The Evolution of Extracellular Hemoglobins of Annelids, Vestimentiferans, and Pogonophorans*

Received for publication, January 19, 2001, and in revised form, April 2, 2001 Published, JBC Papers in Press, April 6, 2001, DOI 10.1074/jbc.M100557200

Enrico Negrisolo‡§, Alberto Pallavicini‡§, Roberto Barbato¶, Sylvia Dewilde||**, Anna Ghiretti-Magaldi‡‡, Luc Moens||, and Gerolamo Lanfranchi‡§§

From the ‡Dipartimento di Biologia and Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative, Università di Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy, the ¶Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale Amedeo Avogadro, Corso Borsalino 54 I-15100 Alessandria, Italy, and the ‡Dipartimento di Biologia and Consiglio Nazionale delle Ricerche Center for the Study of Metalloproteins, Università di Padova, via Ugo Bassi 58/B, 35131 Padova, Italy, university of Antwerp UIA, B-2610 Wilrijk, Belgium

The evolution of extracellular hemoglobins of annelids, vestimentiferans, and pogonophorans was investigated by applying cladistic and distance-based approaches to reconstruct the phylogenetic relationships of this group of respiratory pigments. We performed this study using the aligned sequences of globin and linker chains that are the constituents of these complex molecules. Three novel globin and two novel linker chains of Sabella spallanzanii described in an accompanying paper (Pallavicini, A., Negrisolo, E., Barbato, R., Dewilde, S., Ghiretti-Magaldi, A., Moens, L., and Lanfranchi, G. (2001) J. Biol. Chem. 276, 26384-26390) were also included. Our results allowed us to test previous hypotheses on the evolutionary pathways of these proteins and to formulate a new most parsimonious model of molecular evolution. According to this novel model, the genes coding for the polypeptides forming these composite molecules were already present in the common ancestor of annelids, vestimentiferans, and pogonophorans.

The extracellular hemoglobins of Annelida, Vestimentifera, and Pogonophora together with the chlorocruorin, a variant extracellular hemoglobin restricted to four polychaete families, have been recognized to form a unique group according to their physicochemical properties (1). These molecules are complexes of a large number of polypeptide components that assemble into hierarchical ordered quaternary structures.

The extracellular Hbs and Chls¹ of Annelida are formed by two kinds of polypeptide chains, the globins and the linkers (2). Vestimentifera have two types of extracellular Hbs (2) that are indicated here as heavy hemoglobin (Hb_H) and light hemoglobin (Hb_L). Hb_H is very similar in shape and molecular mass (~ 3,000 kDa) to the Hbs and Chls of Annelida and is also composed of globin and linker chains. In contrast, the Hb_L has a molecular mass of ~ 400 kDa and contains globin chains only. Pogonophora contain an Hb very similar to the Hb_L of Vestimentifera; therefore, it is also indicated here as Hb_L (2).

In their pioneering work on the molecular evolution of extracellular Hbs, Gotoh et al. (3) divided the globin polypeptides into two groups with a common origin. Successively, Suzuki and Riggs (4) demonstrated that linker chains share similarities with the low density lipoprotein receptor and that they cannot be related to the globins because of the high divergence of their primary structures. More recently, Yuasa et al. (5) have proposed a general evolutionary model for the Hbs, Hb_H, and Hb_L. According to these authors the common ancestor of all these molecules was a protein formed by globin chains only. During the successive evolution the linker chains would have been added to the final structure of Hbs and Hb_H. Chls were not included in their analysis, because no sequences were available at that time. Moreover, their model implies an independent evolution of Hbs and Hb_H. As a consequence two points still remain unsettled: (i) the placement of Chl in this scenario and (ii) the evolutionary pathway followed by Hbs, Chls, and Hb_H with respect to Hb_L.

To investigate these points we have sequenced the globin and linker chains of Chl of the polychaete *Sabella spallanzanii* (6). They have been aligned with other available sequences of Annelida, Vestimentifera, and Pogonophora to perform phylogenetic analyses applying both cladistic and distance-based methods. Using this approach we have been able to retrace the molecular evolution of Hbs, Chls, Hb_Hs, and Hb_Ls. The comparison of our new phylogenetic reconstruction with the previously advanced hypotheses (3, 5) lead us to propose an alternative model for the evolution of annelid hemoglobins.

MATERIALS AND METHODS

Sequence Alignment

We aligned all the sequences of extracellular Hbs, Chls, Hb_Hs, and Hb_Ls of Annelida, Pogonophora, and Vestimentifera available in the literature including the newly presented globin and linker sequences of *S. spallanzanii* (6). Globin sequences were aligned manually according to the nonvertebrate globin template (7), and linker sequences were aligned using the ClustalW program (8).

Pairwise Comparisons

^{*} This work was supported in part by the Italian Ministry of University and of Scientific and Technological Research MURST Grant-Cofinanziamento Protocol 9805192993-002 (to A. G.-M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^{\$} Both authors contributed equally to this work.

 $[\]ast\ast$ Post-doctoral fellow of the Fund for Scientific Research-Flanders (FWO).

^{§§} To whom correspondence should be addressed. Tel.: 39-04982-76221; Fax: 39-04982-76280; E-mail: lanfra@cribi.unipd.it.

 $^{^1}$ The abbreviations used are: Chl, chlorocruorin; Hb_H, heavy hemoglobin; Hb_L, light hemoglobin; UTR, untranslated region; NJ, neighborjoining; SC, strict consensus.

Pairwise comparisons were performed on the coding and 3'-UTR portions of the *S. spallanzanii* globin cDNAs and some selected globins of Annelida using the program ALIGN. Similar comparisons were carried out on the linker sequences using the ClustalW program (8).

