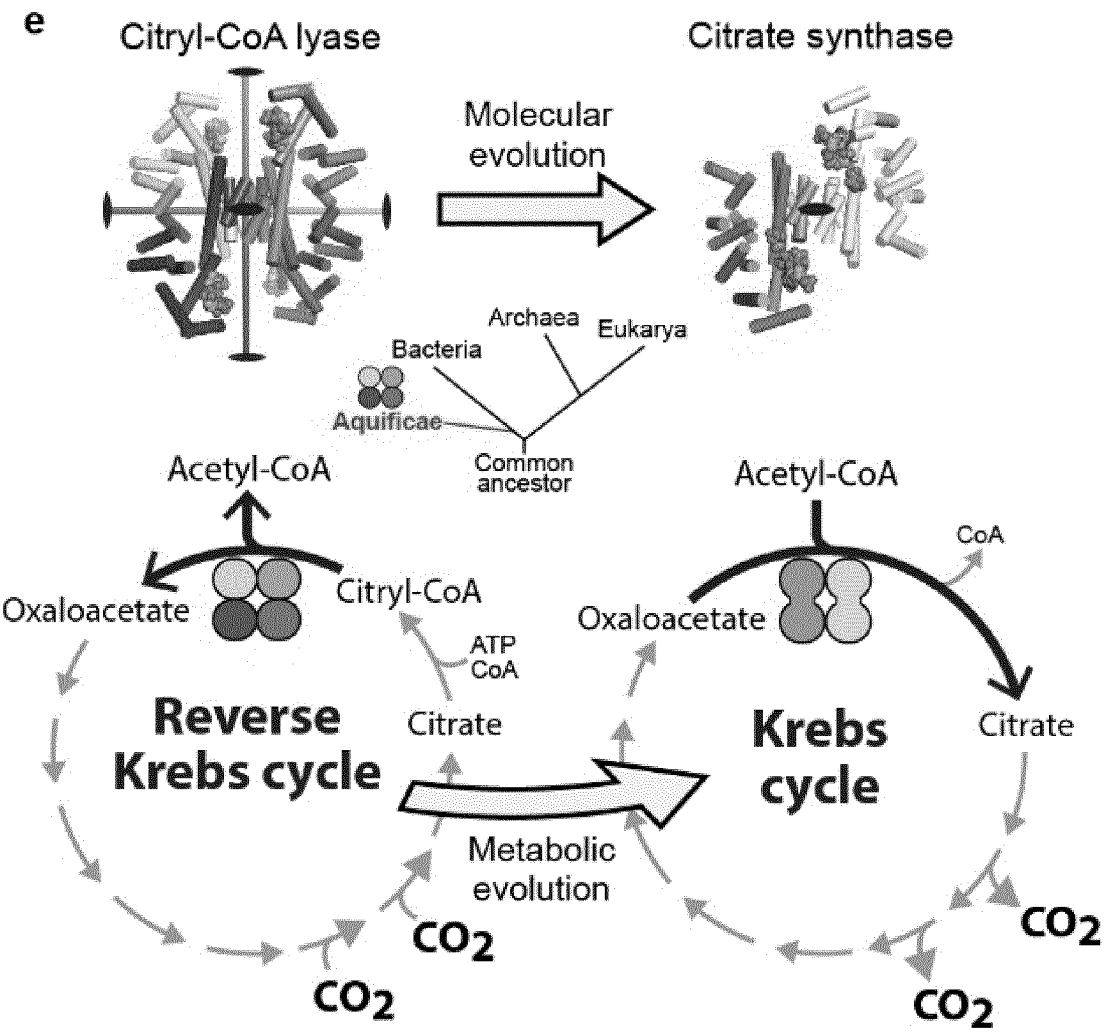


Figure 4 continued

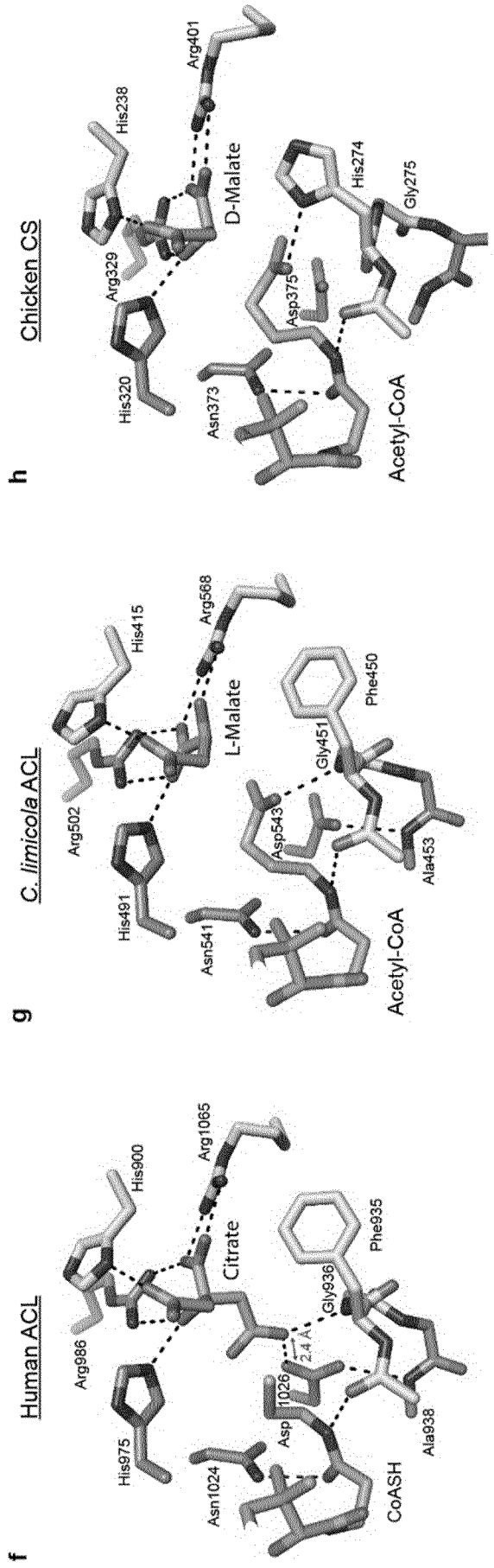


Figure 5

a

ATP + E → E-P + ADP
REACTION 1

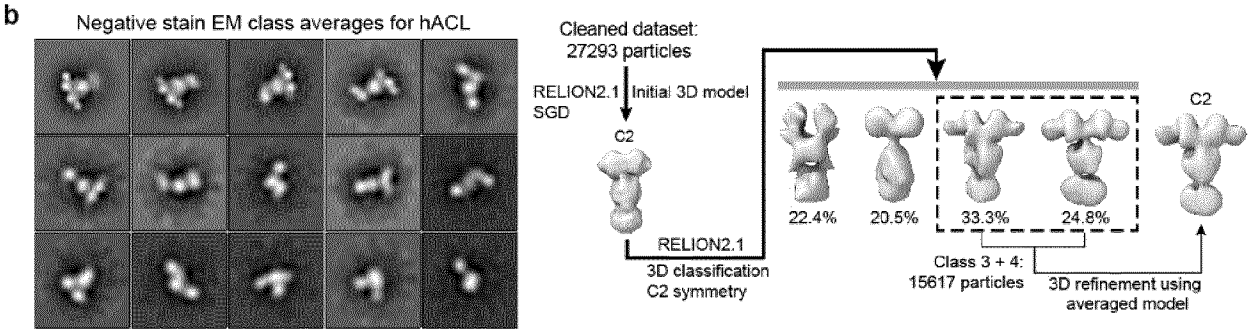
E-P + citrate → E.citryl-P
REACTION 2

E.citryl-P + CoA → E.citryl-CoA + Pi
REACTION 3

E.citryl-CoA → E + acetyl-CoA
REACTION 4 + oxaloacetate

