Chinese Science Bulletin

Evolution of the serum resistance-associated *SRA* gene in African trypanosomes

LAI DeHua[1†], WANG QiaoPing[1], LI Zhi[1], JULIUS Lukeš[2] & LUN ZhaoRong[1†]

[1] Center for Parasitic Organisms, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China;
[2] Biology Centre, Institute of Parasitology, Czech Academy of Sciences; Faculty of Biology, University of South Bohemia, 37005 České Budějovice (Budweis), Czech Republic

Serum resistance-associated (SRA) protein, a protein unique for *Trypanosoma brucei rhodesiense*, is responsible for resistance of this parasite to the lysis by normal human serum (NHS) and is a vital molecular marker to distinguish this species from other African trypanosomes. We cloned and sequenced the *SRA* basic copy (*SRAbc*) gene from *T. b. rhodesiense* and related species and found that this gene is confined to the subgenus *Trypanozoon*. The average 82% identity among the sequenced *SRAbc* genes indicates that they may have a common origin and are highly conserved. Since *SRAbc* coexists in the *T. b. rhodesiense* genome with *SRA*, we propose that *SRAbc* might be the 'donor *VSG*', which after duplication became inserted into the expression site by recombination. Under natural selection, *SRAbc* could reform into *SRA* following mosaic formation.

SRA, *SRAbc*, evolution, African trypanosomes

Kinetoplastid flagellates of the subgenus *Trypanozoon* are widely distributed blood trypanosomes parasitizing a range of mammalian hosts including man and domestic animals. The subgenus includes only subspecies of *Trypanosoma brucei*, *T. b. evansi*, *T. b. equiperdum*, *T. b. brucei*, *T. b. gambiense* and *T. b. rhodesiense*[1,2]. Thanks to the presence of the trypanosome lytic factors (TLFs) in normal human serum (NHS), humans are resistant to these trypanosomes with the important exception of the latter two. Yet only the mechanism by which *T. b. rhodesiense* readily survives in NHS is known. Remarkably, as shown by a series of elegant experiments, the resistance is based solely on the synthesis of serum resistance-associated (SRA) protein, which blocks the TLFs-mediated trypanosome lysis and confers resistance to NHS[3–5].

The *SRA* gene, a vital molecular marker so far unique for *T. b. rhodesiense*[6–9], is an atypical truncated form of the variant surface glycoprotein (*VSG*) gene[3,5,10] that can be amplified as a 1.2 kb-long PCR product. In some isolates of *T. b. brucei*, *T. b. rhodesiense*, primers designed against the *SRA* gene amplified not only the target gene, but also a gene termed *SRA*-related or *SRA* basic copy (*SRAbc*)[11,12]. The *SRAbc* gene retains an average 76% similarity to the *SRA* gene, but is interrupted by frame shifts and a premature stop codon[11]. Structure modeling of the SRA and SRAbc amino acid sequences revealed that the latter has a longer surface domain[12].

Since the *SRAbc* gene may be an evolutionary informative predecessor of the important *SRA* gene, we have attempted to amplify its fragment from DNA of 8 isolates of *T. b. evansi* from South America, Asia and Africa, 4 isolates of *T. b. equiperdum* from Africa and Asia, 2 isolate of *T. b. brucei*, 2 isolates of *T. b. gambiense*

Received November 12, 2008; accepted January 15, 2009
doi: 10.1007/s11434-009-0137-z
†Corresponding author (email: lsslzr@mail.sysu.edu.cn, ltaclai@hotmail.com)
Supported by National Natural Science Foundation of China (Grant Nos. 30570245, 30670275), Changjiang Scholars and Innovative Research Team in University (Grant No. DPCKSCU/IRT0447), International Foundation for Science of Sweden (Grant No. B/4318-1), Grant Agency of the Czech Republic (Grant No. Z60220518) and Education Foundation of the Czech Republic (Grant No. 2B06129)