Evolution of Annelid Hemoglobins

	1 2		3	4	5		6	7		8
	12345678901234567890	1234	567890123	34567890123	45678901	234567	89012345678	39012	345678	901234567
	1 5 10	1	15	1 5 1	0 15 1	5	1	5	1 5	10
Globin fold	AAAAAAAAAA	aaA	AaA	BBBBBBBBb	bBBbbBBC	CCCCC	DDDD	DE	DEEEeEl	EEeEEeeEe
Scapharca I	PSVYDAAAQLTADVKKDLRD	SWK-	VIGS	DKKGNGVAL	MTTLFADN	QETIGY	KRL-GNVS	SQGM-	ANDKLR	GHSITLMYA
Scapharca IIA	VADAVAKVCGSEAIKANLRR	SWG-	VLSA	DIEATGLML	MSNLFTLR	DTKTY	TRL-GDV(KG-K	ANSKLR	GHAITLTYA
Urechis FI	GLTAAOIKAIOD	HWFL	NIKG	CLQAAADSI	FFKYLTAY	GDLAF	HKFSSVPL	YGLF	SNPAYK	AQTLTVINY
Glycera MH4	GLSAAOROVVAS	TWK-	DIAGSD	NGAGVGKEC	FTKFLSAH	HDMAAV	G-FSGASD		PGVA	DLGAKVLAC
Glycera mmc	GLSAAOROVIAA	TWK-	DTAGAD	NGAGVGKDC	LIKFLSAH	OMAAV	G-FSGASD		PGVA	ALGAKVLAC
Glycera Pl	HITADOVAALKA	SOP-	EVSAGD	GGAOLGLEM	FTKYFHEN	OMMET	G-YSGRTEA-	T.K	HSSKLO	HEGKVTTDC
Glycera P2	PLTADOVAALKA	SMP-	EVSAGD	GGGOLGLEL	FTKYFHEN	OMMET	G-YSGRTDA-	T.B	HNAKLO	NHGKVTTDC
Clucora P2		COD-1	EVEACD	-CCAOLCIEM	FTDVFDFN	OMMEU	C-VSCRTSA.	T.K	UNCKLO	NCKTTVHC
Giycela PS	HLIADQVAALKA	D P	OCVC EN	ENDAREDDI	FORT ENTO	CODAT	C	DDM	IN CANET	ALCI DUTCA
The appropriate the second sec	TTO CINCLEQIEVRI		QUVE CE	COMDENTD	FUNEED	DDDCT	N	DDI	VCDF	AMURICAC
Tytorrhynchus I (AI)	IDCGILQRIKVK	Q A-	QVISUGE-	SKIDFAID	VENNEERI	PDRSL	NRVNG-	DIAV	ISPE	ANNOVERC
Tylorrnynchus IIA (A2)	SSDHOGPLQRLKVK	QWA-	KAIGVGH	ERVELGIAL	WKSMFAQD	DARDL	KRVHG-	EDV	HSPALE	APPARIVENC
Lamellibrachia BI (Al)	DONILQRLKVKM	IQWA-	KAYGFG	-AERAKFGNSL	WISIFNYA	DAREL	DSVKS-	KEM	IQS PQ K	AHVARVIGO
Oligobrachia a5 (A2)	DCTSLNRLLVKR	QWA-	EAYG-EG	-TSRELLGNRI	WEDLFANM	DARGL	SRVNG-	NDI	DSSEPQ	ANSLEVLGO
Lamellibrachia AIII (A2)	YECGPLQRLKVKR	QWA-	EAYGSGN	DREEFGHEI	WTHVFKDA	SARDL	KRVRG-	DNI	HIPAPR.	AHATRVLGO
Riftia b (A2)	DYVCGPLQRLKVKR	QWA-1	EAYGSGN	SREEFGHEI	WSHVFQHS	AARDM	KRVRG-	DNI	HTPAFR	AHATRVLGO
Pontodrilus I (A1)	DCDVLARFRVKH	IQWQ-1	EVFLGSN	RMEFSHDL	WKEFFHDH	SDLVAL	KRVHG-	ENI	NSPQ	AHGIRVLAG
Pheretima sie. I (A1)	DONTLKRFKVKH	IQNQ-	QVFSGEH	HRTEFSLHF	WKEFLHDH	DLVSL	KRVQG-	ENI	YSPEPQ	AHGIRVLAG
Pheretima com. I (A1)	DCDVLERFKVKH	QWQ-	AVFSDA	HRTDFSLHF	WKEFLHDH	DLVGL	KRVNG-	DNI	YSPEFQ	AHGIRVLAG
Pheretima hig. I (A1)	ECDVLERFKVKH	IQWQ-	TVFS A	HRTEFSLHF	WKEFLHDH	SIVEL	TRVNG-	ANI	YSPEFQ	AHGIRVLAG
Lumbricus d1 (A1)	ECLVTESLKVKL	QWA-	SAFGHAH-	ERVAFGLEL	WRDIIDDH	EIKAP	SRVRG-	DNI	YSPEEG	AHSQRVLSC
Lumbricus d2 (A1)	DONKLEGLKVKL	ONA-	RAFGTAH-	DRLAFGLEL	WKGILREH	EIKEP	GRVRG-	DNI	YSPE	AHSORVLSC
Tubifex I (A1)	ECDALORFKVKH	IOWA-	EAFGTSH-	HRLDFGLKL	WNSIFRDA	EIRGL	KRVDG-	DNA	YSAEFE	AHAERVLGO
Sabella Glb3 (A1)	GONILORLKVKN		EAFGY	ADDRTSLGTAL	RSIIMOK	ESVDK	FKRVNG-	KDI	SSPARO	AHIORVEGO
Lumbricus II (A2)	KKOCGVLEGLKVKS	ENG-	RAYGSGH-	DREAFSOAT	RATFAOV	ESRSI	KRVHG-	DDT	SHPAT	AHAERVIGO
Lamellibrachia BTV (B1)	SKEOSEGDATIVIK	OWN-	OTYNAGTS	AGSRLTMONKT	FSTLEKLK	ESEAL	SNVNV-	ANN	ISSGATH	AHTVEVISO
Lumbrique III (B1)	DENENCOSEEDNETVOK	D-D-	TIWDDTEC	SKIKICECPIL	LTKLAKDT	DUNDI	KRVDT-	FUZ	FCPK	ALATETING
Tulonnhunchus TID (D2)		NE	CIWEDIES.	TOPNATCOAT	FOFTFALD	NAKCU	C RVNIV	DKE	CENDAR	AUTOUTNO
Tyrorinynchus IIB (B2)	DDCCSAADRHEVED	T D	GIWSAE-F.	I GRAVAI GUAL	FUELFALL	CARKCI	GKVNV-	DDT	UCDENE	AUTINU
Tytorrhynchus IIC (BI)	DICCSIEDRREVQA	L R-	SIWSAED-	IGRRILIGRLL	FEELFEID	SAINGL	RRVNV-	DDI	HSPEER.	ANVLINVING
Lumbricus IV (BZ)	ADDEDCOSYEDRREIRH	ITMD-	DVWSSS-F.	PDRRVALVRAV	FDDLFKHI	TSKAL	EKAKI-	DEF	ESGER	SHLVRVANG
Oligobrachia c (B2)	SSCOSNRDRANVMH	INWD-	AAWSAA-Y	SDRRVALAQAV	FASLFSRD	AAAQGL	SGVSA-	DNE	DSAD R	AHCVRVVNG
Lamellibrachia BII (B2)	SSNSCTTEDRREMQL	MMA-	NVWSAQ-F.	IGRRLAIAQAV	FKDLFAHV	AAIGL	DRVHG-	TDV	NSNEFK	ANCIRVING
Sabellastarte E (B2)	ESCCSMEDRQEVLK	AWE -	TMWSAE-F.	IGRRVTIAQEV	FDGLFKKN	AATKDL	KNVNV-	DDE	SSAEFR	AHCVEVING
Sabella Glb1 (B2)	EGCCSMEDRQEVLN	AWE -	ALWSAE-Y	IGRRVMIAQAA	FQKLFEKA	DSKAL	TRVNV-	DNI	GSPQFR	AHCIRVING
Sabella Glb2 (B2)	EACCSMEDRQEVLK	AWE -	AMWSAE-F	IGRRVIIAQEV	FQRLFEKS	ETKEL	TGVNV-	ANI	DSPERR	AHCVEVTNO
	*	*					* *			* *
	9 10	11	1:	2 13		14	15	1	.6	17
	9 10 89012345678901234567	11 8901	1: 234567890	2 13 01234567890	12345678	14. 901234	15 5678901234!	1 56789	.6 012345	17 678901234
	9 10 89012345678901234567 15 1	11 8901 5	1: 234567890	2 13 01234567890 1 5 1	12345678 0 15	14. 901234	15 56789012345 1 5	1 56789 10	.6 012345 15	17 678901234 20 25
Globin fold	9 10 89012345678901234567 15 1 eEEee FFF	11 8901 5 ffFF	1: 234567890 FF	2 13 01234567890 1 5 1 GGGgGGgGG	12345678 0 15 GgGGggGG	14. 901234 3G	15 56789012345 1 5 HHHHhHI	1 56789 10 HhHHh	.6 012345 15 ЪННННЫ	17 678901234 20 25 HhHHHHH
Globin fold Scapharca I	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK	11 8901 5 ffFF FAVN	1: 234567890 FF ITRI	2 13 01234567890 1 5 1 GGGgGGGGGG KISAAEFGKIN	12345678 0 15 GgGGggGGG GPIKKVLA	14 901234 GG SKN-	15 56789012345 1 5 HHHHhH FGDKYANA	1 56789 10 HhHHh	.6 012345 15 hHHHHH VAVVQA	17 678901234 20 25 HhHHHHH AL
Globin fold Scapharca I Scapharca IIA	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK	11 8901 5 ffff FAVN	1: 234567890 FF ITRI INRI	2 13 01234567890 1 5 1 GGGgGGgGG KISAAEFGKIN KISGDAFGAIV	12345678 0 15 GgGGggGG GPIKKVLA EPMKETLK	14 901234 3G SKN- ARMGN-	15 56789012345 1 5 HHHHhH FGDKYANA YYSDDVAGA	1 56789 10 IhHHh AMAKI	.6 012345 15 hHHHHH VAVVQA	17 678901234 20 25 HhHHHHH AL AL
Globin fold Scapharca I Scapharca IIA Urechis FI	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA	11 8901 5 FfFF FAVN FAVN	1: 234567890 FF HITR HINRI HDAMG	2 13 01234567890 1 5 1 GGGgGGgGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL	12345678 0 15 GgGGggGG GPIKKVLA EPMKETLK KLVGGVFQ	14 901234 SG SKN- ARMGN- EEF-SA	15 56789012343 1 5 HHHHbHH FGDKYANA YYSDDVAGA DPTTVAA	1 56789 10 IhHHh AWAKI AWAAI	.6 012345 15 hhhhhhh VAVVQAJ VGVVQAJ AGVLVAJ	17 678901234 20 25 HhHHHHH AL AL AMK
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAEMKA	11 8901 5 FfFF FAVN FAVN KVPS VGVR	1: 234567890 FF HITR	2 13 01234567890 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAEYFEPLG	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETLKI KLVGGVFQI ASLLSAMEI	14 901234 SKN- ARMGN- EEF-SA HRIGGK	15 56789012345 1 5 HHHHHH FGDKYANA YYSDDVAGZ DPTTVAZ MNAAAKDZ	1 56789 10 Ihhhh WAKI WAAI WAAI	.6 012345 15 NHHHHH VAVVQAJ VGVVQAJ AGVLVAJ YADISG	17 678901234 20 25 HhHHHHH AL AL AMK ALISGLQS-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAEMKA IGVAVSHLGDEGKMVAEMKA	11 8901 5 FAVN FAVN KVPS VGVR	12 234567890 FF HIRR	2 13 01234567890 1 5 1 GGGgGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAOYFEPLG HIKAOYFEPLG	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLKI KLVGGVFQI ASLLSAMEI ASLLSAMEI	14 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK	15 56789012345 1 5 HHHHHH FGDKYANJ YYSDDVAGJ DPTTVAJ MNAAAKDJ MNAAAKDJ	1 56789 10 HhHHh AWAKI AWAAI AWGDA AWAAA	.6 012345 15 hhhhhhi VAVVQA VGVVQA AGVLVA AGVLVA YADISG	17 678901234 20 25 HhHHHHH AL AL AMK ALISGLQS- ALISGLOS-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera Pl	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAEMKA IGVAVSHLGDEGKMVAEMKA	11 5 2ffff FAVN FAVN KVPS VGVR VGVR	1: 234567890 FF HITRI HDAMG HKGYGNK-I HKGYGNK-I HKGFG	2 13 01234567890 1 5 1 GGGgGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAEYFEPLG HIKAQYFEPLG DIRAEFFPALG	12345678 0 15 GgGGggGGG GPIKKVLA EPMKETLKX KLVGGVFQU ASLLSAMEI ASLLSAMEI MCLLDAMEI	14 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG-	15 5678901234 1 5 HHHHbHH FGDKYANJ YYSDDVAGJ DPTTVAJ MNAAAKDJ MNAAAKDJ	1 56789 10 IhHHh WAKI WAAI WAAA WAAA WAAA	6 012345 15 hhhhhhi VAVVQA VGVVQA AGVLVA XADISG XADISG	17 678901234 20 25 HhHHHHH AL AL ALISGLQS- ALISGLQS- ALISGLQS-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera P1 Glycera P2	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGKAVAEMDNAKQMAGTLHA	11 8901 5 FAVN FAVN KVPS VGVR VGVR LGVR	12 234567890 FF HITRI HDAMG HKGYGNK-I HKGYGNK-I HKGFG1	2 13 01234567890 1 5 1 GGGgGGgGG KISAAEFGKIN KISGDAFGAIV HIKAEYFEPLG HIKAQYFEPLG DIRAEFFPALG DIRAEFFPALG	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLK KLVGGVFQ ASLLSAMEI ASLLSAMEI MCLLDAMEI MCLLDAMEI	14 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK EKVPG-	15 5678901234! 1 5 HHHHbH FGDKYANX YYSDDVAGZ DPTTVAZ MNAAAKDZ MNAAAKDZ NNAAKDZ	1 56789 10 Ihhhh WAAI WAAI WAAA WAAA WAAA	6 012345 15 hhhhhhi VAVVQA VGVVQA VGVVQA VADISG VADISG VADISG VREISD	17 678901234 20 25 HhHHHHH AL AL ALISGLQS- ALISGLQS- ALIAGLQS- ALIAGLQS- ALIAGLQS-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LDNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGKAVAEMDNAKQMAGTLHA IGCAVESLDDCSKFETATLHK	11 8901 5 FAVN FAVN KVPS VGVR LVGVR LGVR	12 234567890 FF HITR HDAMG HKGYGNK HKGYGNK HKGFG	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG	12345678 0 15 GgGGggGGG GPIKKVLA EPMKETLK KLVGGVFQ ASLLSAME ASLLSAME MCLLDAME MCLLDAME	14. 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- SKVHG-	15 5678901234! 1 5 HHHHHH FGDKYAN/ YSDDVAGG DPTTVA/ MNAAKD/ NRAAKD/ LNRTI LNRTI	1 56789 10 Ihhhh WAKI WAAI WAAA WAAA WAAA WAAA	.6 012345 15 hhhhhh VAVVQA VGVVQA AGVLVA YADISG YADISG YREISD YREISD	17 678901234 20 25 HhHHHHH AL AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLQS- ALIAGLES-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligoprachia b (M1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAEMKA IGVAVSHLGDEGKMVAEMKA IGKAVAEMDNAKQMAGTLHA IGKAVAEMDNAKQMAGTLHA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDQSKFEATLHK LNELISDLDQOATINVSLA	11 8901 5 ffff FAVN FAVN FAVN KVPS VGVR LGVR LGVR LGVR	1: 234567890 FF HITRI HDAMG HKGYGNK-I HKGYGNK-I HKGFGI HKGFG HKGFG	2 13 01234567890 1 5 1 GGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAQYFEPLG HIKAQYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEYFPALG DIKGEYFPALG	12345678 0 15 GGGGgGGG GPIKKVLA EPMKETLK KLVGGVFQ ASLLSAME MCLLDAME MCLLDAME DALLEAMN OAVMSVVP	14. 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- SKVHG- FH1-D	15 56789012349 1 5 HHHHHH FGDKYANA YYSDDVAGA DPTTVAA MNAAAKDA MNAAAKDA NRTI LNRTI LNRTI	1 56789 10 Ihhhh WAAI WAAI WAAA WAAA WAAA WAAA	.6 012345 15 hhhhhhi VAVVQAJ VGVVQAJ AGVLVA YADISG YADISG YRVISD YRVISD	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLQS- ALIAGLES- GISG
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Twlorrhynchus I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDNVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA EDLIJSVLDDKPUDQDA	11 8901 5 ffFF FAVN FAVN FAVN KVPS VGVR LGVR LGVR LGVR LGVR	1: 234567890 FF HITRI HDAMG HKGYGNK-I HKGYGNK-I HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1	2 13 D1234567890 1 5 1 GGGgGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAEYFPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIKEEYFPALG DILGEYFPALG	12345678 0 15 GgGGgGGG GPIKKVLA EPMKETLKJ KLVGGVFQJ ASLLSAMEJ MCLLDAMEJ DALLEAMN QAVMSVVP T	14. 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- SKVHG- FHLD FHLD	15 5678901234 1 FGDKYANJ FGDKYANJ PTTVAJ DPTTVAJ DPTTVAJ MNAAAKDJ MNAAAKDJ INRTI INRTI INRTI CFNQHJ	1 56789 10 Ihhhh WAXI WAAI WAAA WAAA WAAA WAAA WAAA WAAA	.6 012345 15 hhHHHHH VAVVQAJ VGVVQAJ VAVSA YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD	17 678901234 20 25 HhHHHHHH AL ALISGLQS- ALISGLQS- ALIAGLQS- ALIAGLES- GISG GISG
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKPVLDQALA IDEVISSITDEPUN MADLE	11 8901 5 FfFF FAVN FAVN FAVN KVPS VGVR LGVR LGVR LGVR LGVR LGVR LAG LAG	1234567890 FF HITR HCANG HKGYGNK-I HKGFG HKGFG HKGFG HKGFG HKGFG HKGFG	2 13 01234567890 1 5 1 GGGgGGgGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIKGEYFPALG NLDASNFAAMG FIPFKAFG	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLK: KLVGGVFQU ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFVTIAI	14. 901234 SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK EKVPG- EKVPG- SKVHG- THLD EHI	15 5678901234! 1 5 HHHHbHI FGDKYANJ YSDDVAGJ DPTTVAJ MNAAAKDJ MNAAKDJ INRTI LDRTI CFNQHJ HGADIGJ	1 56789 10 HhHH AMAKI AMAA AMAA AMAA AMAA AMAA AMAA AMA	6 012345 15 NHHHHI VAVVQA AGVLVA YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALIAGLQS- ALIAGLES- GISG TGITA FGITA
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P2 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia PI (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKPVLDQALA LDRVISSLTDEPVLNAQLE LDRVISSLTDEPVLNAQLE	11 8901 5 FfFF FAVN FAVN FAVN KVPS VGVR LGVR LGVR LGVR LGVR LGVR LAG LAG LAG LAG	1: 23456789(FF HITR HCAYGG HKGYGNK HKGFG HKGFG HKGFG HKGFG HKQFG HASR HIKL(DPP	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIKGEYFPALG DIKGEYFPALG SITCHMFNLMR SITCHMFNLMR	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETIKK KLVGGVFQU ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLP; VALEATIOC	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK EKVPG- SKVHG- FHLD EHI AQLG-R SKPC	15 5678901234! 1 5 HHHHHH FGDKYAN/ YSDDVAGG DPTTVA/ MNAAAKD/ LNRTI LNRTI LDRTI CFNCH/ HGADIG/ CFDKEJ	1 56789 10 HhHH AMAKI AMAA AMAA AMAA AMAA AMAA AMAA AMA	6 012345 15 hhHHHHI VGVVQA VGVVQA VGVVQA YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD YREISD	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLQS- GISG GISG FGITA PGIKHD