where E represents ACL

b



c

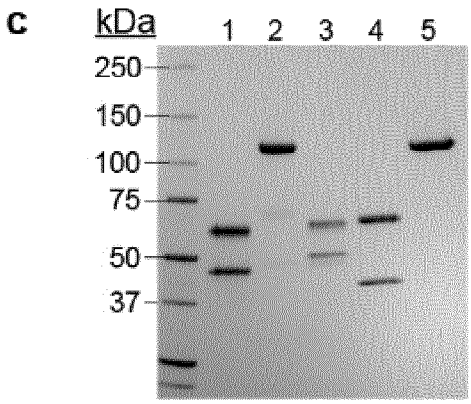


Figure 5 continued

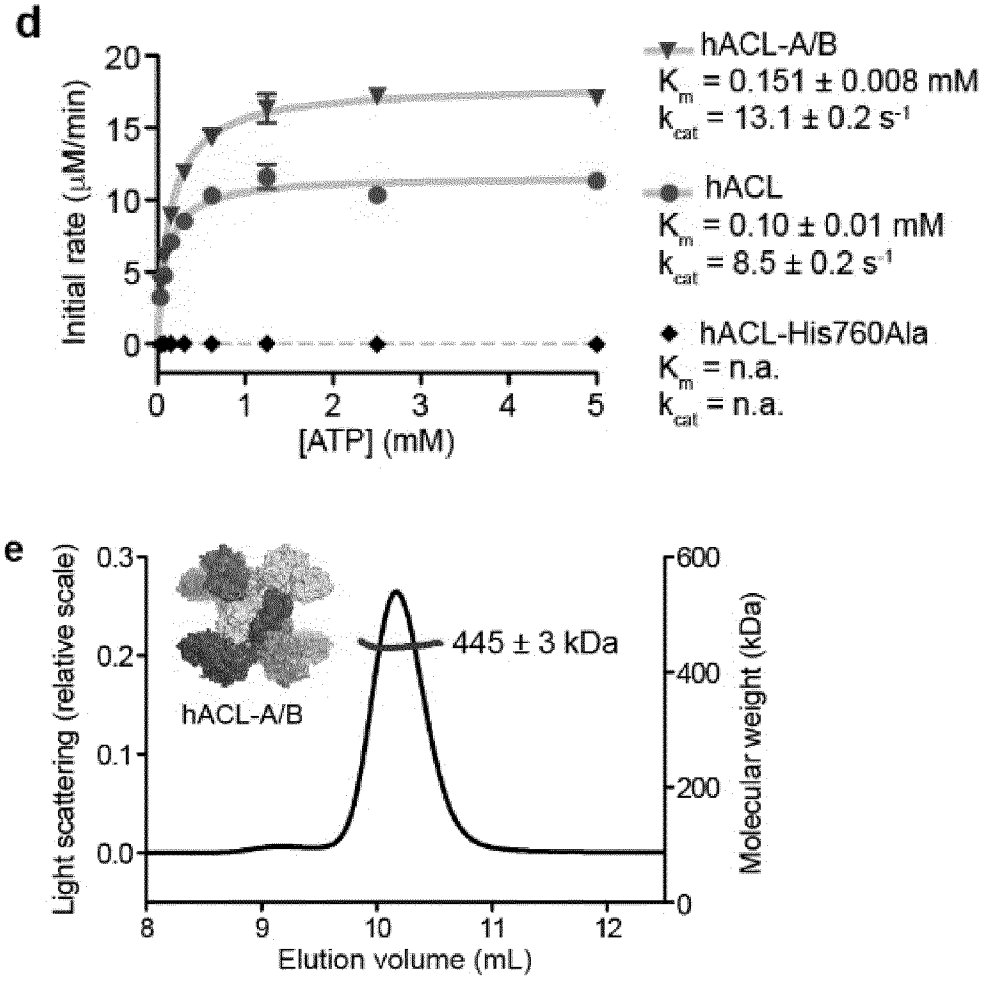


Figure 5 continued

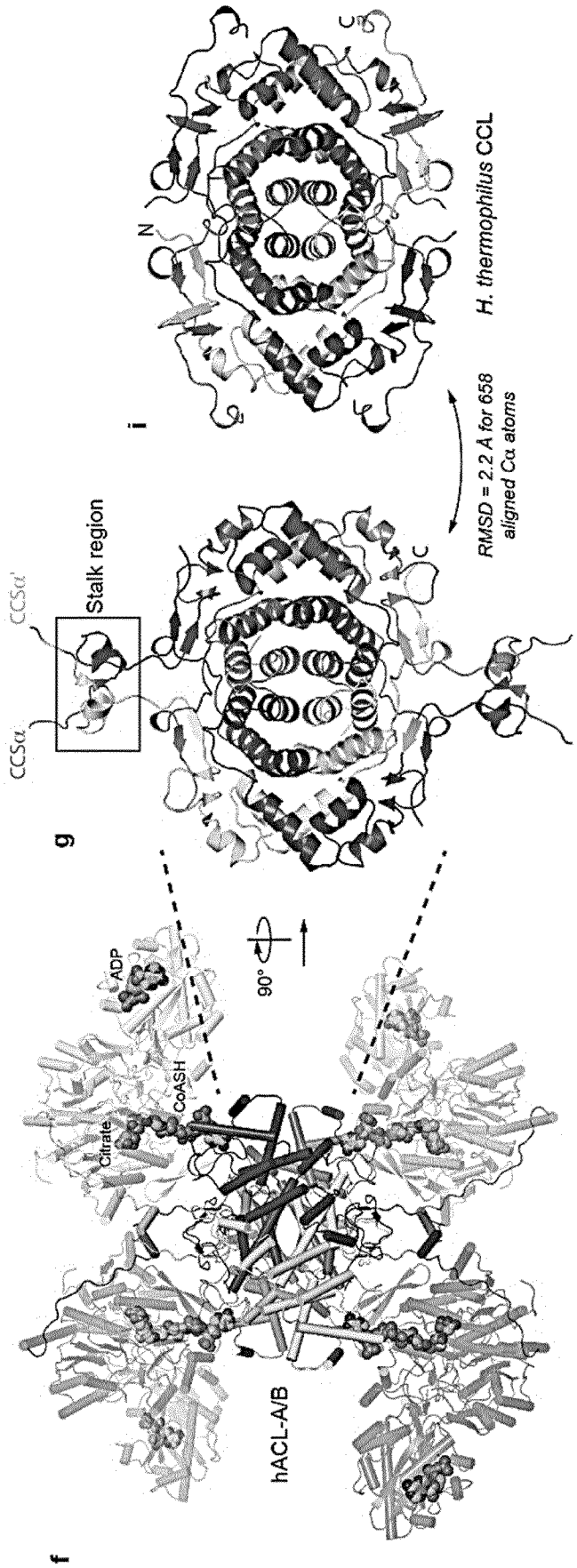


Figure 5 continued

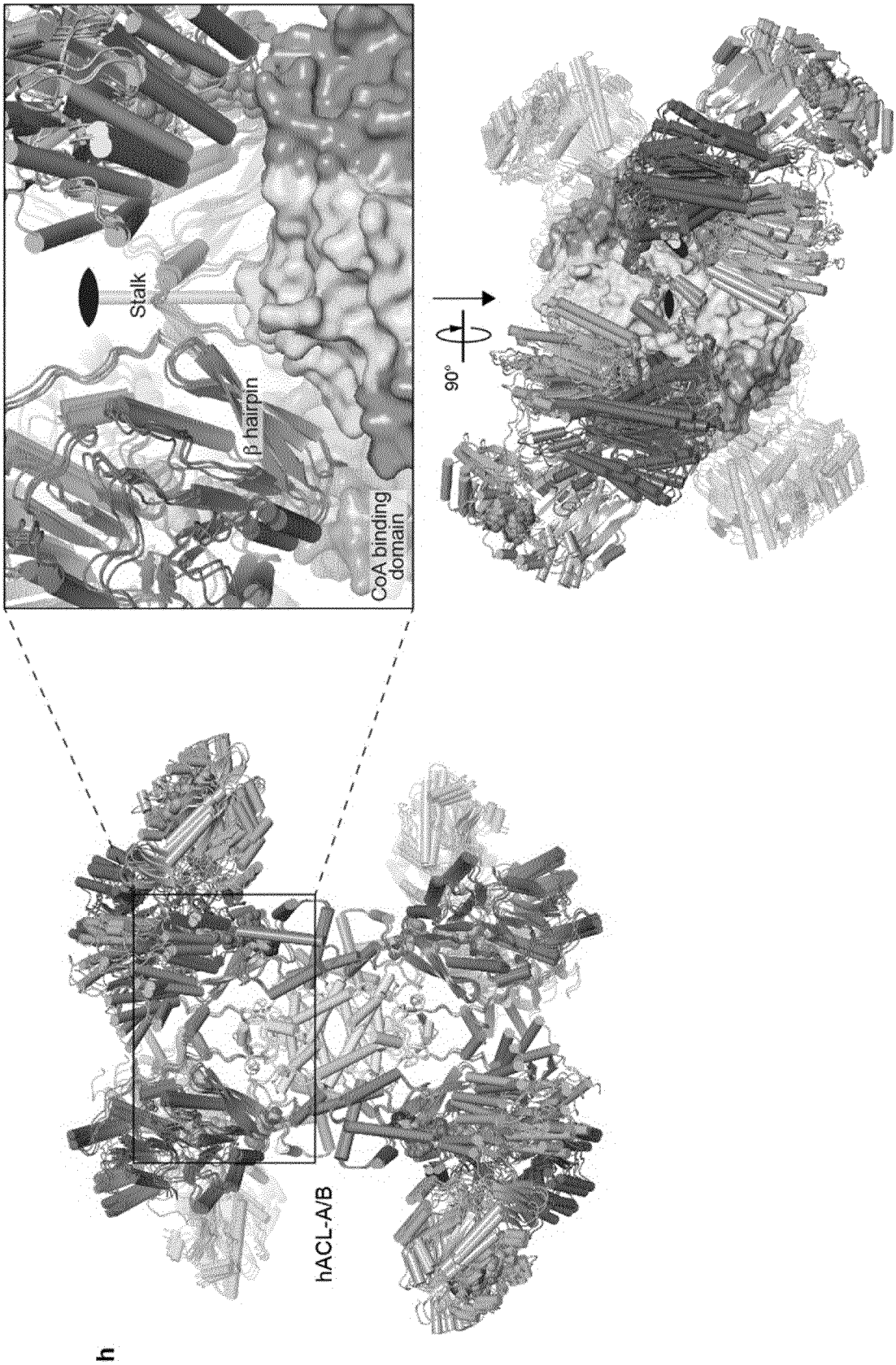


Figure 6

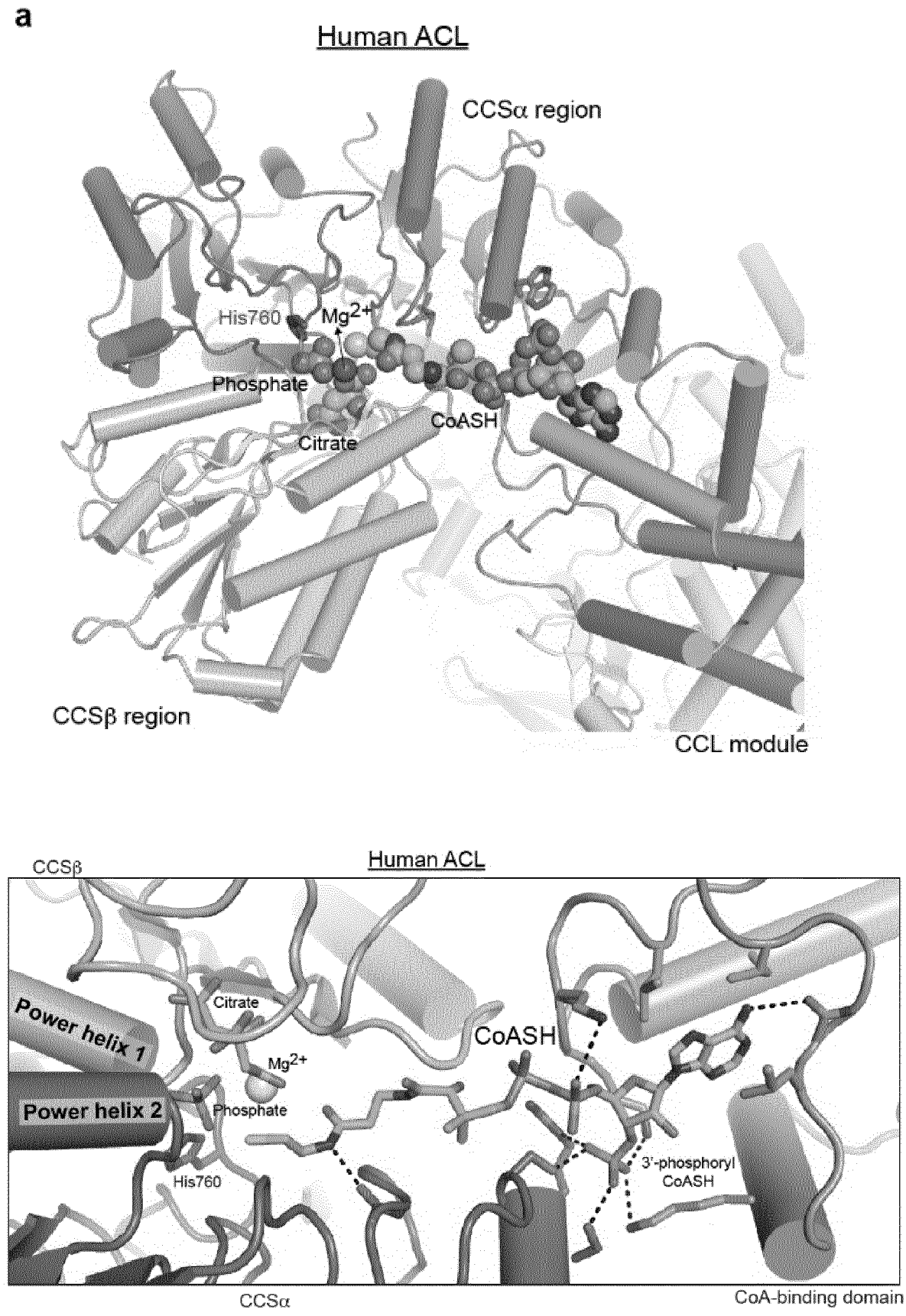
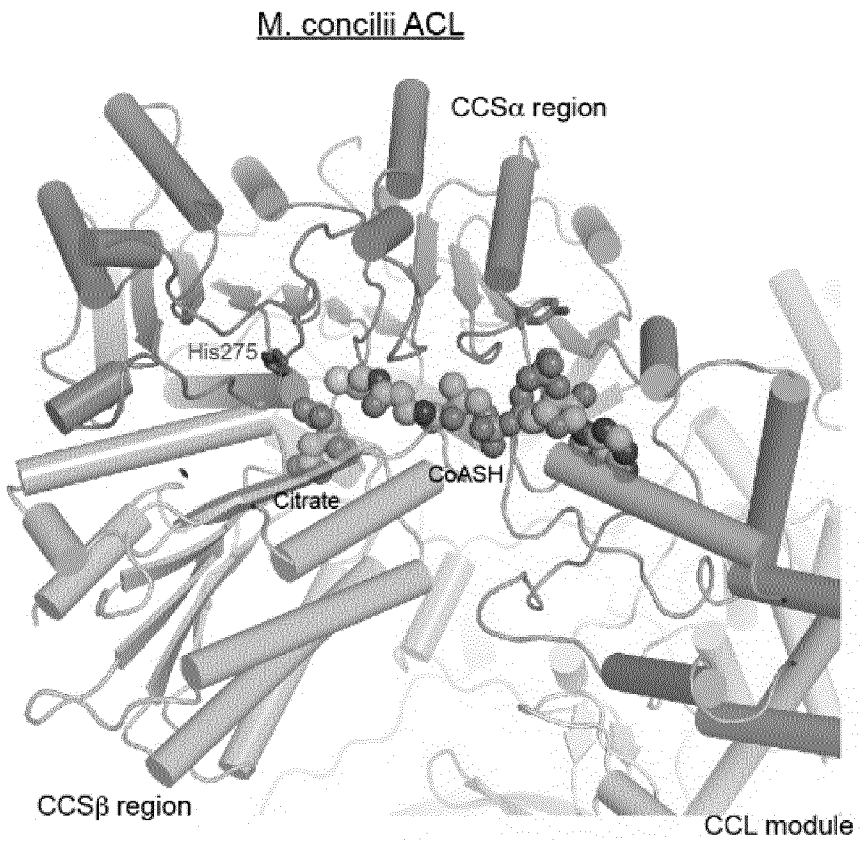