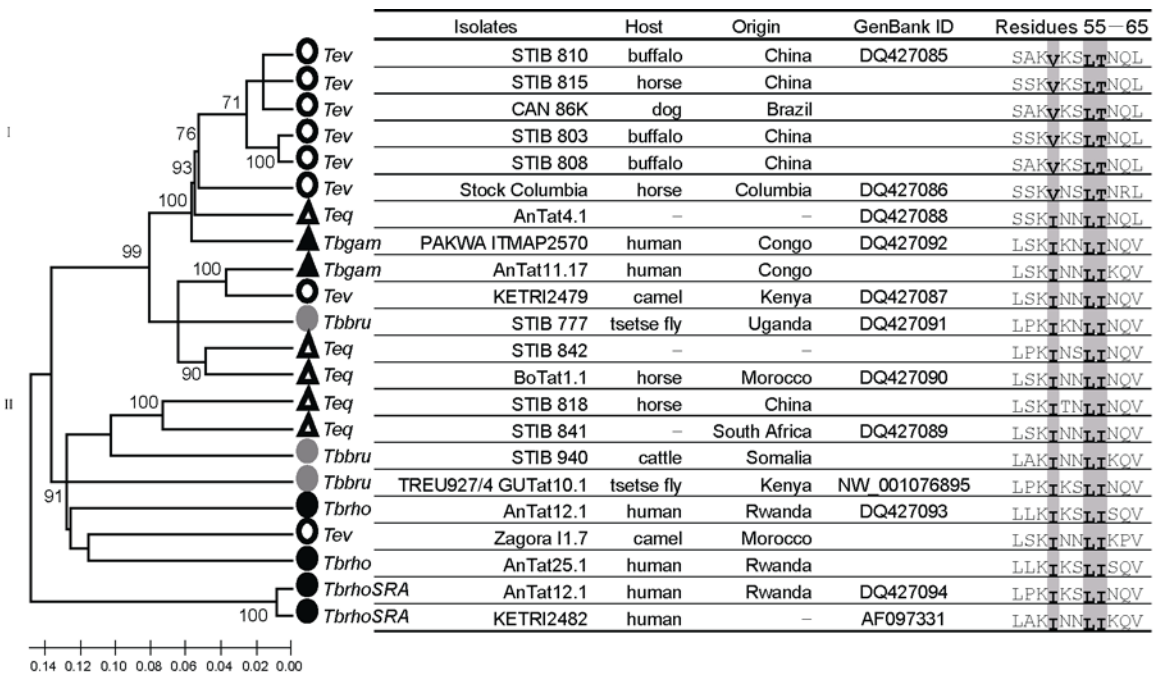
and 2 isolates of *T. b. rhodesiense* from Africa and 1 isolate of *T. congolense* (Figure 1).

To amplify the *SRAbc* gene, forward (5′-CACACC-TCTAAGAATCACAATAG-3′) and reverse (5′-ATTC-TTGTGCCTTGGTGC-3′) primers were derived from the *T. b. brucei* (NW_001076895) and *T. b. rhodesience* sequences (DQ427094). PCR was performed in 25 μL reaction mixture containing about 10 ng of genomic DNA, 0.8 μmol/L of each primer, 200 μmol/L of each dNTP, and 1 U of *Ex* Taq polymerase (TaKaRa, Japan). The PCR program was as follows: 5 min at 94℃; 30 s at 94℃, 1 min at 55℃, 1 min at 72℃, 35 cycles; 10 min at 72℃. Amplicons was seperated in 1.5% agarose gel in the present of ethidium bromide and the selected amplicon were cloned into pBS-T vector (Tiangen, Beijing), and sequenced. Sequence data was analyzed by MEGA 3.1[13].

A *SRAbc* fragment about 1.5 kb was amplified from all isolates except *T. congolense*, which is not a *Trypanozoon* species and the sequence analysis of the *SRAbc* sequences revealed an average 82% similarity. Amplicon of *SRAbc* obtained from *T. b. rhodesiense* was rather unexpected, since it is known to have a functional *SRA* gene in this trypanosome. These results confirmed the presence of the *SRAbc* in all *Trypanozoon* subspecies and its potential use as a marker for this subgenus.

The obtained *Trypanozoon SRAbc* sequences were used for phylogenetic analysis with the *SRA* gene used as an outgroup. All sequences obtained in this study formed two clusters. Six isolates of *T. b. evansi* are closely related and clearly monophyletic, whereas the Kenyan *T. b. evansi* KETRI 2479 appeared in a different branch within cluster I and Zagora I1.7 (from Morocco) falls into cluster II (Figure 1). Substantial differences have been previously described between this type B strain and type A strains of *T. b. evansi* both in their kinetoplastic[14] and nuclear genome[15]. The same applies for the *T. b. equiperdum* isolates used in this study, as they are present in both clusters. Thus, the *SRAbc* sequence further supports the notion that *T. b. evansi* and *T. b. equiperdum* is represented by multiple types which might not be of monophyletic origin[16]. More interestingly, both two *T. b. gambiense* isolates were present in cluster I, showing closer relationship with *T. b. evansi* but not the other human infective subspecies *T. b. rhodesiense*. Actually, our findings support the proposal of different evolutionary pathways in *T. b. gambiense* and *T. b. rhodesiense* to invade human beings[6].