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDGSKFEATLHK LNRLISQLDDQATINVSLA FDILISVLDDKPVLDQALAF LDRVISSLTDEPVLNAQLEF LDRVISSLTDEPVLNAQLEF LDRVISSLDDEPVLNAQLEF	11 8901 5 FfFF FAVN FAVN KVPS VGVR LGVR LGVR LGVR LGVR LAGO LAGO LAGO	1: 234567890 FF HITR HKGYGNK HKGYGNK HKGFG HKGFG HKGFG HKQFG HXQFG HIKL(HDPR(DPR))))))))))))))))))))))))))))))))	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAQYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRASNFAAMG TIFKAFG GITGHMFNLMR SLDALNFAVFG	12345678 GgGGgGGG GPIKKVLA EPMKETLK KLVGGVFQU ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN QAVMSVVP QTMFQTIAI TGLAYVLP KALFATVGG	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- THLD EHI AQLG-R 3KFG	15 56789012349 1 5 HHHHHH FGDKYANA -YYSDDVAGG DPTTVAA MNAAAKDA INRTI LNRTI LNRTI LNRTI CFNQHA HGADIGA CFDKEA FDKEA	1 56789 10 Ihhhh Awaai Awaai Awaai Awaai Awaai Awaai Awaai Awaai Awaai Awaai Awaai Awaai Awaai	6 012345 15 hHHHHH VAVVQAJ VGVVQA VAVVQAJ VAVVQAJ VAVISA VAVISA YAVISA YAEQIV WDEVIY YKVIAK	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- ALIAGLES- GISG TGITA PGIKHD GITGNDMFN GITSC
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDNVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKPVLDQALA LDRVISSLTDEPVLNALE LDRVISSLTDEPVLNALE LDRVISSLDDPVLNALE	11 8901 5 FFFF FAVN KVPS VGVR LGVR LGVR LGVR LGVR LGVR LGVR LAGO LAGO LAGO LAGO LAGO LAGO	1: 234567890 FF HITRI HDAMG HKGYGNK-I HKGYGNK-I HKGFG1 HKGFG1 HKGFG1 HKQFG	2 13 D1234567890 1 5 1 GGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAEYFPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIKGEYFPALG SITGHMFNLMR GIDALNFAVFG GITGHMFNLMR	12345678 0 15 GgGgggGG GPIKKVLA: EPMKETLKJ KLVGGVFQU ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN: QAVMSVP ² QAVMSVP ² TGLAYVLPJ KALFATVG ASAISAVR	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- SKVHG- THLD EHI AQLG-R SKFG ATVGAR	15 5678901234 1 5 HHHHHH FGDKYANA YSDDVAGA DPTTVAA MNAAKDA MNAAKDA MNAAKDA INRTI INRTI INRTI FDNTA CFDNTA CFDNTA CFDNTA SFDNTA	1 56789 10 Ihhhh AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi AwaAi	6 012345 15 NHHHHI VAVVQAJ VGVQAJ VGVQAJ VAVISG YADISG YREISD YR	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALIAGLES- GISG GITGITA GITGNDMFN GITGNDMFN GISG
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I(A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKPVLDQLA LDRVISMLDNAEALNADLE LDRVISMLDNAEALNADLE LDMCVASLDDVPVLNALLAR	11 8901 5 FFFF FAVN KVPS VGVR LGVR LGVR LGVR LGVR LGVR LGVR LLSV LAG LAG LAG LAG LAG LAG LAG	1234567890 FF HITR1 HDAMG HKGYGNK-1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIKGEYFPALG NLDASNFAAMG FI PFKAFG SITGHMFNLMR SLDALNFAVFG SIDALNFAVFG SIPAAGYPAV	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLK: KLVGGVFQÜ ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIA: TGLAYVLPJ: KALFATVG ASAISAVR: HSVMGVE	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK EKVPG- EKVPG- SKVHG- THLD EHI AQLG-R AQLG-R AXFG ATVGAR.	15 5678901234! 1 5 HHHHHH FGDKYAN/ YSDDVAG/ MNAAAKD/ MNAAAKD/ MNAAAKD/ LDRTT CINRTT LDRTT CFNQH/ HGADIG/ CFNKE/ VCFDLE/ SFDND/ FDND/	1 56789 10 HhHH AWAAI AWAAI AWAAA WAAA WAAA WAAA WAAA	6 012345 15 hHHHHI VAVVQA VGVVQA YADISG YADISG YARISD YERISD YERISD YERISD YERISD YERISC YAEQIV WDEVIY YKVIAK MNQIVS LDVITG	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLSS- GISG FGIKHD GITGNDMFN GISG GI GN
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKVVAQMA LDRVISSLTDEPVLNAQLE LDRVISSLTDEPVLNAQLE LDRVISSLDDPVVLNALLA LDMCIALLDDEGVLNQLA LDMCIALLDDEGVLNQLA	11 8901 5 FFAVN KVPS VGVR LGVR LGVR LGVR LGVR LAGO LAGO LAGO LAGO LAGO LAGO LAGO	12 234567890 FF HITR HDAMG HKGYGNK HKGYGNK HKGFG HKGFG HKGFG HKGFG HXQFG HIKL(HDSR(HSSR(HER)	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG DIKGEYFPALG SITAGFFNLMR SLDALNFAVFG SIPAAGYPAFV GVSAAQYDVVE GVEAAHYDTYN	12345678 0 15 GgGGggGG GPIKKVLA EPMKETLK KLVGGVEQ ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN QAVMSVVP QTMFQTIAI TGLAYVLP KALFATVG KALFATVG KASAISAVR	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK SKVHG- SKVHG- THLD EHI SKVHG-R SKFG AQLG-R SKFG ATVGAR HEIGQN	15 5678901234! 1 5 HHHHHH FGDKYAN/ YSDDVAGG DPTTVA/ DPTTVA/ NNAAKD/ NNAAKD/ INRTI INRTI INRTI CFINRJ/ CFDKD/ SFDND/ FDND/ FDND/ FDND/	1 567899 10 HhHHM WAMAKI WAMAAA WAMAAA WAMAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAA AMAAAAA AMAAAA AMAAAAA AMAAAAA AMAAAAA AMAAAAA AMAAAAA AMAAAAA AMAAAAAA	6 012345 15 hHHHHH VAVVQAJ VGVVQAJ AGVLVA YADISG YRVISD YR	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GIGNDMFN GISG GIGGN GIGGN
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLDDEGKFATLHK LNRLISQLDQQATINVSLA FDILISVLDDKFVLDQALA LDRVISSLTDEFVLNAQLE LDRVISSLTDEFVLNAQLE LDMCIALLDDEGVLNIQLA LDMCIALLDDEFVLNIQLA LDGLCALDEPUTLNQLIV	11 8901 5 ffff FAVN KVPS VGVR LGVR LGVR LGVR LLQV LLQV LLSV LLSV LLSV LLSV LLSV LLSV	12 234567890 FF HITR HDAMG HKGYGNKI HKGYGG HKGFG HKGFG HKQFG HKQFG HKQFG HIKL(HDSR(HDSR(HSSR(HETR(HTER())	2 13 D123456780 1 5 1 GGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAQYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG SITAMFNLMR GIDALNFAVFG GIDAAGYPAFV GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE	12345678 GgGGgGGG GPIKKVLA EPMKETLK KLVGGVFQU ASLLSAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLAMEI MCLLDAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLDAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLLAMEI MCLAM	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- SKVPG- SKVHG- THLD SKVG- AQLG-R SKFG- ATVGAR HEIGQN VVIGSE DHLG	15 56789012349 15 HHHHHH FGDKYANA YSDDVAGG DPTTVAA MNAAAKDA INRTI LNRTI LNRTI LNRTI CFNQHA FDQA CFDQA SFDNDA SFDNDA SFDNDA FDNDA FDNDA 	1 567899 10 HHHH AMAKI AMAAA AMAAA AMAAA AMAAA AMAAA AMAA AM	6 012345 15 hHHHHH VAVVQAJ VGVVQA VGVVQA VAVVQAJ VAVIS VADISG VADISG VADIS VADIS VADIS VERIAS VERIAS VERIAS VERIAS VERIAS VERIAS VALUVIS VIVIAN VAVIAN	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- ALIAGLES- GISG GISG GIGNDMFN GISG GIGN GIGG GIGC GIGC
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGKAVAEMDNAKQMAGTLHA IGQAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKPVLDQALA LDRVISSLTDEPVLNAQLE LDMCIALLDDEGVLNQLA LDMCIALLDDEFVLNQLA LDMCIALLDDEFVLNQLA LDSVIGVLD	11 5 fffff FFAVN FFAVN FFAVN KVPS LGQR LLGQE LLGQE LLGQE LLAGO LLSSO LLSSO LLSSO LLSSO LLSSO LLSSO	1234567890 FF HITRI HDAMG HKGYGNK-I HKGFG1 HKGFG1 HKGFG1 HKGFG HIKL(HDR(HDR(HETR(HTER(HTER(HTER(2 13 01234567890 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV HIKAEYFEPLG HIKAQYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIKGEYFPALG SITGHMFNLMR SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SVEAAHYDTVN STKPEYFDIFG STKPEYFDLFG	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLKX KLVGGVFQI ASLLSAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLPJ TGLAYVLPJ KALFATVGG ASAISAVR HSVMGGVEI VHLLRILD TQLFDILG	14. 901234 SK-NN ARMGN EEF-SA HRIGGK EKVPG EKVPG EKVPG EKVPG EHID EHID EHID AQLG-R SKFG PHIG EHIG	15 5678901234! 1 5 HHHHbH FGDKYANJ YYSDDVAGJ DNVAGJ MNAAAKDJ NNAAKDJ INRTI INRTI INRTI CFNQHJ FNQDJ CFNEJ SFDKEJ -VCFDLPJ SFDKDJ -VFDKDJ FDKDJ FDKDJ FDKDJ FDKDJ 	1 10 10 10 10 10 10 10 10 10 10 10 10 10	6 012345 15 hHHHHI VAVVQA VGVVQA YADISG YADISG YALSD YREISD YREISD YREISD YREISD YREISD YREISD YRUSD YERAS YAEQIV WDEVIY YKVIAK MNQIVS LDVITG LDVITG LNVITM YAVIAA	17 678901234 20 25 HhHHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GISG GITGNDMFN GISG GIGN GING GIKG GIKB
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA LDKVISULDQQATINVSLA FDILISVLDDKPVLDQLA LDRVISMLDNAEALNADLE LDMCIALLDDEGVLNIQLA LDMCIALLDDEGVLNIQLA LDGLCALDE DTTINQLIV LDSVIGVLD DDTFTVQLA	11 8901 5 FIFFF FAVN KVPS LGVR LGVR LGVR LGVR LLQV LLQV LLQU LLQU LLQU LLQU LLQU LLQU	12 234567890 FF HITRI HDAMG HKGYGNK-I HKGFGI HKGFGI HKGFGI HKGFGI HKGFGI HKGFG HSSR(HDSR(HDSR(HTER(HTER(HTER(HTER(HTER(HTER())	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG SITGHMFNLMR SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SIDAALYDTVN SVEAAAYDTVN STKPEYFDLFG STKPEYFDLFG	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETLK: KLVGGVFQI ASLLSAMEI ASLLSAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLP ³ QTMFQTIAI TGLAYVLP ³ MAVMSVE HAVMSVE HAVMSVE VHLLRILD KHLADHLG KHLADHLG	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- THLD PHLD CHI AQLG-R SKOG ATVGAR HEIGQN WVIGSE HLG KLG KLG	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YSDDVAG/ DPTTVA/ NNAAKD/ NNAAKD/ NNAAKD/ INRTI INRTI INRTI CFNRT/ CFNRD/ FDKD/ VFDKD/ VFDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ 	1 566789 10 1hHhh Amakii Amada	6 012345 15 hHHHHI VAVVQAJ VGVVQAJ AGVLVAJ YADISG YADISG YREISD YREISD YERIASI YAEQIV YRVIAK YAQUISI LOVITG LOVITG LOVITG LOVITG YAVIAA YAVIAA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISGN
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA LDRVISLDDGEKFEATLHK LDRVISSLTDEPVLNAGLE LDRVISSLTDEPVLNAGLE LDRVISSLDDEPVLNAGLE LDMCIALLDDEGVLN QLA LDMCIALLDDE PVLN QLA LDMCIALLDDE PVLN QLA LDSVIGVLD BDTFNVQLA LDSVIGVLD BDTFNVQLA LDSVIGVLD BDTFNVQLA	11 8901 5 fffF FAVN KVPS VGVR LGVR LGVR LGVR LGVR LAG0 LAG0 LAS0 LNS0 LNS0 LAS0 LKS0 LKA0	12 234567890 FF HITR	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG SITGHMFNLMR SLDALNFAVE GUEAAHYDTVN GVEAAHYDTVN GTKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETLK: KLVGGVEQ ASLLSAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI QAVMSVVP QTMFQTIAI TGLAYVLP: KALFATVG KALFATVG KALFATVG VHLLFILD TQLFDILG KHLASHLG KHLASHLG	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK SKVHG- SKVHG- SKVHG- SKFG AQLG-R SKFG ATVGAR HEIGQN NVIGSE DHLG SKLG SKLG SKLG	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YSDDVAGG DPTTVA/ DPTTVA/ NRAAKD/ INRTI INRTI INRTI INRTI CFINR/ FINR/ FINR/ SFINR/ SFINR/ FINR/ FINR/ SFINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ FINR/ 	1 567393 10 HhHHH AvaAi AvaAa A AvaAa AvaAa AvaAa AvaAa AvaAa AvaAa AvaAa AvaAa AvaA	6 012345 15 hHHHHH VAVVQAJ VGVVQAJ AGVLVAJ YADISG YRVISD YREISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YAEQIV WDEVIY YNVIAK MNQIVS LDVITG LNVITM YAVIAA YAVIAA YDFIAS	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GIGNDMFN- GISG GIGN GIGN GIKB GIKB GIKB
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Lumbricus dl (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LNKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKINVAQLA LDRVISSLTDEFVLNAQLE LDRVISSLTDEFVLNAQLE LDRVISSLTDEFVLNAQLE LDRVISSLTDEFVLNAQLE LDMCIALLDDEGVLNVQLA LDGLCALDED DTTTVQLA LDSVIGVLD EDTFTVQLA LDSVIGVLD EDTFTVQLA LDSVIGVLD EDTFTVQLA LDSVIGVLD EDTFTVQLA	11 8901 5 FFAVN KVPS VGVR LGVR LGVR LGVR LGQE LASO LASO LKSO LKSO LKSO LKSO LKSO LKSO	12 234567890 FF HITR	2 13 D123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG DIRGEYFPALG SITGHMFNLMR SLDANFAVFG SIDAAGYPAFV GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE GTKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG	12345678 0 15 GGGGgGGGG GPIKKVLA EPMKETLK KLVGGVFQU ASLLSAMEI ASLLSAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI ASLSAMEI ASLSAMEI HAVMGVEI VHLLRILD TQLFDILG KHLADHLG KHLAHLG KHLAHLG	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- SKVPG- SKVPG- AQLG-R SKFG-A AQLG-R SKFG-A ATVGAR HEIGQN VVIGSE DHLG- CKLG CKLG CKLG CKLG	15 56789012349 15 HHHHHH FGDKYAN/ YSDDVAGG DPTTVA/ MNAAAKD/ INRTI LNRTI LNRTI LNRTI CFNCH/ HGADIG/ CFDKD/ SFDKD/ SFDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ 	1 56789 10 Hhithi Awaki	6 012345 15 hHHHHH VAVVQAJ VGVVQAJ GVVVQAJ VGVVQAJ VGVVQAJ YADISG YAVISD YRVISD YRVISD YRVISD YERIAS YAEQIV WDEVIY YKVIAK MNQIVS LDVITG LDVITG LDVITM YAVIAA YAVIAA YAVIAA YDFIAS VDQIID	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLSS- GISG GISG GIGN GIGG GIGG GIK GIK GIK GIK GIK
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia AIII (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKNVAQMKA IGVAVSHLGDEGKNVAQMKA IGVAVSLDDQATINVSLA FDILISVLDDKPVLDQALA LDRVISSLTDEPVLNAQLE LDMCIALLDDEGVLNQLA LDMCIALLDDEFVLNQLA LDMCIALLDDEFVLNQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD IFTLVQLA LDSVIGVLD IFTLVQLA LDTISMLD PDMLTAQLA	11 8901 5 FFFF FAVN KVPS VGVR LGVR LGVR LGVC LASO LASO LKSO LKSO	1234567890 FF HITR1 HDAMG HKGYGNK-1 HKGYGNK-1 HKGFG1 HKGFG1 HKGFG1 HKGFG HIKL HDPR HDPR HDPR HETR HTER HTER HTER HTER HVER1 HVER1 HVER1	2 13 01234567890 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV ITPKHFGQLL HIKAQYFEPLG DIRADFFPALG DIRADFFPALG DIKGEYFPALG SITGHMFNLMR SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLKV ASLLSAMEI ASLLSAMEI MCLLDAMEI DALLEAMN QAVMSVVP QTMFQTIAI TGLAYVLPJ TGLAYVLPJ TGLAYVLPJ TGLAYVGVEI VHLLRIDD TQLFDILG KHLADILG KHLADILG NHLLEVLG	14. 901234 SKN ARMGN EEF-SA HRIGGK HRIGGK EKVPG EKVPG EKVPG EHID EHID EHID KIG KLG ELG BRLG BRLG BRLG	15 5678901234! 1 5 HHHHbHH FGDKYANJ YYSDDVAGJ DPTTVAJ MNAAKDJ MNAAKDJ INRTI LDRTI CFNQHJ HGADIGJ CFNQHJ GFDLPJ SFDKDJ -VFDKDJ VFDKDJ FDKDJ FDKDJ FDKDJ 	1 10 10 10 10 10 10 10 10 10 10 10 10 10	6 012345 15 hHHHHI VAVVQA VGVVQA YADISG YADISG YADISG YREISD YREISD YREISD YREISD YRVISD YERIAS YAUXA IDVITG LDVITG LDVITG LDVITG LDVITG YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YDFIAS	17 678901234 20 25 HhHHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GISG GIGN GIGN GIGGN GIGG GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK GIK