a
Human ACL
CCSα region
His760
Mg2+
Phosphate
CoASH
Citrate
CCSβ region
CCL module
b
Human ACL
CCSβ
Power helix 1
Power helix 2
Citrate
Mg2+
Phosphate
His760
CoASH
3'-phosphoryl CoASH
CCSα
CoA-binding domain

Figure 6 continued

c



d

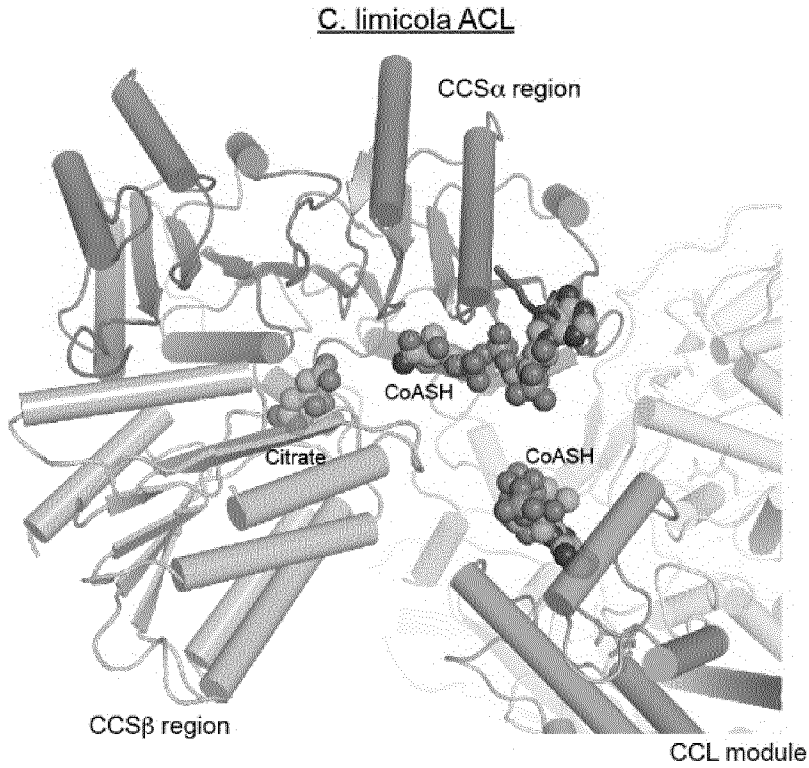


Figure 6 continued

e

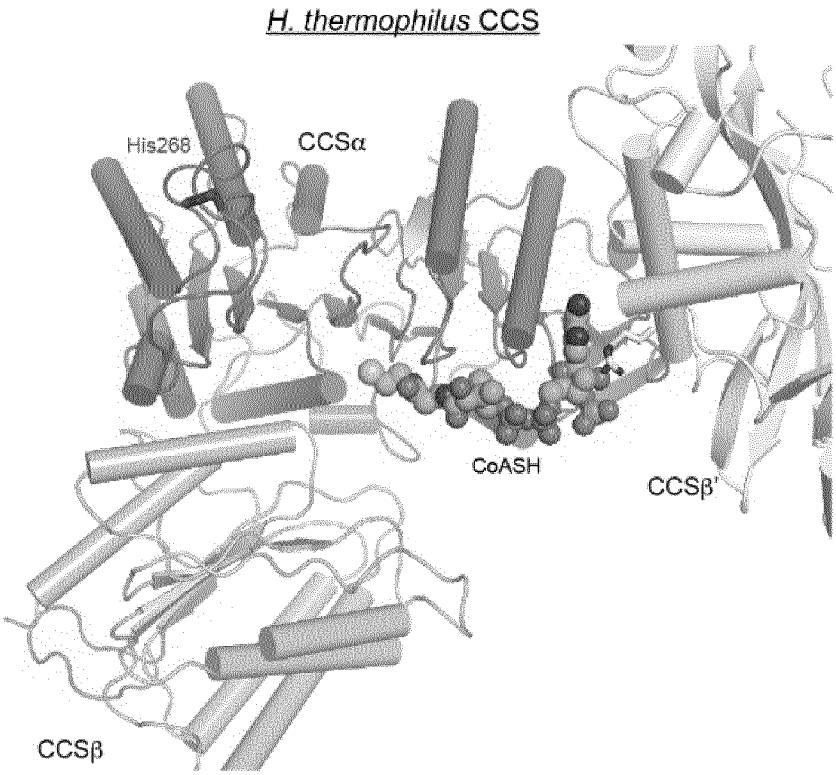


Figure 6 continued

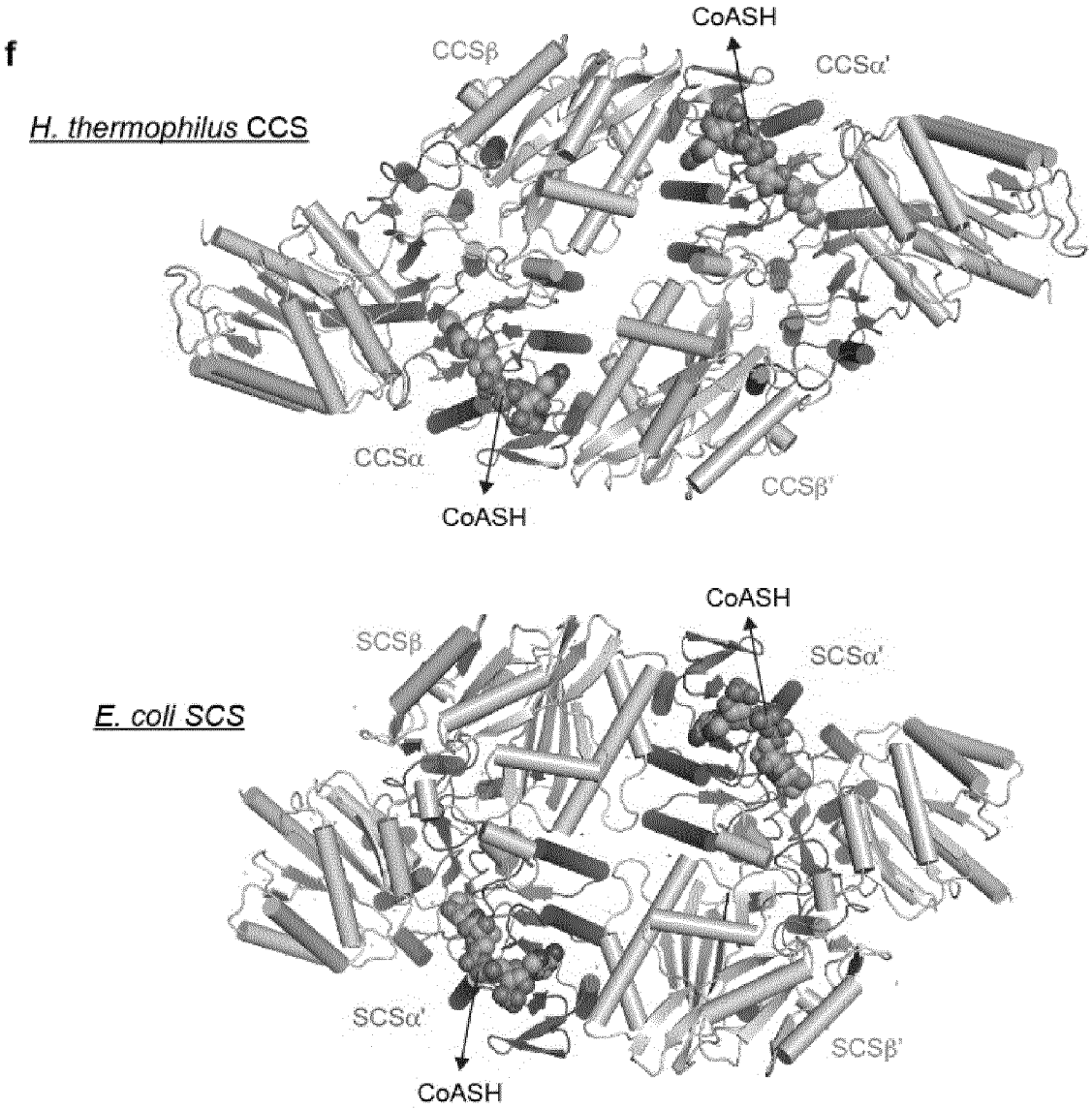


Figure 7

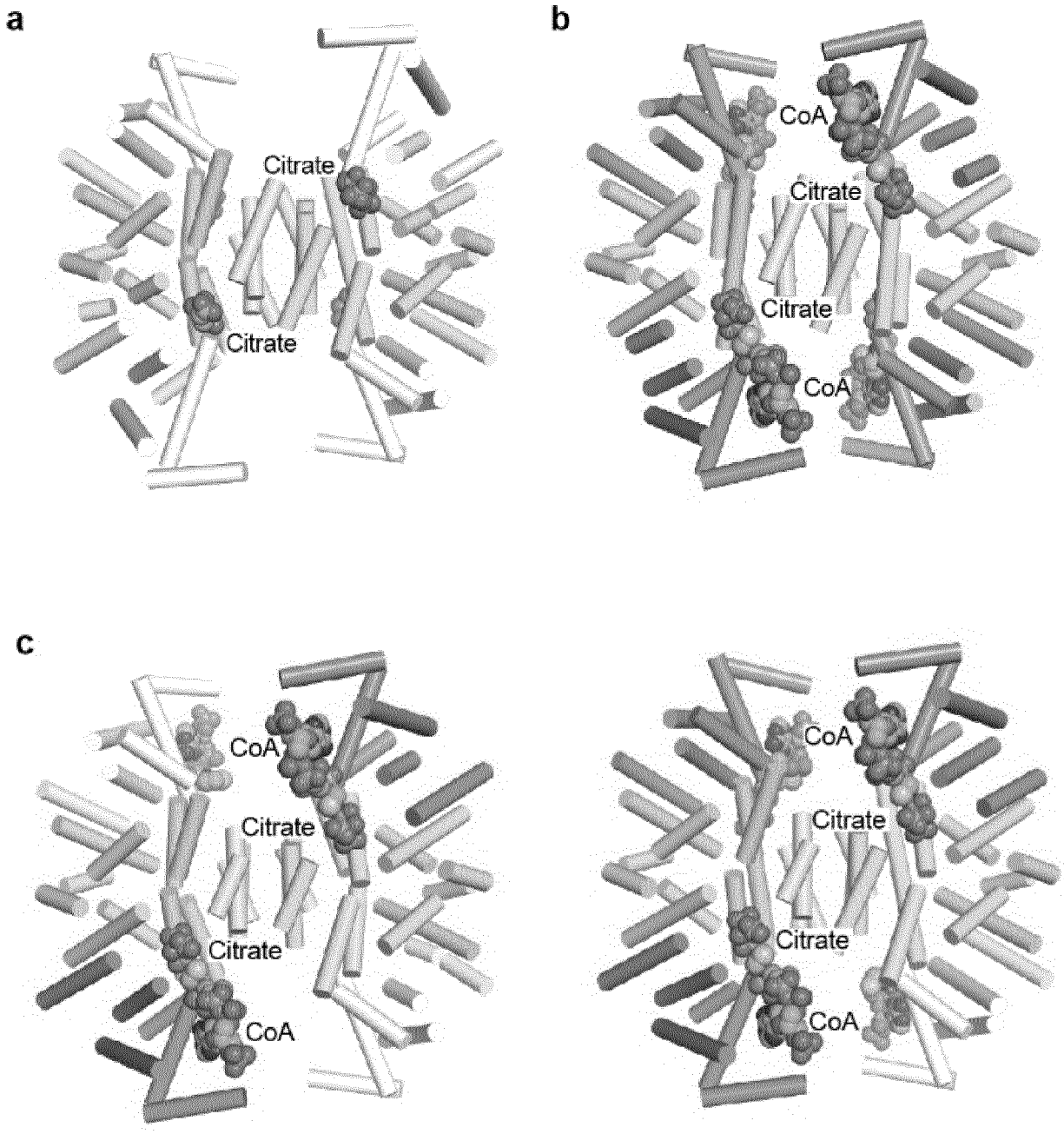

a
Citrate
Citrate
b
CoA
Citrate
Citrate
CoA
c
CoA
Citrate
Citrate
CoA
CoA
Citrate
Citrate
CoA