The tertiary structure of the VSG protein, composed of four parts (a signal peptide in the N-terminal region, a



Isolates	Host	Origin	GenBank ID	Residues 55–65
STIB 810	buffalo	China	DQ427085	SAKVKSLTNQL
STIB 815	horse	China		SSKVKSLTNQL
CAN 86K	dog	Brazil		SAKVKSLTNQL
STIB 803	buffalo	China		SSKVKSLTNQL
STIB 808	buffalo	China		SAKVKSLTNQL
Stock Columbia	horse	Columbia	DQ427086	SSKVNSLTNRL
AnTat4.1	–	–	DQ427088	SSKTNNLINQL
PAKWA ITMAP2570	human	Congo	DQ427092	LSKTKNLINQV
AnTat11.17	human	Congo		LSKTNNLIKQV
KETRI2479	camel	Kenya	DQ427087	LSKTNNLINQV
STIB 777	tsetse fly	Uganda	DQ427091	LPKTKNLINQV
STIB 842	–	–		LPKTNSLINQV
BoTat1.1	horse	Morocco	DQ427090	LSKTNNLINQV
STIB 818	horse	China		LSKTTNLINQV
STIB 841	–	South Africa	DQ427089	LSKTNNLINQV
STIB 940	cattle	Somalia		LAKTNNLIKQV
TREU927/4 GUTat10.1	tsetse fly	Kenya	NW_001076895	LPKTKSLINQV
AnTat12.1	human	Rwanda	DQ427093	LLKTKSLISQV
Zagora I1.7	camel	Morocco		LSKTNNLIKPV
AnTat25.1	human	Rwanda		LLKTKSLISQV
AnTat12.1	human	Rwanda	DQ427094	LPKTKSLINQV
KETRI2482	human	–	AF097331	LAKTNNLIKQV

Figure 1 *SRAbc*-based phylogenetic tree. Neighbor-Joining Phylogenetic tree as inferred from the *SRAbc* sequences of the *Trypanozoon* species, using MEGA3.1 software[13] with Bootstrap 10000 replicates, pairwise deletion treatment for gaps/missing data. *T. b. evansi* isolates are marked as Tev; *T. b. equiperdum* isolates are marked as Teq; *T. b. brucei*, *T. b. rhodesiense* and *T. b. gambiense* isolates are marked as Tbbru, Tbrho and Tbgam, respectively. Residues essential for the function of SRA are shadowed.

coiled coil formed by two long α-helices, surface loops and the C-terminal region) is well known. The predicted tertiary structure of the SRA protein is surprisingly similar to VSG but lacks the surface loops[11]. We have translated the *SRAbc* sequences into amino acid sequence by removing frame shifts and premature stop codons (data not shown). The predicted 3D structure of the SRAbc amino acid sequences revealed all domains characteristic for VSG[17]. The signal peptide (residues 1—31) and two long α-helices (residues 32—81 and 87—139) are conserved in all isolates (Figure 2), whereas the surface loops are highly variable. Furthermore, the model predicted disulfidic bond between residues 149—211 that is conserved between the SRAbc and SRA (represented here by AF097331) (Figure 2). A comparison of their nucleotide sequences showed that both genes can be subdivided into three regions. The regions A (496—497 bp) and C (436—440 bp) share about 70% (64%—78%) and 76% (72%—79%) homology with the *SRA* gene, respectively. However, the region B is substantially variable in sequence and length, since it comprises 470—628 bp in *SRAbc* but only 174—286 bp in *SRA*. This result confirmed previous observations based on a narrower dataset[12].

A total of eight sequences similar to the *SRA* gene, termed *SRAbc*, of which two share high similarity with the former gene, have been found in the *T. b. brucei* genome[12]. However, no transcripts have been detected for these two *SRAbc* gene and since their predicted reading frames contain frameshifts, it has been proposed that they lack any function[12]. Yet, since in the present study *SRAbc* gene was found in 17 isolates representing all *Trypanozoon* species with the average 82% nucleotide identity, it is reasonable to assume that due to its con-