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I(A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Tubifex I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA LDRVISQLDQQATINVSLA FDILISVLDDKPVLDQLA LDRVISMLDNAEALNADLE LDMCIALLDDEGVLNIQLA LDMCIALLDDEGVLNIQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDTISMLDPDMLAQLA	11 8901 5 FFFF FAVN VGVR VGVR LGVR LLGVR LLGVR LLGVE LLGQE LLSQ LLSQ LKSQ LKSQ LKSQ LKSQ	12 234567890 FF HITRI HDAMG HKGYGNK-I HKGFGI HKGFGI HKGFGI HKGFGI HKGFGI HKGFG HIKL(HSSR(HDSR(HTER(HTER(HTER(HTER(HVER1 HVER1 HVER1 HVER1	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG SITGHMFNLMR GIDALNFAVFG GIDAGYPAFV GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVVE GVSAAQYDVE GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG GTKPEFFDIFL MLKPEFFDIFL NLKPEFFDIFL	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETLK: KLVGGVFQU ASLLSAMEI ASLLSAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLPJ QTMFQTIAI TGLAYVLPJ KALFATVG ASAISAVR: HSVMYGVEI HAVMGVEI VHLLRILD VHLLRILD KHLASHIG KHLASHIG KHLLAVLP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EKVPG- EKVPG- PHLD PHLD THLD EHI AQLG-R SKFG ATVGAR HEIGQN VVIGSE DHLG D KLG D RLG D RLG YLG	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YSDDVAG/ MNAAAKD/ MNAAAKD/ MNAAAKD/ INRTI INRTI INRTI CFNQH/ HGADIG/ CFDKD/ VFDKD/ VFDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ 	1 10 10 10 11 10 10 10 10 10 10 10 10 10	6 012345 15 hHHHHI VAVVQAJ VGVVQAJ AGVLVAJ YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YRVIAK YAVIAJ YAVIAJ YAVIAJ YAVIAJ YAVIAJ YAVIAJ YDQIID INHIID LGVITG	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLQS- GISG
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Lumbricus d2 (A1) Tubifex I (A1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA LDRVISLDDQATINVSLA FDLLISVLDDQATINVSLA FDLLISVLDDKPVLDQALA LDRVISSLTDEPVLNAQLE LDRVISSLDDEPVLNAQLE LDMCIALLDDEGVLNQLA LDMCIALLDDEGVLNQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDEDTFNVQLA LDSVIGVLDDEDTFNVQLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLDEDTFNVGLA LDSVIGVLD	11 8901 5 FFFF FAVN VGVR VGVR LGVR LGVR LGVR LLGVE LLGVE LLGVE LLGVE LLGVE LLGVE LLGVE LLGVE LLGVE LKS0 LKS0 LKS0 LKS0 LKS0 LKS0	12 234567890 FF HITR	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG SITAGFFPALG SITAGFFPALG SITAGFFPALG SITAGFFPALG SITAGFFPALG SIPAGYPALG SIPAGYPALG SIPAGYPALG SIPAGYPALG SIPAGYPALG SIPAGYPALG SIPAGYPALG SIFPEYFDLFG SIFFE SIFFE	12345678 0 15 GgGGggGG GPIKKVLA EPMKETLK KLVGGVEQ ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN QAVMSVVP QTMFQTIAI TGLAYVLP KALFATVG KALFATVG KALFATVG KALFATVG KHLAFILG KHLAFILG KHLAFILG KHLAVLG NHLLEVLG SLMLAVES	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK SKVHG- EKVPG- SKVHG- EHI SKFG AQLG-R SKFG ATVGAR HEIGQN NVIGSE HLG SKLG DELG DRLG DHLG SHL	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YSDDVAGG DPTTVA/ MNAAKD/ NRTI INRTI INRTI INRTI CFINR/ CFINR/ CFINR/ SFINR/ SFINR/ SFDND/ FDND/ FDND/ FDND/ FDND/ FIND/ FIND/ FIND/ FIND/ 	1 10 10 10 11 11 11 11 11 11 11 11 11 11	6 012345 15 hHHHHH VAVVQAJ VGVVQAJ AGVLVAJ YADISG YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVIA YAUIAA YAVIAA YAVIAA YDFIAS VDJID INHID LGVITG LGVITG TKVISS	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISG GIGNDMFN- GISG GIGN GISG GISG GISG GISG GISG GISG GISG GISGSGV GISGSGV
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I(A1) Tylorrhynchus I(A1) Uligobrachia a5 (A2) Lamellibrachia aBI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LNKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKNVAQMKA IGVAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKPVLNALLA LDRVISSLTDEPVLNAQLE LDMCIALLDDEGVLNIQLA LDMCIALLDDEFVLNIQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDSVIGVLD=DTFTVQLA LDIXISMLDDPDLAAQLA LDIXISLLDDQAAFDAQLA FDMCISMLDDSDVLASQLA	11 88901 5 FFAVN KKVPS VGVR LGVR LGVR LGVR LLGVR LLGVC LLRQ LLRQ LLRQ LLRQ LLRQ LLRQ LLRQ LLR	12 34567890 FF HITRI HNRI HXGYGNK-I HKGYGNK-I HKGFG1 HKGFG1 HKGFG1 HKGFG1 HXGFG1 HIKL(HDPR(HDPR(HTER(HTER) HTER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1 HVER1	2 13 0123456780 1 5 1 GGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG SITGHMFNLMR SLDASNFAAMG TIPFKAFG SITGHMFNLMR SLDASNFAAMG TIPFKAFG SITGHMFNLMR SLDALNFAVFG GVEAAHYDTVN STKPEYFDLFG STKPEYFD STKPE ST	12345678 0 15 GGGGgGGGG GPIKKVLA EPMKETLK KLVGGVFQU ASLLSAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI DALLEAMI QAVMSVVP QTMFQTIAI TGLAYVLP KALFATVG ASAISAVG VHLLFATVG NHLLRVLG KHLANHLG KHLANLG KHLLAVLG NHLLEVLG NHLLEVLG NHLLAVLS SLMLAVES TAILHVVA	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK EKVPG- EKVPG- SKVPG- SKVPG- SKVPG- AQLG-R SKFG-A ATVGSE DHLG- DKLG DKC	15 5678901234! 1 5 HHHHHH FGDKYAN/ YSDDVAGG DPTTVA/ MNAAAKD/ LNRTI LNRTI LNRTI LNRTI CFDQH/ HGADIG/ CFDKD/ VFDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDG/ FDG/ FDFG/ 	1 10 10 10 10 10 10 10 10 10 10 10 10 10	6 012345 15 hHHHHH VAVVQAJ VGVVQAJ GVVVQAJ VGVVQAJ YAVISG YAVISD YRVISD YRVISD YRVISD YRVISD YRVISD YRVISD YAEQIV WDEVIY YAVIAJ YAVIAJ YAVIAJ YAVIAJ YAVIAJ YDFIAS VDQIID INHID LGVITG INHID	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GITGNDMFN GISG GITGNDMFN GISG GIGGN GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKE GIKGHH
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima com. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSHLGDEGKVVAQMKA LDRVISVLDDQATINVSLA LDRVISVLDDQATINVSLA LDMCIALLDDEGVLNQLA LDMCIALLDDEFVLNQLA LDMCIALLDDEFVLNQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DTTINQLA LDSVIGVLD IFTLVQLA LDSVIGVLD IFTLVQLA LDTISMLDDPDMLAQLA LDMTISLLDDQAAFDAQLA FDMCISMLDDSDVLASQLA LDMGINLLMDAATITSLD	11 8901 5 FFAVN KKVPS UGVR LGVR LGVR LGVR LLGVR LLGVC LLSU LKSO LKSO LKSO LKSO LKSO LKSO LKSO LKSO	12 234567890 FF HITR1 HNR1 HKGYGNK-1 HKGYGNK-1 HKGYG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HIKL0 HIKL0 HDPR0 HDPR0 HTER0 HTER0 HTER0 HTER0 HTER0 HTER1 HVER1 HVERG1 HVERG1 HVAR1	2 13 01234567890 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV ITPKHFGQLL HIKAEYFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG DIKGEYFPALG SITGHMFNLMR SLDALNFAVFG SIDALNFAVFG SIDALNFAVFG STKPEYFDLFG STKPEYFD STFF STFF STFF STFF STFF STFF STFF ST	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETLK: KLVGGVFQU ASLLSAMEI ASLLSAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLPJ TGLAYVLPJ TGLAYVGVEI HAVMGVEI HAVMGVEI HAVMGVEI KHLADHIG KHLADHIG KHLADHIG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG	14. 901234 SG SKN ARMGN EEF-SA HRIGGK HRIGGK EKVPG EKVPG EKVPG EHID EHID EHID CHID NVIGSE SKFG SKFG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG CHIG SHG-	15 5678901234! 1 5 HHHHDHI FGDKYANJ YGDKYANJ PTVAJ MNAAAKDJ MNAAAKDJ MNAAAKDJ INRTI INRTI INRTI CFNQHJ GPDLPJ SFDNDJ FDKEJ FDKEJ FDKEJ FDKEJ FDKEJ FDKEJ FDKEJ FDKEJ 	1 10 10 11 10 10 11 11 10 10 10 10 10 10	6 012345 15 hHHHHI VAVVQA VGVVQA YADISG YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD YRVIAS YAUAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISGV GISGV AIHD AIAEGLP
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKNVAQMKA IGVAVSHLGDEGKNVAQMKA LDRVISQLDQQATINVSLA FDILISVLDDKPVLDQLA LDRVISMLDNAEALNADLE LDMCIALLDDEGVLNIQLA LDMCIALLDDEGVLNIQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDTISMLDDPDMLAAQLA FDMCISMLDDSDVLASQLA FDMCISMLDDSDVLASQLA LDMGINYLNDAATLTSITS	11 8901 5 FFAVN FFAVN VGVR VGVR LGVR LLGVR LLGVR LLGVR LLGVE LLAVO LKSO LKSO LKSO LKSO LKSO LKSO LKSO LKS	12 234567890 FF HITRI HDAMG HKGYGNK-I HKGFGI HKGFGI HKGFGI HKGFGI HKGFGI HKGFG HTER0 HDPR0 HDPR0 HTER0 H	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG SITGHMFNLMR SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG STKPEYFDLFG STKPEYFD STKP	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETIK: KLVGGVEQI ASLLSAMEI ASLLSAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI QAVMSVVP QTMFQTIAI TGLAYVLP KALFATVGG ASAISAVR HAVMGVEI HAVMGVEI HAVMGVEI HALABILG KHLADHLG KHLANIG KHLAVLS SLMLAVES TAILHVVA EILATGLP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK EVPG- EKVPG- EKVPG- THLD EHI SKVHG THLD EHIG SKG SKLG	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YSDDVAG/ DPTTVA/ NNAAKD/ NNAAKD/ NNAAKD/ INRT1 INRT1 CFNR/ CFNR/ CFDKD/ VFDKD/ VFDKD/ VFDKD/ HFDQA/ HFDQA/ HFDYAQ/ HFDYAQ/ HFDYAQ/ HFDYA/ HFDYA/ HFDYA/ FDKD/ CFDKD/ CFDKD/ CFDKD/ CFDKD/ C	1 10 10 10 11 10 10 10 10 10 10 10 10 10	6 012345 15 hHHHHI VAVVQAJ VGVVQAJ AGVLVAJ YADISG YADISG YREISD YREISD YREISD YREISD YREISD YRVISK YAUAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YDFISS UDVITG INHIID LGVITG TKVISS IDHIED LLPLKS LLPLKS	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISG GIGN GIGN GIGN GIK
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia A5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima com. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIB (B2)	9 10 89012345678901234567 15 1 eEee FFF LQNFIDQLDNPDDLVCVEK LNNFVDSLDPSRLKCVVEK LNNFVDSLDPSRLKCVVEK LOKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAEMKA IGVAVSHLGDEGKMVAEMKA IGVAVSHLGDEGKMVAEMKA IGVAVSHLGDEGKMVAEMKA IGKAVAEMDNAKQMAGTLHA IGAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKFVLDQALA LDRVISSLTDEFVLNAQLE LDRVISSLTDEFVLNAQLE LDRVISSLTDEFVLNALA LDRVISSLTDEFVLNALA LDRVISLDDEFVLNQLA LDRVIGVLD DTFTVQLA LDSVIGVLD DTFTVQLA LDISSLDDPDLAAQLA LDISSLDDPDLAAQLA LDISSLDDPDLAAQLA LDISSLDDPALAAQLA LDIAISTLDQPATLKEELD LDLAINLLDDPALDAAL	11 8901 5 FFAVN KVPS VGVR LGVR LGVR LGVR LLGVC LLGVC LLGVC LKS0 LKS0 LKS0 LKS0 LKS0 LKS0 LKS0 LKS0	12 234567890 FF HITR	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG SICALNFAVE GUEAAHYDTW GVEAAHYDTW GVEAAHYDTW GVEAAHYDTW GYEAAHYDTW GYEAAHYDTW GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG GVEAHYKFG GVQKAHFKKFG GVAVYFDEME	12345678 0 15 GgGGggGG GPIKKVLAS EPMKETIKK KLVGGVEQ ASLISAMEI MCLIDAMEI MCLIDAMEI MCLIDAMEI DALLEAMN QAVMSVVP QTMFQTIAI TGLAYVLP KALFATVGG KHLASHIG KHLASHIG KHLASHIG KHLASHIG KHLASHIG KHLAVLP SIMLAVES TAILHVVA KVLMTVLP KALFKVLP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK SKVHG- EKVPG- SKVHG- EHI SKIG KLG ELG CKLG	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YGDKYAN/ PTVA/ DPTTVA/ DPTTVA/ NRTI INRTI INRTI INRTI CFNRI/ HGADIG/ CFDKD/ VFDKD/ VFDKD/ VFDKD/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ 	1 56789 10 HhHH Amaki Am	6 012345 15 hHHHHH VAVVQAJ VGVVQAJ AGVLVAJ YADISG YRVISD Y	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISG GIGNDHFN GISG GIGNDHFN GISG GISG GISG GISG GISG GISG GISGN
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera mmc Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I(A1) Tylorrhynchus I(A2) Lamellibrachia AIII (A2) Lamellibrachia AIII (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus III (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIC (B1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LNKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSELDDGSKFEATLHK LNKLISQLDQQATINVSLA FDILISVLDDQATINVLA IDISSLTDEFVLNAQLE LDMCIALLDDEFVLNQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD IPTLTVQLA LDTISMLDD PDMLTAQLA FDMCISMLDDSDVLASQLA LDMISLLDDQAAFDAQLA FDMCISMLDDSDVLASQLA LDMGINYLNDAATLTSLTS LDLAINLLDDPFALDAALD LDTISLLDQPALDAALD LDAVNLEDFKALQEELK	11 88901 5 FFAVN KKVPS VGVR LGVR LGVR LGVR LLGVR LLRQ LKSO LKSO LKSO LKSO LKSO LKSO LKSO LKSO	12 334567890 FF HIRRI HNRI HKGYGNK-I HKGYGNK-I HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HVER0 HTER0 HTER0 HTER0 HTER0 HTER1 HVER1 HVERG1 HVERG1 HVERG1 HVERG1 HVERG1 HVERG1 HRER2 HRER2	2 13 01234567890 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV HIKAEYFEPLG HIKAEYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG SITGHMFNLMR GIDALNFAVFG SITABAGYPAFV GVEAAHYDTYN GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG GTKPEYFDLFG STKPEYFD STKPE	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLKX KLVGGVFQ ASLLSAMEI ASLLSAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFTIAL MCLLDAMEI DALLEAMN: QAVMSVVP QTMFTIA GLAYVLP TGLAYVLP TGLAYVLP TGLAYVLP KALFATAG KHLANLG KHLANLG KHLANLG KHLANLG KHLLAVLG NELLAVLP SIMLAVES TAILHVVA KVLMTVLP EILATGLP KALLKVLP KALLKVLP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK EKVPG- EKVPG- SKVHG- THL-D EHID EHID EHID EHID EHID EHID EHID EKUG SKUG SKL	15 5678901234! 