Figure 7 continued

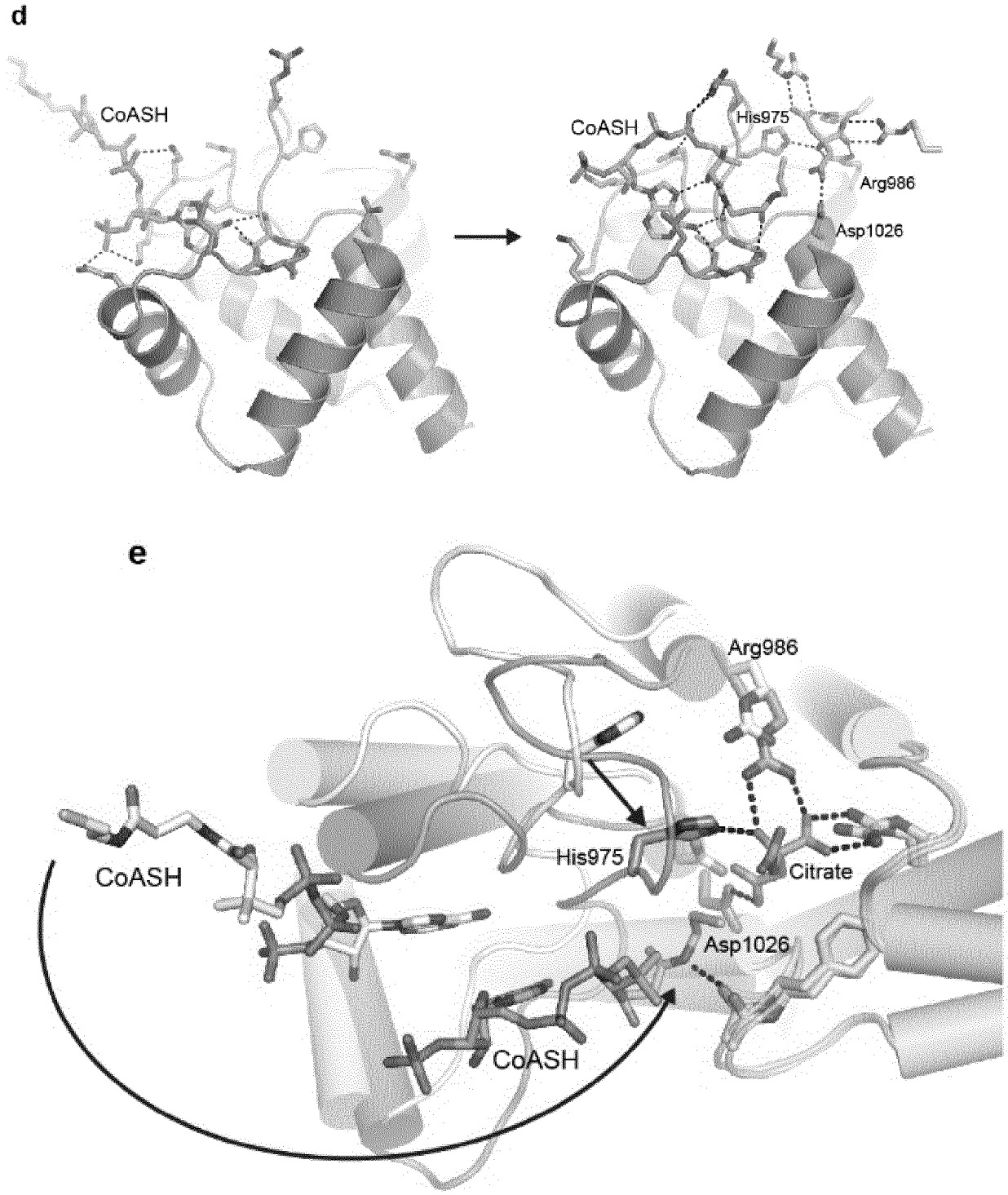


Figure 7 continued

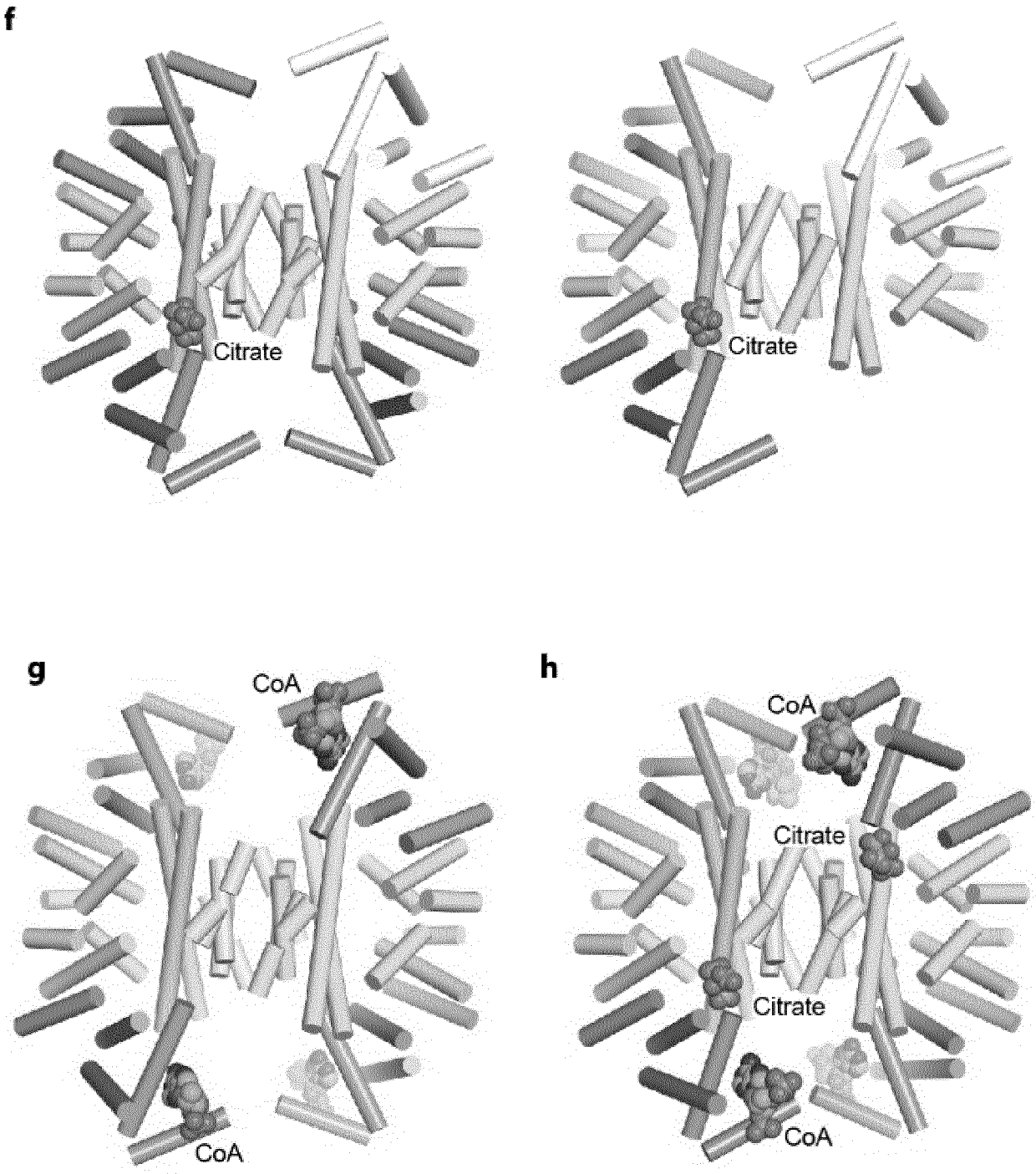


Figure 7 continued

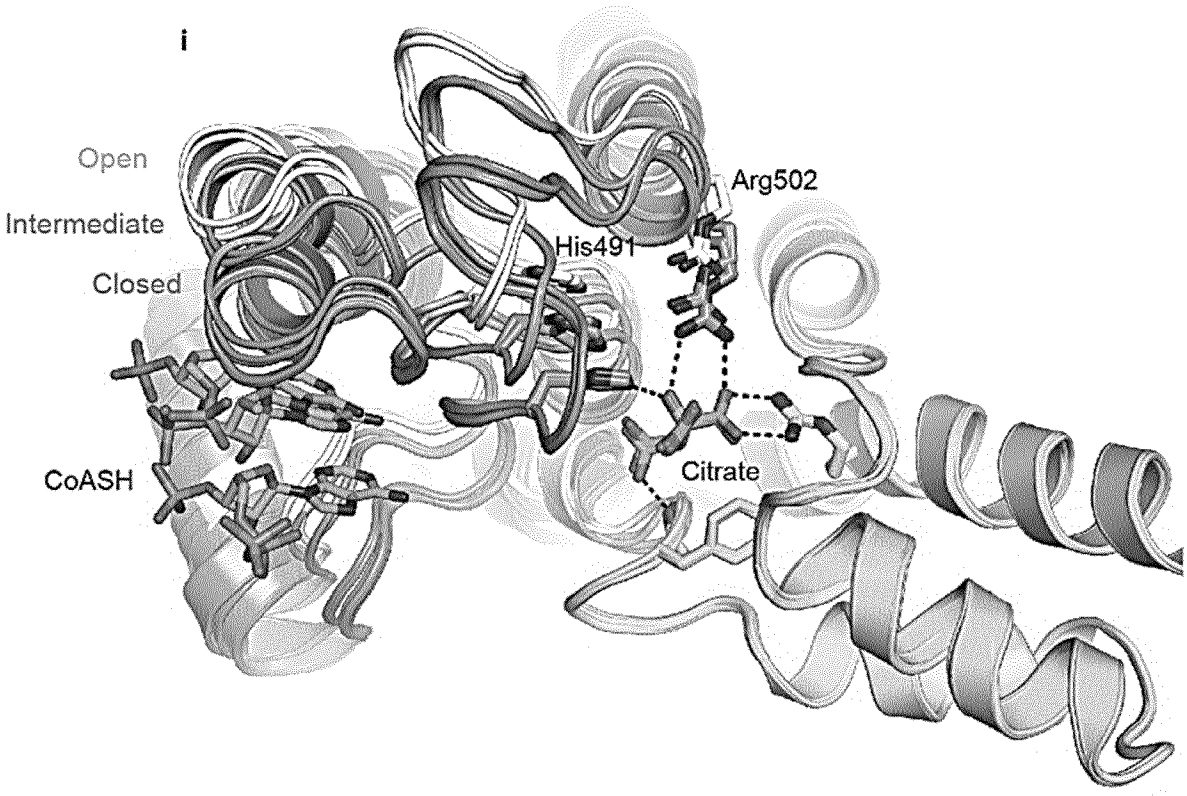


Figure 8

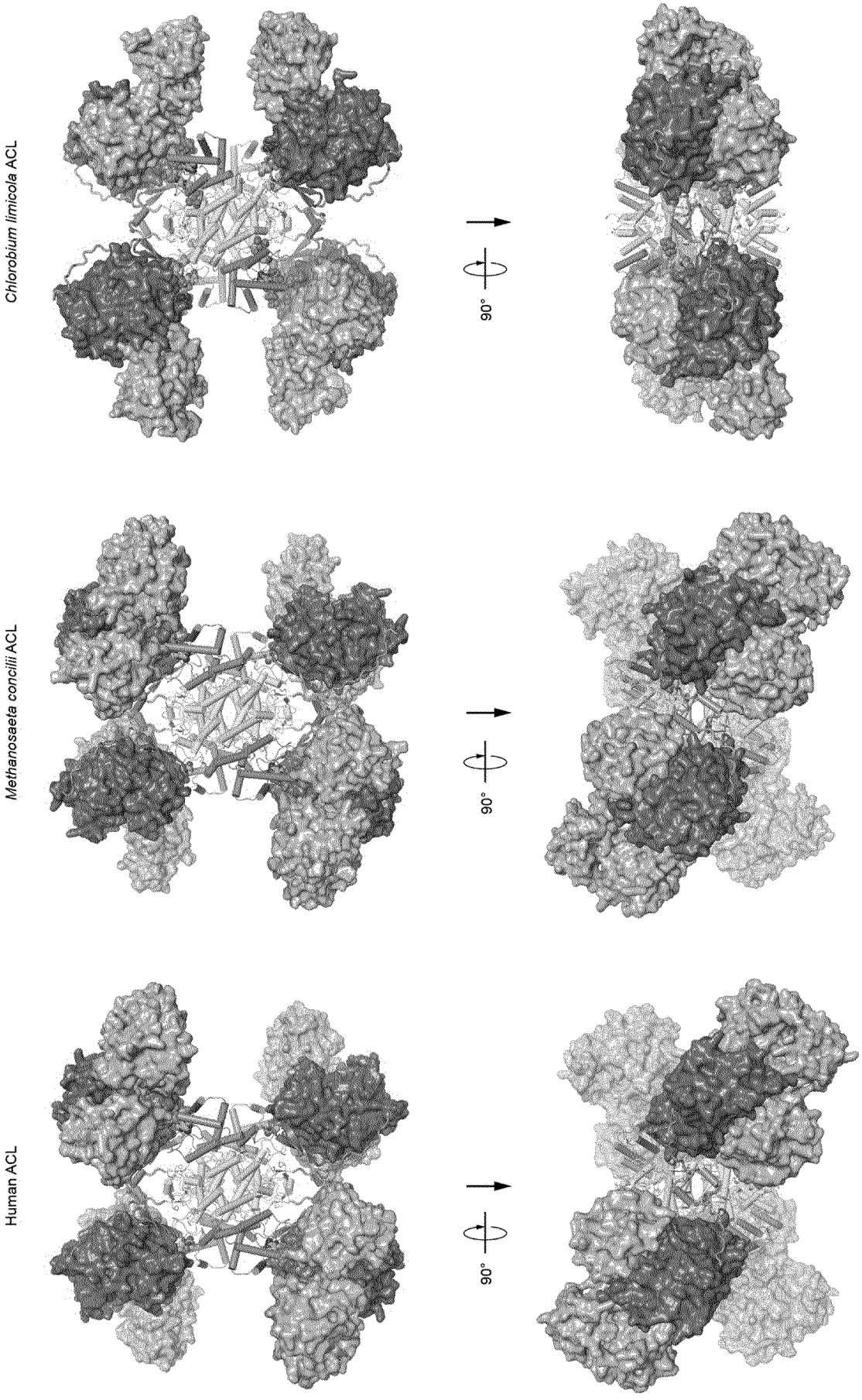