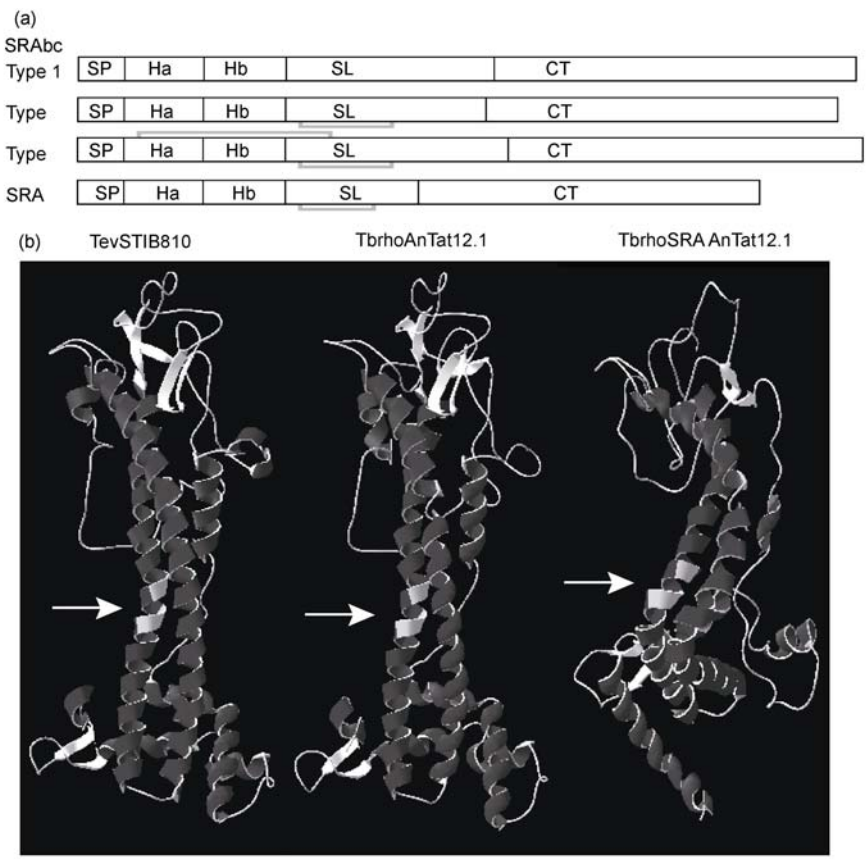


Figure 2 2D and 3D putative structure of *SRAbc* and *SRA*. (a) Variant patterns of the 2D structure of *SRAbc* and *SRA*. Signal peptide (SP) and two alpha-helix regions (Ha and Hb) are conserved. High variation in length of the surface loops (SL) and sulfur-sulfur bonds (gray line) is indicated. The length of the C-terminal region (CT) is conserved. Type 1: TevSTIB803, TevCAN86k, TbgamPAKWAITMAP2570, TbrhoAnTat25.1; Type 2: TevSTIB808, TevStockColumbia; TeqBoTat1.1, TeqSTIB841, TbbSTIB940; Type 3: TevSTIB815, TevSTIB810, TevKETRI2479, TeqSTIB818, TeqSTIB842, TbrhoAnTat12.1, TbgamAnTat11.17, TbbNW_001076895; and *SRA*: AF097331. (b) 3D structures of *SRAbc* and *SRA* were calculated by MODELLER (http://salilab.org/modweb/[18]), using 1VSG and 2VSG as models. Alpha-helix regions were in gray, and beta-sheet in white. The three residues of *SRAbc* (58, 61 and 62) shown in arrows may be more exposed in *SRA*, which lost most of the surface loop leading to a bent framework.

servation, this gene retains some functions.

The *SRAbc* gene shares many conserved amino acid residues with the *VSG* terminus, implying a genetic recombination between *SRAbc* and *VSG* in the past. There are two equally plausible relationships between the *SRAbc*, *SRA* and *VSG* genes: i/ *SRAbc* and *SRA* arose from a common *VSG* ancestor gene and evolved in parallel; ii/ *SRA* is derived from *SRAbc*, which in turn originates from a *VSG* ancestor. In our data we find support for the latter possibility. The mechanism of gene conversion, widespread with the *VSG* genes of *T. brucei*, may be invoked also in the case of the *SRAbc* and *SRA* genes. Upon duplication, *VSG* in an active expression site could be replaced with a duplicated copy of an unexpressed donor *VSG*[19] . Since *SRAbc* coexists in the *T. b. rhodesiense* genome with *SRA*, we propose that an ancient *SRAbc* represented the 'donor *VSG*', which was first duplicated and via recombination subsequently inserted into the expression site. This scenario is supported by the evidence that a fragment of 1.7 kb, which is 85% identical to the *SRA* co-transposed region, was found upstream of *SRAbc*[12]. Under natural selection (such as NHS), *SRAbc* mutated into *SRA* following multiple deletions[11,20]. Relieved from evolutionary pressure, the original *SRAbc* accumulated frame shifts and lost the ability to encode *VSG*. Yet the hypothesis of mosaic *VSG* formation[21,22] raises the possibility to reform a functional *SRA* through segmental gene conversion. So far, no sound reasons are at hand that would explain why the extremely important conversion of *SRAbc* into *SRA* occurred only in *T. b. rhodesiense*.