1 5 HHHHbH FGDKYANJ YYSDDVAG DPTTVAJ MNAAAKDJ MNAAAKDJ INRTI INRTI INRTI CFNQH HGADIG SFNC SFNC SFNC SFNC SFNC SFNC SFNC HFDCG 	1 10 10 10 10 10 10 10 10 10 10 10 10 10	6 012345 15 NHHHHI VAVVQA VGVVQA YADISG YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD YREIST YAUIAA YAVIAA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GISG GISGN GISGN GISGN GISGN GIKH GIKH GIKH GIKH GIKH GIKH GIKH GIKH GIKH GIKH GIKH VIKAELP VIKAELP VISHRIDG-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima com. I (A1) Pheretima hig. I (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIE (B2) Tylorrhynchus IIC (B1)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSHLGDEGKVVAQMKA IGVAVSLDDQATINVSLA LDRVISNLDNAEALNADLE LDMCIALLDDEFVLNQLA LDMCIALLDDEFVLNQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDTISSLDDPDMLAQLA LDMTISLLDDQAAFDAQLA LDMTISLLDDQAAFDAQLA LDMISLDQAATLTSIS LDLAINLLDDPALDAALD LDLANLLEDFKALQEELK LDTIGVLGSDTINSLD	11 8901 5 FFAVN FRAVN KVPS UGVR LGVR LGVR LGVR LLGVR LLGVR LLGVC LKS0 LKS0 LKS0 LKS0 LKS0 LKS0 LKS0 LKS0	12 234567890 FF HITR1 HITR1 HAGYGNK-1 HKGYGNK-1 HKGYGS1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HIKL0 HDPR0 HTER0	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV ITPKHFGQLL HIKAEYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIKGEYFPALG SITGHMFNLMR SLDALNFAVFG SIDALNFAVFG SIDALNFAVFG STKPEYFDIFG STKPEYFDIFG STKPEYFDIFG STKPEYFDIFG STKPEYFDIFG STKPEYFDIFG SIKAUYGVFV ISAEYFDVFAE SLAVYFDAMG SVCAHFKKFG SVTKEYFRGIG SVTKEYFRGIG	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETLK: KLVGGVFQU ASLLSAMEI ASLLSAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLP TGLAYVLP TGLAYVLP TGLAYVLP KALFATVG KHLADHLG KHLAHLG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG ELLAYLP ELLATGLP KALKVLP KALNVLP EAFARVLP	14. 901234 SG SKN ARMGN EEF-SA HRIGGK EKVPG EKVPG EKVPG THLD EHI AQLG-R- SKFG OKLG	15 5678901234! 1 5 HHHHDHI FGDKYANJ YGDKYANJ YSDDVAGJ MNAAAKDJ MNAAAKDJ MNAAAKDJ MNAAKDJ INRTI INRTI INRTI CFNQHJ FNQHJ FNQHJ FNCH VFDKDJ VFDKDJ VFDKDJ FDKDJ FDKDJ FDKDJ 	1 10 10 11 11 11 11 12 12 12 12 12 12 12 12 12	6 012345 15 NHHHHI VAVVQA VGVVQA YADISG YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD YREISD YRVIAK MNQIVS LDVITG INVITM YAVIAA	17 678901234 20 25 HhHHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GIKG GIKG GIKGH GIKGH GIKGH KISSRLNA- VISHRIDG VISHRIDG VISHRIDG VISHRIDG
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIB (B2) Tylorrhynchus IV (B2) Oligobrachia c (B2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGVLVAQHA IDNCIALDDEGVLNAQLA IDMCIALLDDEGVLNAQLA IDMCIALLDDEGVLNAQLA IDSVIGVLD DDTFTVQLA IDSVIGVLD DDTFTVQLA IDSVIGVLD DDTFTVQLA IDSVIGVLD DDTFTVQLA IDTISMLDPPDLAAQLA FDMCISMLDDSDVLASQLA IDMINILDDPALDAQLA IDMGINYINDAATLTSIS IDLAINLLDDPALDAALD LDLAVNLEDFKALQEELK IDTLIGVLGSSDTINSLID IDLINLLDDTIVLQSHLG IDVAINMLDPAVLNECLA	11 8901 5 FFAVN FAVN VGVR UGVR LGVR LGQE LAGO LKSO LKSO LKSO LKSO LKSO LKSO LKSO LKS	12 234567890 FF HITR	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG DIRADFFPALG SITGMFNLMR SIDALNFAVFG SIDAGYPAFV GVSAAQYDVE SVEAAHYDTVN STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG SVCAVYFDAMG SVQKAHFKKFG GVXAVYFDEME SFXTYYFKEFG SVTKEYFRGG SVAAHFDVMA	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETIK: KLVGGVEQ ASLLSAME ASLLSAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLAME MCLLAME MCLLAME QAVMSVVP QTMFQTIA TGLAYVLP KALFATVG KHLADHLG KHLADHLG KHLAVLG KHLAVLG KHLAVLG KHLAVLG KHLAVLG KHLAVLC SLMLAVES TAILHVVA SLMLAVES TAILHVVA ELLAVLP KALLKVLP KALLKVLP KALLKVLP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK HRIGGK HRIGGK HRIGGK HIG DKUG DKLG DKLG DKLG DKLG DKLG DKLG DKLG DHLG DKLG DHLG QLG SKG QLG SC SC SC SC SC SC SC SC SC SC	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YGDKYAN/ YGDKYAN/ MNAAKD/ MNAAKD/ MNAAKD/ NRTI INRTI INRTI INRTI FNC/ CFNC/ FDKD/ FDKD/ FDKD/ FDKD/ FDKD/ 	1 56789 10 HhHH An AKI An AKI An AAA An AA An An Aa An Aa An Aa An An Aa An Aa An An A	6 012345 15 hHHHHI VAVVQAJ VGVVQAJ AGVLVAJ YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD YRVISG YAUAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA YAVIAA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISG GIGN GIGN GIGN GIGN GIGN GIGN GIGN GIK
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIC (B2) Tylorrhynchus IIC (B2) Lamellibrachia BIU (B2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSELDDGSKFEATLHK LNRLISQLDQQATINVSLA FDILISVLDDKVLVALA IDVISSLTDEFVLNALE LDRVISSLTDEFVLNALE LDRVISLDDEGVLNIQLA LDRVIGVLD DTFTVQLA LDSVIGVLD DTFTVQLA LDSVIGVLD DTFTVQLA LDSVIGVLD DTFTVQLA LDSVIGVLD DTFTVQLA LDTISMLDTPDLTAQLA LDTISMLDTPDLAQLA LDMISLDDQAFDAQLA LDMISLDDQAFTAQLA LDMISLDDQAFTAQLA LDMISLDDQAFTAQLA LDMISLDDQAFTAQLA LDIXISLDDQAFTAQLA LDMISLDDQAFTAQLA LDIXISLDQAFTAQLA LDIXISLDA LDI	11 88901 5 FFAVN KKVPS VGVR LGVR LGVR LGVR LLGVR LLGVR LLRQ LLRQ LLRQ LLRQ LLRQ LLRQ LLRQ LL	12 234567890 FF HITR	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRGEYFPALG MILDASNFAAMG FIPFKAFG SITAGDFFPALG SITAGTFP	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETIK: KLVGGVEQI ASLLSAMEI MCLLDAMEI MCLLDAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLP: KALFATVG KALFATVG KHLASHLG KHLASHLG KHLASHLG KHLASHLG KHLAVLP SLMLAVES TAILHVVA: KVLMTVLP: EAFARVLPP EAFARVLPP EAFARVLPP EAFARVLPP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK HRIGGK SKVHG- SKVHG- EHI SKIG SCIG SSKIG SSSKIG SS	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YGDKYAN/ DPTTVA/ DPTTVA/ NRT1 INRT1 INRT1 INRT1 CFNR1/ CFNR2/ CFDKD/ VFDKD/ VFDKD/ VFDKD/ VFDKD/ HFDGG/ HFDFGG/ HFDFGG/ HFDFGG/ FKD/ CFNKD/ CFNKD/ CFNPD/ CFNPD/ CFNPD/	1 56789 10 40 40 40 40 40 40 40 40 40 40 40 40 40	6 012345 15 hHHHHH VAVVQAJ VGVVQAA AGVLVA YADISG YREISD YREISD YREISD YREISD YREISD YREISD YRVISD YERIAS YAEQIV WDEVIY YAVIAA YD	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GISGSQV GIK GIK GIK GIK GIK GIK GIK GIK GIK GISAGL VISHRIDG- RIAKDLP GISAGL
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus I (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Lamellibrachia AIII (A2) Niftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima ig. I (A1) Pheretima hig. I (A1) Lumbricus d1 (A1) Lumbricus d2 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIC (B1) Lumbricus IV (B2) Oligobrachia c (B2) Lamellibrachia BII (B2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LNKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGKTVAQMKA IGVAVSHLGDEGVLNQLA LDRVISSLTDEFVLNQLEH LDMCIALLDDEFVLNQLA LDSVIGVLD DDTFTVQLA IDSVIGVLD DDTFTVQLA IDSVIGVLD DDTFTVQLA IDSVIGVLD IFTLVQLA LDSVIGVLD IFTLVQLA LDTISMLD PDMLTAQLA LDTISMLDD PDMLTAQLA LDMTISLLDDQAAFDAQLA FDMCISMLDDSDVLASQLA LDMGINYLNDAATLTSLSS LDLAINLLDDPFALDAALD LDLAINLLDDFALDAALD LDLAINLLDDFALDALD LDLAINLLDDFALDALD LDLAINLLDDFALDALD LDLAINLLDDFALDALD LDLAINLLDDFALDALD LDLAINLLDDFALDALD LDLAINLLDDFALDALD LDLINLLDTIVLQSHLG LDVAINMINDPAVLNEQLA	11 8901 5 FFAVN KKVPS UGVR LGVR LGVR LGVR LGVR LLGVC LNS0 LNS0 LNS0 LNS0 LNS0 LNS0 LNS0 LNS0	12 234567890 FF HIRRI HKGYGNK-I HKGYGNK-I HKGYGS1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HKGFG1 HVER0 HETR0 HTER0	2 13 01234567890 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV ITPKHFGQLL HIKACYFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG DIRASNFAAMG FIPFKAFG SITGHMFNLMR SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG SVKAVYFDAFK SLAVYFDAFK SVKAVYFDAFK SVKAVYFDEME SVKAVYFDEME SVKAVYFDEME SVKAVYFDEME SVKAVYFREGG SVXANFDVMA SVTKGGFSAIA	12345678 0 15 GgGGggGGG GPIKKVLA: EPMKETLK', KLVGGVFQI ASLLSAMEI MCLLDAMEI MCLLDAMEI DALLEAMN: QAVMSVVP QTMFQTIAI TGLAYVLPJ TGLAYVLPJ TGLAYVLPJ TGLAYVGVEI VHLLRILD TQLFDILG KHLADIGG KHLADIGG KHLADIGG KHLAVLS KHLHVVG NHLLEVLG NHLLEVLG NHLLEVLG NHLLEVLG NHLLAYLPG EILAYLPG EILAYLPG EAFARVLPG EAFARVLPG EAFARVLPG ESFAEVLPG	14. 901234 SG SKN ARMGN EEF-SA HRIGGK HRIGGK EKVPG SKVHG PHLD PHLD EHI AQLG-R SKFG KLG SHLG SHLG HLG SHLG HLG SHLG SHLG SHLG PHLG SHLG-	15 5678901234! 1 5 HHHHbHH FGDKYANJ YYSDDVAGJ DPTVAJ NNAAKDJ NNAAKDJ NNAAKDJ INRTI INRTI INRTI FNQHJ FNQHJ FNQHJ FNQHJ FNQHJ 	1 567389 10 10 10 10 10 10 10 10 10 10 10 10 10	6 012345 15 NHHHHI VAVVQA VGVVQA YADISG YADISG YADISG YALSD YREISD YREISD YREISD YRVISD YERIAS YAUAA YAVIAA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GISG GITGNDMFN GISG GITGNDMFN GISG GITGNDMFN GISG GISGN GISGN GISGN GISGN GISGN GISGN GISGN GITGGLA GMTEGLA EIGSOLAD-
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Riftia b (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus III (B1) Tylorrhynchus IIB (B2) Tylorrhynchus IIC (B1) Lumbricus IV (B2) Oligobrachia c (B2) Lamellibrachia BII (B2) Sabellag Glb1 (P2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDKVVDALG-GNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IGVAVSHLGDEGKVAQMKA IDVISULDQQATINVSLA IDVISSLTDEPVLNAQLE LDMCIALLDDEGVINQLA LDMCIALLDDEFVLNQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DTTINVLA LDTISSLDDQAAFDAQLA IDMTISLLDDQAAFDAQLA IDMISLDDQAAFDAQLA IDMISLDDPALAAAL IDMISLDDPALAAAL IDMISLDDPALAAAL IDMISLDDPALAAAL IDIANNLDDPALDAALD LDLANLLDDFKALQEELK LDVAINMLNDPAVLNEQLA LDSAIGLLSPSTLNDQLI HDTIINMA DDTDVLFLIT	11 8901 5 FFAVN FFAVN KVPS UGVR LGVR LGVR LGVR LLGVR LLGVR LLGVC LKS0 LKS0 LKS0 LKS0 LKS0 LKS0 LKS0 LKS0	12 234567890 FF HITR1 HITR1 HITR1 HAGYGNK-1 HKGYGNK-1 HKGYGG1 HKGYGG1 HKGFG1 HKGFG1 HKGFG1 HIKL0 HETR0 HETR0 HTER0 HTER0 HTER1 HVER	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKAEYFEPLG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG DIRAEFFPALG SITGHMFNLMR SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG SIDALNFAVFG STKPEYFDIFG STKPEYFDIFG STKPEYFDIFG STKPEYFDIFG SIKAUYFDAMG SURAVYFDAMG SURAVYFDAMG SVKAEYFALFR SVKAEYFALFR SVKAEYFALFR SVKAEYFALFR	12345678 3 3 3 3 3 3 3 3 3 3 3 3 3	14. 901234 SG SKN ARMGN EEF-SA HRIGGK EKVPG EKVPG EKVPG THLD EHI AQLG-R- SKVHG AQLG-R- SKFG CKLG	15 5678901234! 1 5 HHHHDHI FGDKYANJ YGDKYANJ YSDDVAGJ MNAAAKDJ MNAAAKDJ MNAAAKDJ MNAAAKDJ INRTI INRTI INRTI FNQHJ FNQHJ FNQHJ FNQHJ 	1 567389 10 HhHH: A AKI A GDA A GDA	6 012345 15 hHHHHI VAVVQA VGVVQA YADISG YADISG YADISG YREISD YREISD YREISD YREISD YREISD YREISD YREISD YREISD YREISD YAUXA YAVIAA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALIAGLES- GISG GISG GISG GISG GISG GIGGN GISGN GISG GISG GISG GISG GISG GIKG GIKG GIKG GIKG GIKG GIKGH GIKGH GIKGH GIKGH GIKGH GIKGH GIKGH GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL GISAGL
Globin fold Scapharca I Scapharca IIA Urechis FI Glycera MH4 Glycera P1 Glycera P2 Glycera P3 Oligobrachia b (A1) Tylorrhynchus I (A1) Tylorrhynchus IIA (A2) Lamellibrachia BI (A1) Oligobrachia a5 (A2) Lamellibrachia AIII (A2) Pontodrilus I (A1) Pheretima sie. I (A1) Pheretima hig. I (A1) Pheretima hig. I (A1) Uumbricus d1 (A1) Lumbricus d2 (A1) Tubifex I (A1) Sabella Glb3 (A1) Lumbricus II (A2) Lamellibrachia BIV (B1) Lumbricus III (B1) Tylorrhynchus IIB (B2) Tylorrhynchus IIC (B1) Lumbricus IV (B2) Oligobrachia c (B2) Lamellibrachia BII (B2) Sabella Glb2 (B2)	9 10 89012345678901234567 15 1 eEEee FFF LQNFIDQLDNPDDLVCVVEK LNNFVDSLDDPSRLKCVVEK LNNFVDSLDDPSRLKCVVEK LDNVVDALGGNAGALMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKMVAQMKA IGVAVSHLGDEGKNVAQMKA IGVAVSHLGDEGKNVAQMKA IGVAVSLDDQATINVSLA FDILISVLDDKPVLDQLA LDRVISMLDNAEALNADLE LDMCIALLDDEGVLNIQLA LDMCIALLDDEGVLNIQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDSVIGVLD DDTFTVQLA LDTISMLDDPDMLAQLA FDMCISMLDDSDVLASQLA FDMCISMLDDSDVLASQLA LDIAISTLQPATLKEELD LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDLAINLLDDPALDAQLA LDSAIGLLSDPSTLNQLI HDTIINMA DSDALAQQLI FDTIINMA DSDALAQQLI FDTIINMA DSDALAQQLI	11 8901 5 FFAVN FAVN VGVR UGVR LGQE LAGO LKSO LKSO LKSO LKSO LKSO LKSO LKSO LKS	12 234567890 FF HITR	2 13 0123456780 1 5 1 GGGGGGGGG KISAAEFGKIN KISGDAFGAIV -ITPKHFGQLL HIKA2YFEPLG DIRADFFPALG DIRADFFPALG DIRADFFPALG DIRADFFPALG SITGMFNLMR SIDALNFAVFG SIDAGYPAV GVSAAQYDVE SVEAAHYDTVN STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG STKPEYFDLFG SVKAVYFDAFK SLKAVYFDAFK SVKAVYFDAFK SVKAVYFDAFK SVXAVYFDEME SFXTVYFKEFG SVXAAHFDVMA SVTKGGFSAIA SVKASYIALFR	12345678 0 15 GgGGggGG GPIKKVLA: EPMKETIK: KLVGGVEQ ASLLSAME ASLLSAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLDAME MCLLAME MCLLAME QAVMSVVP QTMFQTIA TGLAYVLP KALFATVG KHLADHLG KHLADHLG KHLADHLG KHLADHLG KHLAVLP SLMLAVES TAILHVVA KVLMTVLP ELFATGLP KALKVLP KALKVLP KALKVLP KALKVLP KALKVLP CSFAEVLP QSFLRVMP QSFLRVMP	14. 901234 SG SKN- ARMGN- EEF-SA HRIGGK HRIGGK HRIGGK SKVPG- EKVPG- SKVPG- THLD EHI SKUPG- THLD EHI SKG OKLG OK-	15 5678901234! 1 5 HHHHHH FGDKYAN/ YGDKYAN/ YGDKYAN/ WNAAAKD/ MNAAAKD/ MNAAKD/ MNAAKD/ NRT1 INRT1 CFNCH/ CFNCH/ FOKD/ FOKD/ FDKD/ FDKD/ 	1 567389 10 HhHH Ar AKI Ar GAA Ar GAA Ar GAA Ar GAA Ar GAA Ar GAA Ar GAA Ar GAA Ar CAA Ar CAAA Ar CAA Ar CAAA Ar CAA Ar CAAA Ar CAAA Ar CAAA Ar CAAA Ar CAAA Ar CAAAA Ar CAAAAA Ar CAAAAAA Ar CAAAAA Ar CAAAAAA Ar CAAAAAA Ar CAAAAAAAA Ar CAAAAAAAAAA	6 012345 15 hHHHHI VAVVQAJ AGVLVAJ YADISG YADISG YREISD YERISD YERISD YERISD YERISD YERISD YERISD YERISD YERISD YERISD YAUIAS YAUIAS YAUIAS YAVIAS YA	17 678901234 20 25 HhHHHHH AL ALISGLQS- ALISGLQS- ALISGLQS- ALISGLQS- GISG GISG GISG GIGN GIGN GIGN GIGN GIGN GIGN GIGN GIGN GIGN GIK