Chlorobium limicola ACL
Methanosaeta concilii ACL
Human ACL
90°
90°
90°

Figure 9

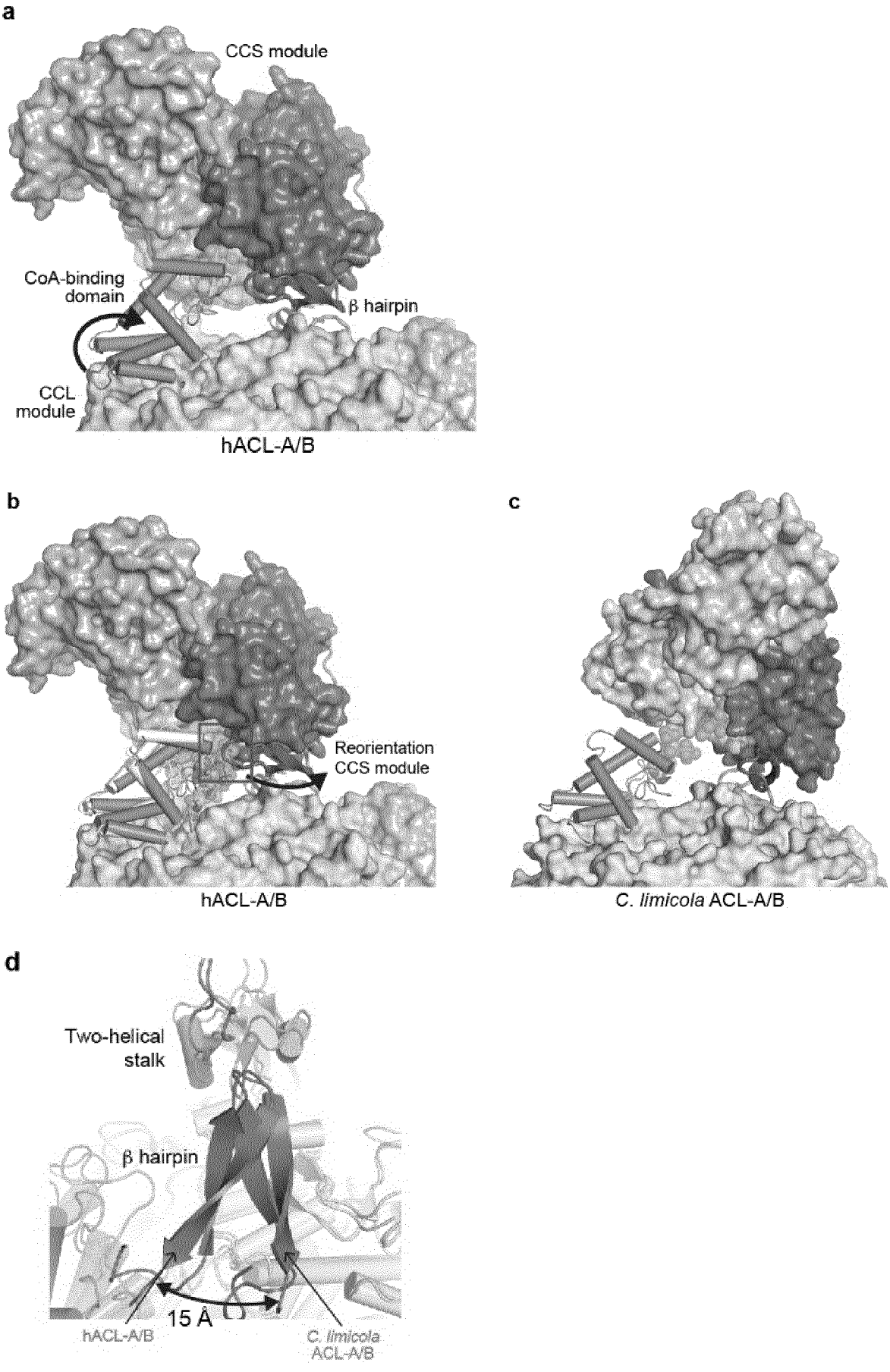
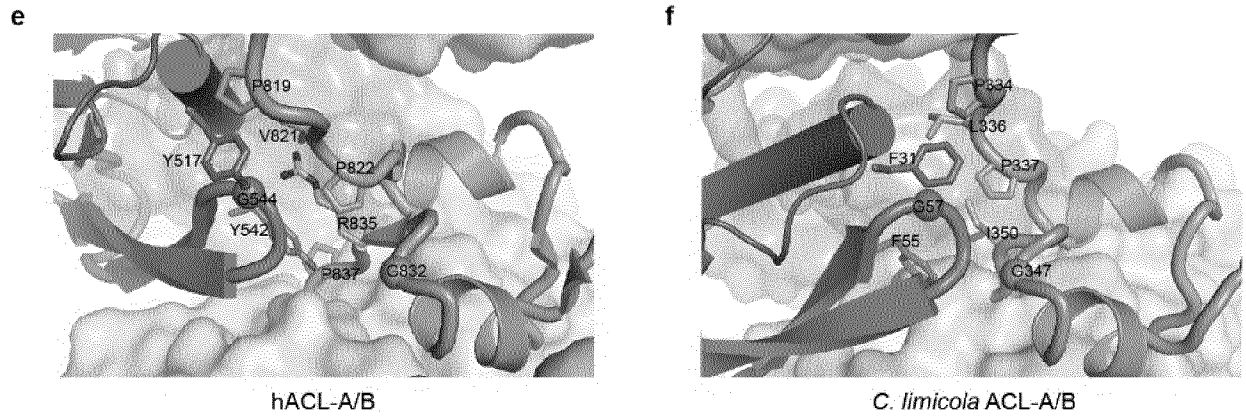
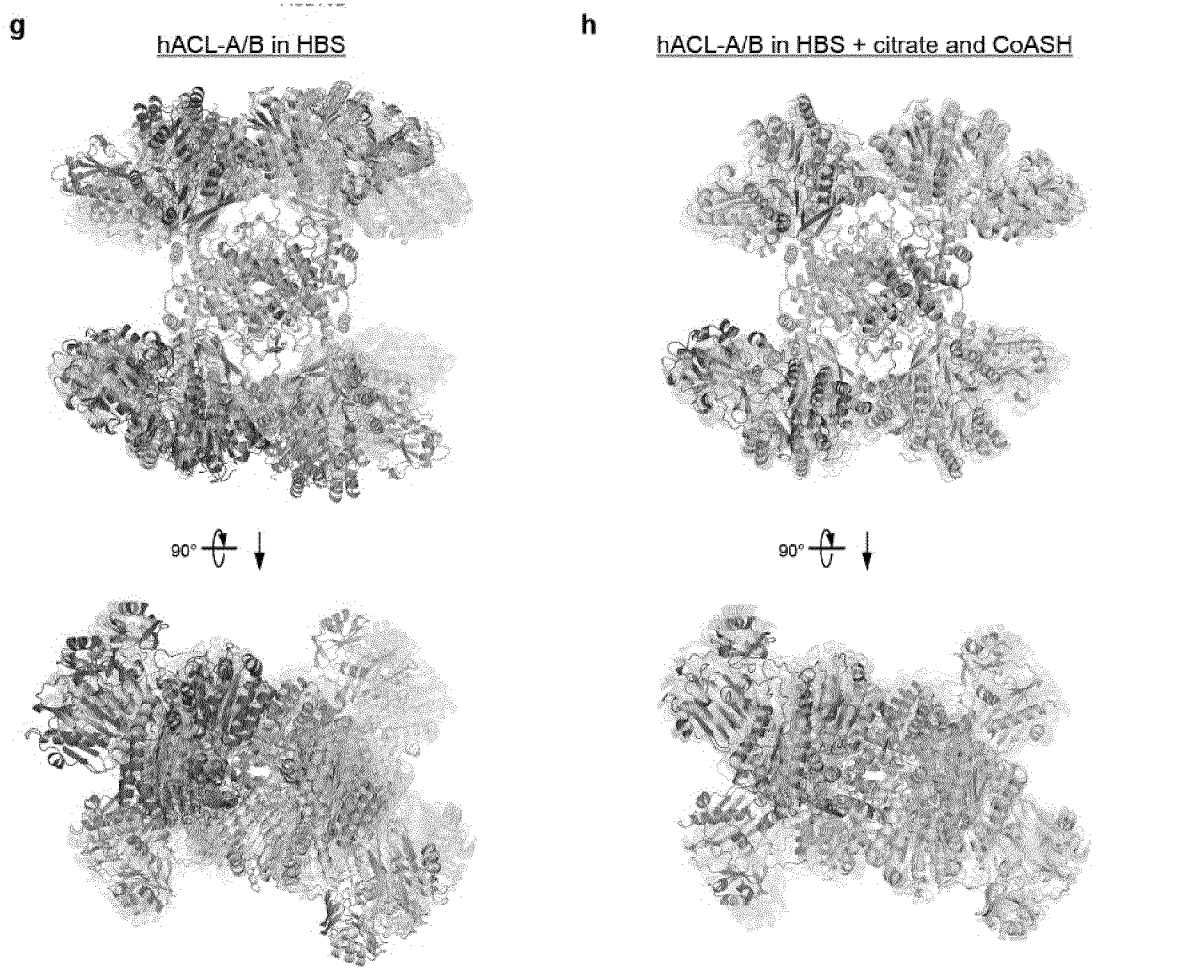