1 Hoare C A. The Trypanosomes of Mammals. Oxford: Blackwell Scientific Publications, 1972
2 Maudlin I, Holmes P H, Miles M A. The Trypanosomiases. Oxfordshire: CABI publishing, 2004
3 Xong H V, Vanhamme L, Chamekh M, et al. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma brucei rhodesiense*. Cell, 1998, 95(6): 839－846[DOI]
4 Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, et al. Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature, 2003, 422(6927): 83－87[DOI]
5 Pays E, Vanhollebeke B, Vanhamme L, et al. The trypanolytic factor of human serum. Nat Rev Microbiol, 2006, 4(6): 477－486[DOI]
6 De Greef C, Imberechts H, Matthyssens G, et al. A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. Mol Biochem Parasitol, 1989, 36(2): 169－176[DOI]
7 De Greef C, Chimfwembe E, Kihang'a Wabacha J, et al. Only the serum-resistant bloodstream forms of *Trypanosoma brucei rhodesiense* express the serum resistance associated (SRA) protein. Ann Soc Belg Med Trop, 1992, 72(1): 13－21
8 Gibson W, Backhouse T, Griffiths A. The human serum resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infect Genet Evol, 2002, 1(3): 207－214[DOI]
9 Radwanska M, Chamekh M, Vanhamme L, et al. The serum resistance associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. Am J Trop Med Hyg, 2002, 67(6): 684－690
10 De Greef C, Hamers R. The serum resistance associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. Mol Biochem Parasitol, 1994, 68(2): 277－284[DOI]
11 Campillo N, Carrington M. The origin of the serum resistance associated (SRA) gene and a model of the structure of the SRA polypeptide from *Trypanosoma rucei rhodesiense*. Mol Biochem Parasitol, 2003, 127(1): 79－84[DOI]
12 Vanhamme L, Renauld H, Lecordier L, et al. The reference *Trypanosoma brucei* strain TREU927/4 contains *T. brucei rhodesiense*-specific SRA sequences, but displays a distinct phenotype of relative resistance to human serum. Mol Biochem Parasitol, 2004, 135(1): 39－47[DOI]
13 Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinfo, 2004, 5(2): 150－163[DOI]
14 Borst P, Fase-Fowler F, Gibson W C. Kinetoplast DNA of *Trypanosoma evansi*. Mol Biochem Parasitol, 1987, 23(1): 31－38[DOI]
15 Li F J, Gasser R B, Zheng J Y, et al. Application of multiple DNA fingerprinting techniques to study the genetic relationships among three members of the subgenus *Trypanozoon* (Protozoa: Trypanosomatidae). Mol Cell Probes, 2005, 19(6): 400－407[DOI]
16 Lai D H, Hashimi H, Lun Z R, et al. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *T. equiperdum* and *T. evansi* are *petite* mutants of *T. brucei*. Proc Natl Acad Sci USA, 2008, 105(6): 1999－2004[DOI]
17 Carrington M, Miller N, Blum M, et al. Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. J Mol Biol, 1991, 221(3): 823－835[DOI]
18 Lambert C, Léonard N, De Bolle X, et al. ESyPred3D: Prediction of proteins 3D structures. Bioinfomatics, 2002, 18(9): 1250－1256[DOI]
19 Donelson J E. Antigenic variation and the Africa trypanosome genome. Acta Trop, 2003, 85(3): 391－404[DOI]
20 Gibson W, Ferris V. Conservation of the genomic location of the human serum resistance associated gene in *Trypanosoma brucei rhodesiense*. Mol Biochem Parasitol, 2003, 130(2): 159－162[DOI]
21 Roth C, Bringaud F, Layden R E, et al. Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. Proc Natl Acad Sci USA, 1989, 86(23): 9375－9379[DOI]
22 Barry J D, Marcello L, Morrison L J, et al. What the genome sequence is revealing about trypanosome antigenic variation. Biochem Soc Transact, 2005, 33(5): 986－989[DOI]