FIG. 1. Alignment of globin chains. The sequences have been aligned in the coding portions using as reference the tertiary structure template of invertebrate globins (7). They are shown with the accession numbers for the Swiss-Prot, NCBI, and EMBL data banks in parentheses: *Scapharca* I (P02213), *S. inaequivalvis*; *Scapharca* IIA (P14821), *S. inaequivalvis*; *Urechis* F1 (P06148), *U. caupo*; *Glycera* MH4 (P15447), *G. dibranchiata*; *Glycera* mmc (P02216), *G. dibranchiata*; *Glycera* I (P23216), *G. dibranchiata*; *Glycera* III (P21659), *G. dibranchiata*; *Glycera* III (P21660), *G. dibranchiata*; *Oligobrachia* b (5), *Oligobrachia* mashikoi; *Tylorrhynchus* I (P02219), *Tylorrhynchus* heterochaetus; *Tylorrhynchus* IIA (P09966), *Tylorrhynchus* heterochaetus; *Lamellibrachia* BI (23), *Lamellibrachia* sp.; *Oligobrachia* a5 (5), *Oligobrachia* mashikoi; *Lamellibrachia* AIII (P15469), *Lamellibrachia* sp.; *Riftia* b (P80592), *Riftia* pachyptila; *Pontodrilus* I (24), *Pontodrilus* matsushimensis; *Pheretima* sie. I (P11740), *Pheretima* sieboldi; *Pheretima* com. I (24), *Pheretima* comunissima; *Pheretima* hil. I (25), *Pheretima* hilgendorfi; *Lumbricus* d1 (U55073),