a
CCS module
CoA-binding domain
β hairpin
CCL module
hACL-A/B
b
Reorientation CCS module
hACL-A/B
c
C. limicola ACL-A/B
d
Two-helical stalk
β hairpin
15 Å
hACL-A/B
C. limicola ACL-A/B

Figure 9 continued

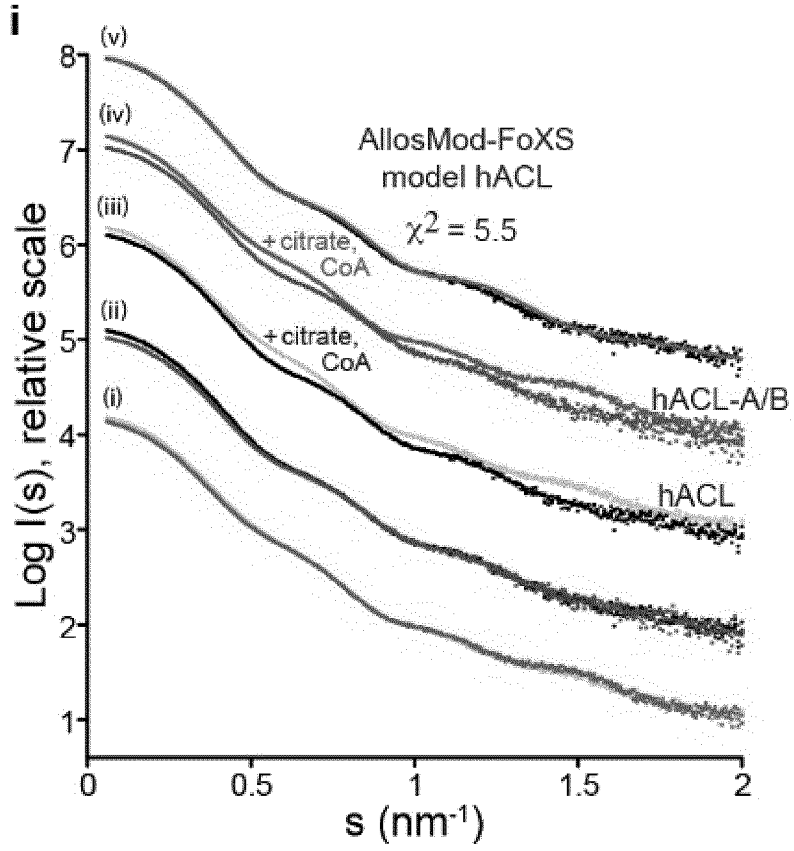




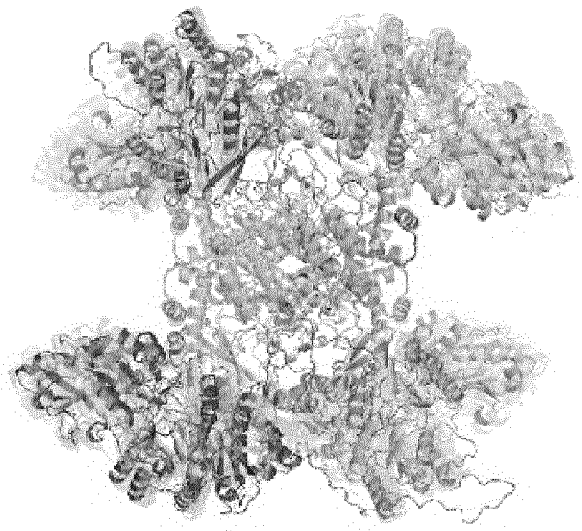
	All-residue rmsd (Å)						SAXS χ^2-values		
							FoXS	*Crysol*	*Crysol 3.0*
P1_hACL-A/B_1	0.0	3.3	3.3	4.9	11.7	13.4	150.5	147.9	67.3
P1_hACL-A/B_2	3.3	0.0	3.9	3.9	10.2	13.1	141.1	137.0	49.6
C2_hACL-A/B_1	3.3	3.9	0.0	3.2	10.0	12.2	73.0	77.7	53.9
C2_hACL-A/B_2	4.9	3.9	3.2	0.0	8.0	11.7	45.0	52.7	36.1
MultiFoXS_1	11.7	10.2	10.0	8.0	0.0	10.2	9.9	12.6	17.2
MultiFoXS_2	13.4	13.1	12.2	11.7	10.2	0.0	17.4	27.2	11.9
→ Two-state model							2.8	7.8	10.2

	All-residue rmsd (Å)					SAXS χ^2-values		
						FoXS	*Crysol*	*Crysol 3.0*
P1_hACL-A/B_1	0.0	3.2	3.3	4.9	4.5	34.4	38.3	6.3
P1_hACL-A/B_2	3.3	0.0	3.9	3.9	5.4	51.8	57.3	9.9
C2_hACL-A/B_1	3.3	3.9	0.0	3.2	2.3	10.7	13.9	4.0
C2_hACL-A/B_2	4.9	3.9	3.2	0.0	4.3	17.1	22.3	8.6
MultiFoXS	4.5	5.4	2.3	4.3	0.0	2.8	4.4	5.1

Figure 9 continued



j hACL in HBS + citrate and CoASH



	All residue rmsd (Å)	SAXS χ^2-values		
		FoXS	*Crysol*	*Crysol 3.0*
P1_hACL-A/B_1	5.2	115.9	130.0	38.3
P1_hACL-A/B_2	5.1	170.0	188.6	71.4
C2_hACL-A/B_1	4.9	57.1	63.7	33.9
C2_hACL-A/B_2	3.9	72.7	86.1	47.7
AllosMod-FoXS	0.0	5.5	8.8	5.2

Figure 10

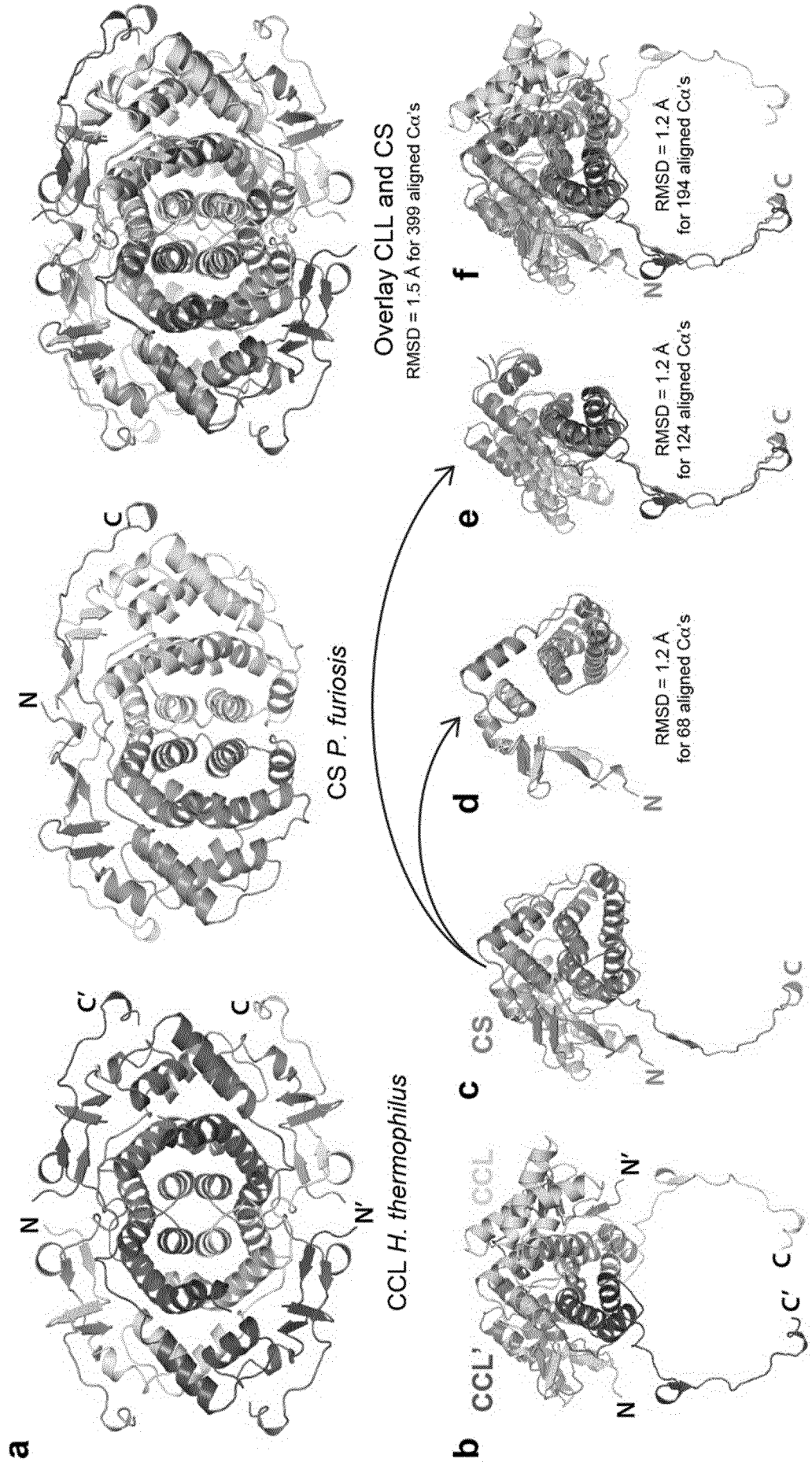

a
CCL H. thermophilus
N
C'
C
N'
CS P. furiosis
N
C
Overlay CLL and CS
RMSD = 1.5 Å for 399 aligned Cα's
b
CCL'
CCL
N'
N
C
C'
c
CS
N
C
d
N
RMSD = 1.2 Å
for 68 aligned Cα's
e
RMSD = 1.2 Å
for 124 aligned Cα's
C
f
N
RMSD = 1.2 Å
for 194 aligned Cα's
C

Figure 10 continued

G

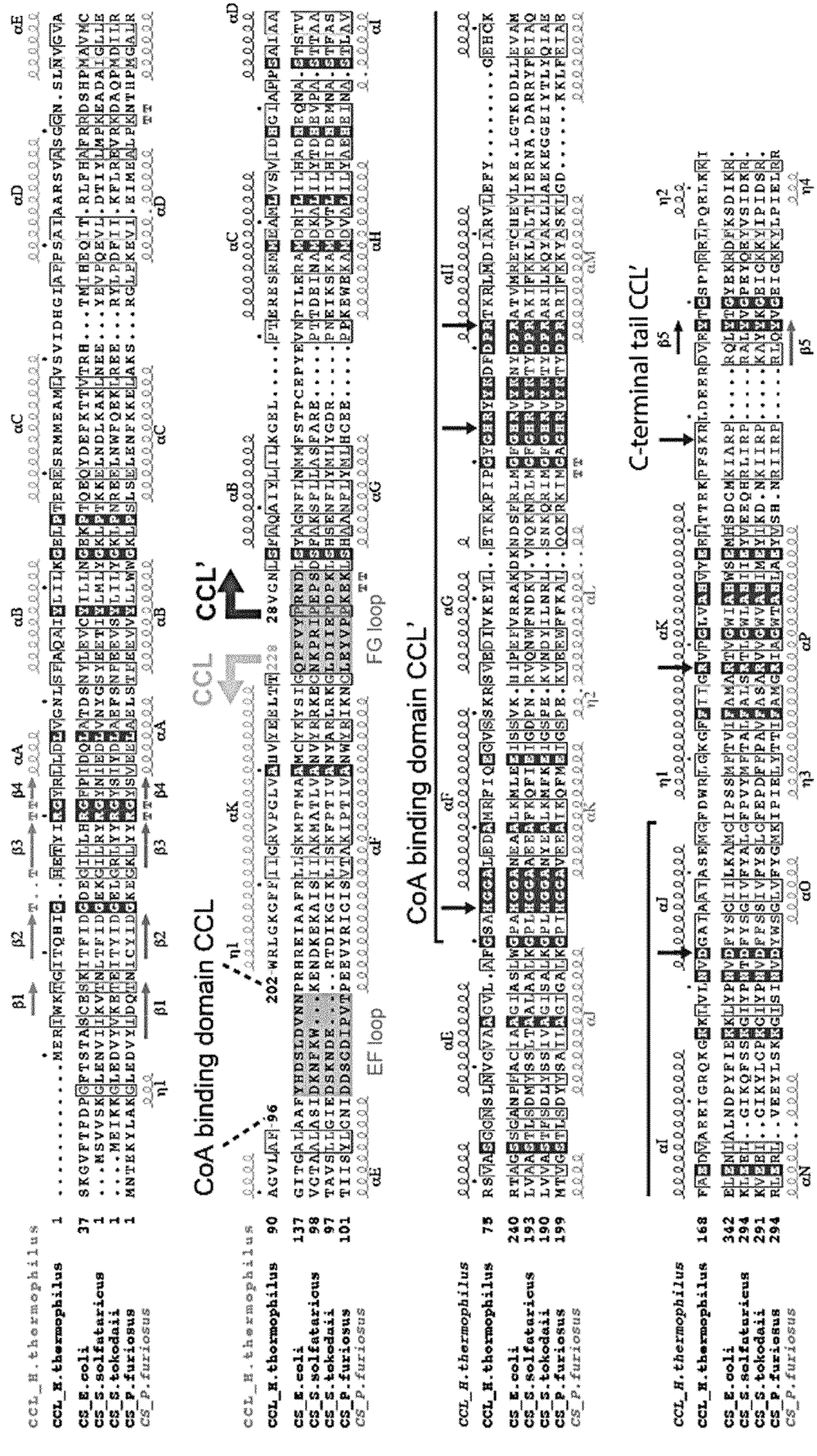


Figure 10 continued

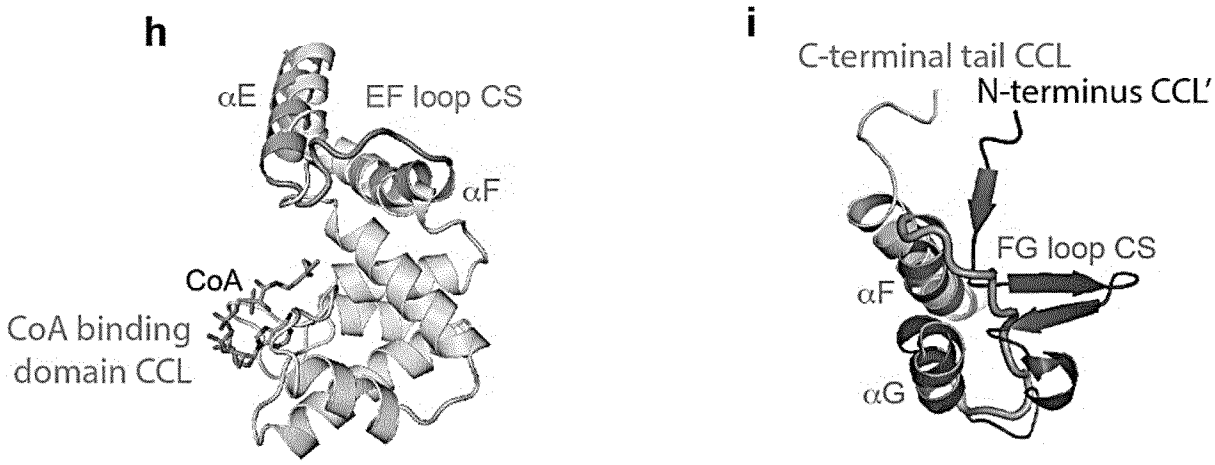


Figure 11

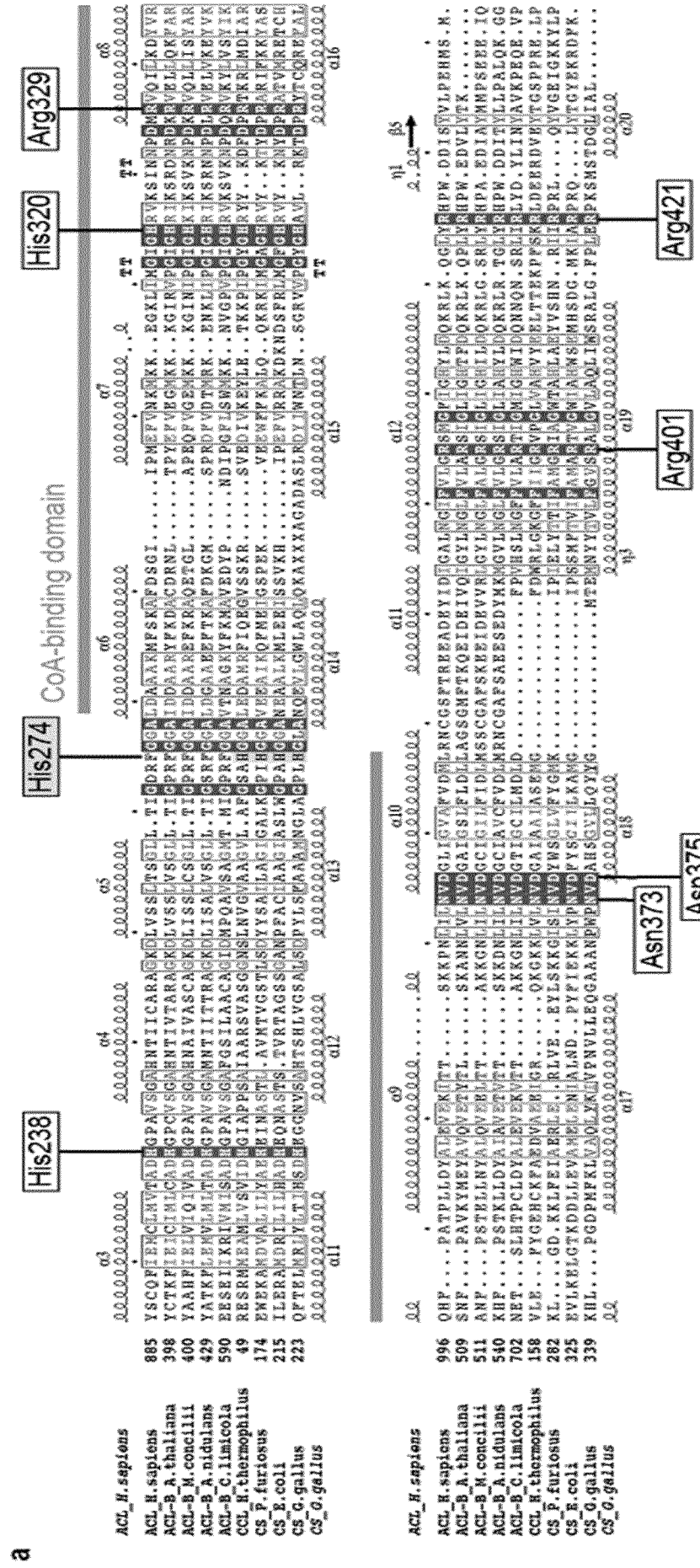

a
His238
His274
CoA-binding domain
His320
Arg329
Arg401
Arg421
Asn373
Asp375
ACL_H.sapiens
ACL_H.sapiens
ACL-B_A.thaliana
ACL-B_M.conciliI
ACL-B_A.nidulans
ACL-B_C.limicola
CCL_H.thermophilus
CS_P.furiosus
CS_E.coli
CS_G.gallus
CS_G.gallus