Phylogenetic Analyses

Sabella

diversicolor;

common to the majority of sequences.

Lamellibrachia

Neanthes

heterochaetus;

Sequence alignments were used to build phylogenetic trees using the maximum parsimony² (9) and neighbor-joining (NJ) methods (10).

Outgroup Choice-Globin sequences with known x-ray structures were chosen as outgroups. Based on the phylogeny of Metazoa (11), the sequences of the closest relatives of Annelida, Vestimentifera, and Pogonophora were selected as outgroups. These groups are (i) the globins of the mollusk Scapharca inaequivalvis, (ii) the globin of the echiuran Urechis caupo, and (iii) the intracellular globins of the annelid Glycera dibranchiata. No outgroup for the linker sequences could be selected because their evolutionary position is unknown (4).

Cladistic Analysis-The cladistic analyses were done using the program PAUP, version 3.1.1. (12). Globin evolution was studied using a heuristic approach because the number of sequences involved excludes the exhaustive approach (12). All characters were weighted equally, and those uninformative were excluded from the analysis. After a series of trials, we set the PAUP options to perform more efficiently in finding the shortest trees. For the optimization of characters, the ACCTRAN option was set. For the heuristic searches, the options set were: (i) keep minimal trees only, (ii) collapse zero-length branches, (iii) starting tree sources, get by stepwise addition, (iv) swap on, minimal trees only, (v) addition sequence, random, (vi) replications, 50, (vii) swapping algorithm, TBR tree bisection-reconnection, (viii) MULPARS on, and (ix) steepest descent on.

For the phylogenetic analysis of linker chains, we applied an exhaustive approach because of the smaller size of the data set (12). Uninformative characters were also excluded, and the remaining characters were weighted equally. The PAUP optimization of characters (ACCT-RAN on) and exhaustive search (keep minimal trees only and collapse zero-length branches on) options were set.

Tree Indexes in the Cladistic Analyses-Two indexes are traditionally used to test the robustness of the most parsimonious tree obtained by cladistic analyses, the consistency and retention indexes (13, 14). These indexes were calculated as implemented in PAUP (12). Values of the consistency index and retention index > 0.5 indicate that convergent/ parallel evolution does not affect strongly the phylogenetic reconstruction and that the obtained topologies of trees are reliable. Molecular synapomorphies in the cladistic analysis were detected using the program MacClade (Version 3), which allows the tracking of the evolution of each character (15).

Neighbor-joining Analysis-The distance-based phylogenetic reconstructions were performed on globin and linker alignments according to the NJ method (10) as implemented in the TREECON program (Version

Lumbricus terrestris; Lumbricus d2 (U55074), Lumbricus terrestris; Tubifex 1 (P18202), Tubifex tubifex; Sabella Glb3 (AJ131285), S. spallanzanii; Lumbricus II (P02218), Lumbricus terrestris; Lamellibrachia BIV (23), Lamellibrachia sp.; Lumbricus III (P11069), Lumbricus terrestris; Tylorrhynchus IIB (P13578), Tylorrhynchus heterochaetus; Tylorrhynchus IIC (P02220), Tylorrhynchus heterochaetus; Lumbricus IV (P13579), Lumbricus terrestris; Oligobrachia c (5), Oligobrachia mashikoi; Lamellibrachia BII (23), Lamellibrachia sp.; Sabellastarte E (D58418), Sabellastarte indica; Sabella Glb1 (AJ131283), S. spallanzanii; and Sabella Glb2 (AJ131284), S. spallanzanii. The amino acids common among all the globins of Annelida, Pogonophora, and Vestimentifera are indicated by asterisks. Red, invariant amino acids; blue, amino acids common to the majority of sequences; green, synapomorphic amino acids; black, autoapomorphic amino acids.

² Glossary of phylogenetic terms: apomorphy, a derived state of a character that represents an evolutionary novelty with respect to the ancestral state; autoapomorphy, a derived character state that is unique for a particular sequence; bootstrap test, a statistical test used to verify the robustness of the topology of an evolutionary tree; bootstrap value, the result of the bootstrap test (in this paper it is expressed as a percentage); cladistic analysis, a phylogenetic reconstruction based on the principle of maximum parsimony; cladogram, a tree depicting the phylogenetic relationships that results from a cladistic approach; clade, a monophyletic group in a cladistic context; ingroup, a set of sequences that are considered the focus of interest: monophyletic group. a group that includes all the sequences that originate from a common ancestor; most parsimonious tree, a tree produced in a cladistic analysis that has the shortest length; orthologous sequences, two sequences derived from a speciation event (i.e. the same sequence in different species); outgroups, a set of sequences that are brought into the analysis to determine the root of the ingroup and ancestral states; strict consensus tree, a tree derived from a set of trees in which all conflicting branching patterns are collapsed into multifurcations; synapomorphy, a derived state of a character that is shared by all the taxa belonging to a clade; tree length in the cladistic analysis, the value obtained by computing the sum of the minimum numbers of substitutions in all the positions of the alignment.

Percentage of amino acids

1.3b) (16). The NJ trees were created by applying the following settings: (i) distance calculation, *Kimura 83*; (ii) alignment positions, *all*; and (iii) insertions and deletions, *not taken into account*.

Bootstrap Test—The bootstrap resampling (17) was performed to test



FIG. 3. Pairwise comparison of open reading frames and 3' UTRs of globin cDNAs. Black bars, globin open reading frame; white bars, globin 3' UTRs; SAS.g1, Sabella Glb1 globin; SAS.g2, Sabella Glb2 globin; SAS.g3, Sabella Glb3 globin; SAI.gl, Sabellastarte E globin; LUT.d1, Lumbricus d1 globin; LUT.d2, Lumbricus d2 globin; LUT.III, Lumbricus III globin.

the robustness of the trees obtained by cladistic and neighbor-joining phylogenetic reconstructions. In both cases, 500 replicates were run.

RESULTS

Sequence Alignments

Globin Alignment-In Fig. 1 we present the results of a global alignment of the Sabella globin sequences with the globins of other annelids, vestimentiferans, pogonophorans, and the canonical globin fold (7). Eight sequences chosen according to the criteria described under "Materials and Methods" were added as outgroups. An inspection of the alignment clearly indicates that the key residues Trp (A12), Phe (CD1), His (E7), and His (F8), consistent with the canonical globin fold, are absolutely conserved. Conversely, the presence of Pro (C2), which usually determines the bend of the corner between helices B and C, is not universal to all the globins. In two Sabella globins (Glb1 and Glb2), the small residues Ala and Cys are found at positions B6 and E8. In these sites the majority of annelid globins have a Gly residue, with the exception of globin I of *Tylorrhynchus*, in which a Phe is found at position B6. These substitutions in the Sabella globins would result in a closer crossing of the B and E helices. All the extracellular annelid globins shown in the alignment have in common the Cys residues at positions NA2 and H11. These amino acids are important for the formation of the first supramolecular aggregate in the assembly of the whole Hb (18), which is confirmed by the fact that they are replaced in the intracellular globins of Glycera, Urechis, and Scapharca by other residues. Four other residues, namely Arg (E10), Phe (E4), His (F3), and Gln (F7), are highly conserved in the extracellular globins, but the structural and functional significance of this conservation remains to be clarified.

Linker Alignment—A similar alignment was carried out for the linker sequences of Sabella with the five other linkers available in the data bases (Fig. 2). The main feature that seems to be conserved is a cysteine-rich segment ((Cys- X_6)₃-Cys- X_5 -Cys- X_{10} -Cys) that is typical for all the linker chains sequenced thus far and has been related to the low density lipoprotein-receptor motif (4). Considering the current global



Pairwise alignments

FIG. 4. Pairwise comparison of amino acid sequences of linker chains. Black bars, identical amino acids; striped bars, strongly similar amino acids; dotted bars, weakly similar amino acids; white bars, different amino acids; TY2, Tylorrhynchus L2 linker; NEA, Neanthes L2 linker; SA1, Sabella L1 linker; LUM, Lumbricus L1 linker; TY1, Tylorrhynchus L1 linker; LAM, Lamellibrachia LAV1 linker; SA3, Sabella L3 linker.



FIG. 5. Phylogenetic analysis of globin polypeptides. A, the SC tree derived from merging 18 most parsimonious cladograms. The statistic analyses of the most parsimonious cladograms gave the following values: length, 1,689 steps; consistency index, 0.601; retention index, 0.587. B, NJ tree. Both trees are based on the alignment shown in Fig. 1. The *numbers* on the branches refer to the bootstrap values expressed as a percentage after 500 replicates; only values \geq 50% are reported. *Lowercase letters* located close to the nodes refer to monophyletic clades that are supported by the molecular synapomorphies reported in Table I. The standard abbreviations of globins are shown in *parentheses* according to the nomenclature previously suggested by Gotoh *et al.* (22).

alignment (Fig. 2), the cysteine-rich segment can be written more precisely as $((Cys-X_{5}-_{6})_{2}-Cys-X_{6}-Cys-Asp-X_{3}-Asp-Cys-X_{4}-Asp-Glu-X_{4}-Cys).$

Pairwise Comparison—The analysis of the pairwise comparison of the globin cDNAs (Fig. 3) shows a higher percentage of identity in the open reading frame portion (57.59% mean value) than in the 3'-UTR portion (47.82% mean value). This is consistent with a higher degree of variability of the globin 3' UTRs that are less subject to structural constrains. However, it should be noted that *Lumbricus* d1 globin and d2 are more similar in 3' regions than in the coding portions.