Figure 11 continued

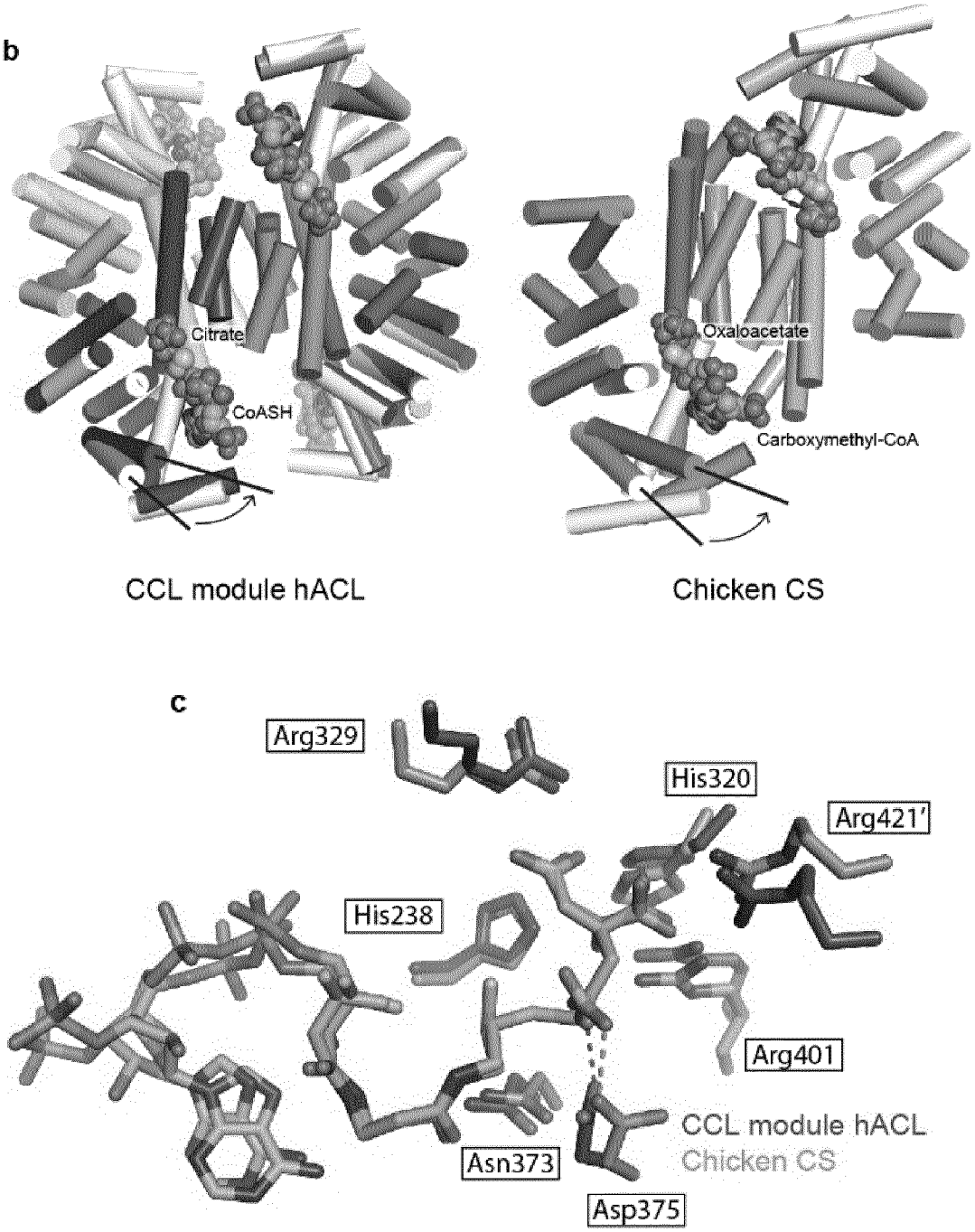
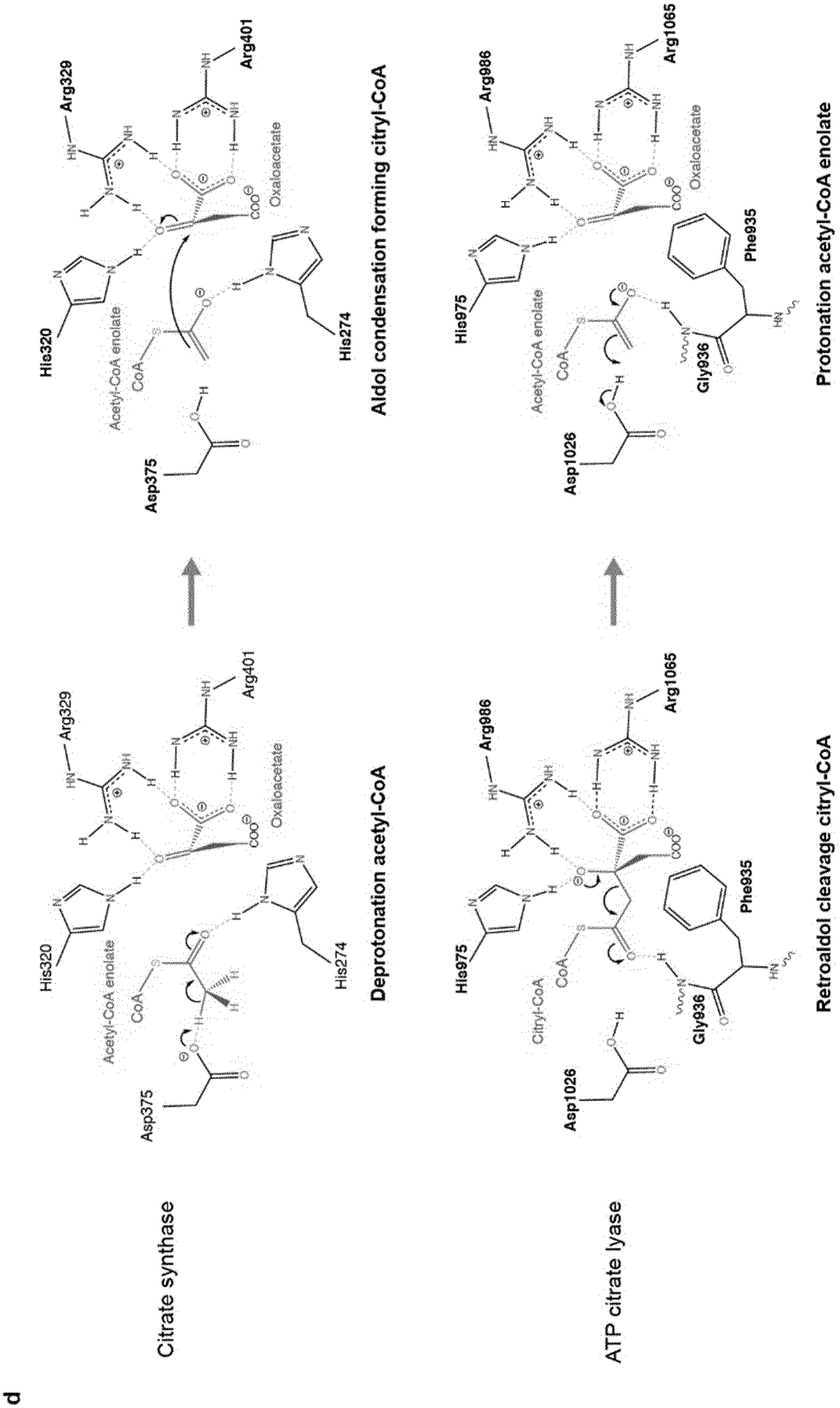


Figure 11 continued



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/055859

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/88 C30B29/58
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K C30B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JINHONG HU ET AL: "Binding of hydroxycitrate to human ATP-citrate lyase", ACTA CRYSTALLOGRAPHICA / D. SECTION D, BIOLOGICAL CRYSTALLOGRAPHY, vol. 73, no. 8, 1 August 2017 (2017-08-01) , pages 660-671, XP055685399, Oxford ISSN: 2059-7983, DOI: 10.1107/S2059798317009871 cited in the application abstract page 661, left-hand column, line 21 - line 25 page 661, right-hand column, line 40 - page 662, left-hand column, line 9 page 660, line 7 - line 11 ----- -/--	1-11

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 April 2020	23/04/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Mundel, Christophe

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Jinhong Hu ET AL: "Determination of the Binding Mode of Hydroxycitrate, and Investigation of the Binding Site of 14-3-3 Protein on Human ATP Citrate Lyase", , 1 January 2017 (2017-01-01), XP055685352, DOI: 10.11575/PRISM/25607 Retrieved from the Internet: URL:https://prism.ucalgary.ca/bitstream/handle/11023/3549/ucalgary_2017_hu_jinhong.pdf?sequence=3&isAllowed=y [retrieved on 2020-04-10] abstract Introduction; page 1 page 2, line 18 - line 21 page 3, line 6 - line 10 page 10, line 1 - line 11 -----	1-11
Y	TIANJUN SUN ET AL: "Identification of the Citrate-binding Site of Human ATP-Citrate Lyase Using X-ray Crystallography", JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 285, no. 35, 27 August 2010 (2010-08-27), pages 27418-27428, XP055685359, US ISSN: 0021-9258, DOI: 10.1074/jbc.M109.078667 cited in the application page 27419, left-hand column, line 23 - line 35 page 27426, right-hand column, line 54 - page 27427, left-hand column, line 2 -----	1-11
Y	TIANJUN SUN ET AL: "ADP-Mg 2+ bound to the ATP-grasp domain of ATP-citrate lyase", ACTA CRYSTALLOGRAPHICA SECTION F STRUCTURAL BIOLOGY AND CRYSTALLIZATION COMMUNICATIONS, vol. 67, no. 10, 1 October 2011 (2011-10-01), pages 1168-1172, XP055685383, DOI: 10.1107/S1744309111028363 cited in the application page 1168, line 10 - line 16 page 1169, left-hand column, line 5 - line 8 -----	1-11

1