The pairwise comparisons performed on the linker chains (Fig. 4) give the following main results. (i) *Tylorrhynchus* L2 and *Neanthes* L2 markedly differ from the general structure of other linkers. In fact they share 74.58% of identical amino acids, whereas the mean for the whole data set is 25.79%. The latter value decreases to 23.35%, if the comparison of *Tylorrhynchus* L2 versus Neanthes L2 is excluded from the computation. (ii) Roughly 40% of the amino acids are specific for each chain (mean of dissimilar amino acids = 39.45%). (iii) Identical amino acids represent one fourth of the whole data when sequences are pairwise-compared. However, the residues common to all the linker chains in the global alignment are only 15 (5.7%) (Fig. 2), and they are mainly restricted to the cysteinerich segment.

Phylogenetic Analyses

Globin Phylogenetic Analysis-We have applied both cladistic and distance-based methods to study the molecular evolution of annelid globins. The cladistic analysis results in 18 equally parsimonious cladograms. The resulting strict consensus (SC) tree is presented in Fig. 5A together with the distancebased tree (Fig. 5B). Both phylogenetic reconstructions recognize a monophyletic origin of Hbs, Chls, Hb_Hs, and Hb_Ls. A more detailed analysis of the ingroup shows that the extracellular Hbs, Chls, Hb_Hs, and Hb_Ls can be divided into two distinct groups. However, the actual position of the Lumbricus III globin cannot be resolved in the SC tree. The topologies of the two trees show some discrepancies. In particular the placement of the Lamellibrachia BIV globin seems controversial. In fact, in the cladogram it is placed into the A group according to the classification of Gotoh et al. (3), whereas in the NJ tree it is included in the B group. Both hypotheses are poorly supported by bootstrap values, revealing the weakness of the more basal nodes. In Table I we list the clades that belong to the ingroup and are supported by one or more molecular synapomorphies. Several of them are also sustained by high bootstrap values.

Linker Phylogenetic Analysis—The cladistic analysis of linker sequences produced two equally parsimonious cladograms. The SC tree is showed in Fig. 6, where the NJ tree is

B

Evolution of Annelid Hemoglobins

TABLE I

Globin clades

The globin sequences aligned in Fig. 1 have been grouped according to the maximum parsimony criterion, and the common amino acids that represent the molecular synapomorphies for the different clades are listed together with their relative positions in the alignment. i, clade positions on the trees reported in Fig. 5. BT, bootstrap values supporting the different clades. Only values $\geq 50\%$ are indicated.

i	Clade	BT	Molecular synapomorphies
		%	
a	Tylorrhynchus IIA + Tylorrhynchus I		Gln (20), Val (29), Asn (52)
b	Lamellibrachia AIII + Riftia b	100	Tyr (8), Phe (42), His (46), Val (47), Thr (74), Thr (82), Thr (103), Val (129), His (131), Met (135), Gln (168)
с	Pontodrilus I + (Pheretima sie. I + (Pheretima com. I + Pheretima hig. I))	95	Leu (54), Glu (97), Pro (169)
d	Pheretima sie. $I + (Pheretima com. I + Pheretima hig. I)$	96	His (42), Phe (43)
е	$Hbs + Chls + Hb_{H}s + Hb_{L}s$	98	Cys (10), Arg (83), Cys (159)
f	Pheretima com. I + Pheretima hig. I	68	Glu (29), His (136), Phe (156)
g	Pontodrilus I to Lumbricus d2;		Asp (139), Thr (149)
ĥ	Lumbricus $d1 + Lumbricus d2$	98	Pro (57), Thr (97), Met (100)
i	Tylorrhynchus IIB to Sabella Glb1		Ser (35)
j	Sabellastarte E + (Sabella Glb2 + Sabella Glb1)	100	Met (12), Ala (21), Glu (23), Lys (50), Ala (95), Phe (96), Met (121), Leu (126), Ala (140), Pro (144), Ala (162), Met (163)
k	$Sabella \ Glb2 + Sabella \ Glb1$	98	Asn (110), Ser (170)



FIG. 6. Phylogenetic analyses of linker polypeptides. A, the SC tree derived from merging two most parsimonious cladograms The statistic analyses of the most parsimonious cladograms gave the following values: length, 326 steps; consistency index, 0.862; retention index, 0.612. B, NJ tree. Both trees are based on the alignment reported in Fig. 2. Numbers on the branches refer to the bootstrap values expressed as a percentage after 500 replicates; only values \geq 50% are reported.

Sabella L1

Sabella L3

also presented. Cladistic and distance-based analyses were performed only on the available sequences of linker chains because no convincing putative outgroups are known. In fact, with the exception of the cysteine-rich segment that relates the linkers with the low density lipoprotein receptor (4), no other significant similarity is known between linker chains and other proteins. Nevertheless, *Tylorrhynchus* L2 and *Neanthes* L1 were used to root the trees, because they are much more similar to each other than to other linkers.

The SC and NJ trees show some discrepancies in their topology. The cladogram structure favors a strict relationship between *Tylorrhynchus* L1, *Lamellibrachia* LAV1, and *Sabella* L3. This clade is also supported by a very high bootstrap value. Conversely, the NJ tree supports the two groups *Tylorrhynchus* L1 + *Lumbricus* L1 and *Sabella* L1 + *Sabella* L3. The first is also corroborated by the bootstrap value (72%), whereas the second does not receive strong support by the bootstrap test (48%). Both cladistic and distance-based analyses strongly favor the grouping of *Tylorrhynchus* L2 and *Neanthes* L2 with respect to other linker chains.

DISCUSSION

The cladistic and distance-based analyses that we performed on the extracellular Hbs of annelids, pogonophorans, and vestimentiferans confirm the previous results of Gotoh *et al.* (3). These authors proposed the division of the globin chains into two main groups, A and B, each divided further into two subgroups, A1/A2 and B1/B2. This classification was successively applied to Hb_Hs and Hb_Ls (19, 20). Recently, a correction of this nomenclature has been proposed that recommends an inversion of the names of the two main strains (21).

However the identification of homologous chains, *i.e.* globins that are the products of orthologous genes, does not seem to be a trivial task. In both our analyses *Tylorrhynchus* IIA and *Tylorrhynchus* I are grouped together more closely than the respective "homologous" sequences. Therefore it seems problematic to name them *Tylorrhynchus* A2 and *Tylorrhynchus* A1 as previously suggested (22). In this light *Lumbricus* d1 and *Lumbricus* d2 can be defined either as two allelic forms of the same gene or as the products of a very recent gene duplication. The same reasoning can be applied to *Sabella* Glb1 and *Sabella* Glb2.

The subdivision into four homologous groups of globins could be an oversimplification of the real situation. In fact we have found a higher number of globin chains in the Chl, purified from a single specimen of *Sabella* (6). A study of the entire portion of the genome coding for these proteins should be the best way to understand these discrepancies.

The globins and linkers, forming the Chls, are tightly associated in the phylogenetic reconstructions with those included in the Hbs. This clearly identifies the Chl as a variant of Hb. On this assumption the name chlorocruorin must be considered only as a descriptive term.

The phylogenetic reconstruction presented in this paper reveals that the genes coding for the polypeptides that form Chls, Hbs, Hb_Hs, and Hb_Ls appeared before the separation of Vestimentifera and Pogonophora phyla from Annelida (*Sensu stricto*) (11). The evolutionary pathway does not change even if we consider pogonophorans and vestimentiferans as members of the class Opisthochaeta within the phylum Annelida (20). The critical question is: what was the scenario in which Chls, Hbs, Hb_Hs, and Hb_Ls evolved?

Yuasa *et al.* (5) previously suggested that Hb, Chl, and Hb_H evolved from Hb_L, which is made only of globin chains, adding linker chains to form the final hexagonal bilayered structure. If



FIG. 7. Evolutionary models of extracellular hemoglobins. The two alternative hypotheses for the evolution of extracellular hemoglobins are shown. Letters in parentheses correspond to the different assumptions that support the two different models. See the text for discussion.

we accept this hypothesis we must make four assumptions (Fig. 7A): (i) Hb_L was present in the common ancestor of Annelida, Vestimentifera, and Pogonophora, (ii) in the Annelida phylum Hb₁ disappeared, originating Hb and Chl, (iii) in Vestimentifera Hb_H evolved from Hb_L but the latter did not disappear, and (iv) Hb_H evolved separately from Hb and Chl. The last assumption is supported by the fact that vestimentiferans and pogonophorans are sister groups, independent from their systematic position. As a consequence, Hb_H could not be present in their common ancestor (an annelid) because otherwise we should admit its subsequent loss in Pogonophora, which contrasts with the starting hypothesis.

Our analyses performed on linkers of Hb, Chl, and Hb_H show that the genes coding for these polypeptides were already present in the common ancestor of Annelida, Vestimentifera, and Pogonophora. We therefore suggest an alternative scenario for the evolution of this group of respiratory pigments. Our most parsimonious hypothesis is based only on the following three assumptions (Fig. 7B). (i) The common ancestor of Annelida, Vestimentifera, and Pogonophora had in its genome the whole set of genes coding for globin and linker chains. As a consequence Hb, Chl, and Hb_H did not evolve independently. (ii) The common ancestor of Vestimentifera and Pogonophora evolved the Hb_L from Hb_H, starting from the set of genes coding for the latter. (iii) Pogonophora lost Hb_H during evolution.

The data deduced from our phylogenetic analyses fit nicely in this new hypothesis, which requires fewer ad hoc assumptions to explain the origin of this group of proteins. The new evolutionary model that we propose can be applied also to the Hbs of leeches, which were not considered in our study because the available sequences are not sufficiently complete. However, the data based on the amino termini of some leech globins clearly show that they are homologous to other extracellular Hbs (21).

Acknowledgments-We thank Dr. Lucio Cariello, Director of the Stazione Zoologica "A. Dhorn" (Napoli, Italy), for providing the specimens of S. spallanzanii.

REFERENCES

- 1. Vinogradov, S. N. (1985) Comp. Biochem. Physiol. B 82, 1–15
- Vinogradov, S. N. (1985) Comp. Buchem. Infisiol. B 22, 1–15
 Lamy, J. N., Green, B. N., Toulmond, A., Wall, J. S., Weber, R. E., and Vinogradov, S. N. (1996) Chem. Rev. 96, 3113–3124
 Gotoh, T., Shishikura, F., Snow, J. W., Ereifeji, J. W., Vinogradov, S., and Walz, D. A. (1987) Biochem. J. 241, 441–445
- Suzuki, T., and Riggs, A. F. (1993) *J. Biol. Chem.* **268**, 13548–13555 Yuasa, H. J., Green, B. N., Takagi, T., Suzuki, N., Vinogradov, S., and Suzuki,
- T. (1996) Biochim. Biophys. Acta 1296, 235–244 6. Pallavicini, A., Negrisolo, E., Barbato, R., Dewilde, S., Ghiretti-Magaldi, A.,
- Moens, L., and Lanfranchi, G. (2001) J. Biol. Chem. 276, 26384–26390
 Kapp, O. H., Moens, L., Vanfleteren, J., Trotman, C., Suzuki, T., and Vinogradov, S. (1995) Protein Sci. 4, 2179–2190
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, $46\overline{7}3 - 4680$
- 9. Li, W. H. (1997) Molecular Evolution, Sinauer Associates, Inc., Sunderland, MA

- Saitou, N., and Nei, M. (1987) Mol. Biol. Evol. 4, 406-425
 Conway Morris, S. (1993) Nature 361, 219-225
 Swofford, D. L. (1993) Phylogenetic Analysis Using Parsimony, Center for Biodiversity, Illinois Natural History Survey, Champaign, IL 3. Kluge, A. G., and Faris, J. F. (1969) Syst. Zool. 18, 1–32
- 14. Faris, J. F. (1989) Cladistics 5, 417-419
- 15. Madison, W. P., and Madison, D. R. (1992) MacClade: Analysis of Phylogeny and Character Evolution, Version 3, Sinauer Associates, Inc., Sunderland, MA
- Van de Peer, Y., and De Wachter, R. (1994) Comput. Appl. Biosci. 3, 227–230
 Felsenstein, J. (1985) Evolution 39, 783–791 Martin, P. D., Kuchumov, A. R., Green, B. N., Oliver, R. W. A., Braswell, E. H.,
- 18. Wall, J. S., and Vinogradov, S. N. (1996) J. Mol. Biol. 255, 154-169 19. Suzuki, T., Takagi, T., and Ohta, S. (1990) Biochem. J. 266, 221-225
- Zal, F., Suzuki, T., Kawasaki, Y., Childess, J. J., Lallier, F. H., and Tulmond, 20. A. (1997) Proteins 29, 562-574
- 21. Shishikura, F., Ochiai, T., and Yamanaka, I. (1997) Zool. Sci. 14, 923-930
- Gotoh, T., Suzuki, T., and Takagi, T. (1991) in Structure and Function of Invertebrate Oxygen Carriers (Vinogradov S. N., and Kapp O. H., eds) pp. 279-283, Springer-Verlag New York Inc., New York
- 23. Takagi, T., Iwaasa, H., Ohta, S., and Suzuki, T. (1991) in Structure and Function of Invertebrate Oxygen Carriers (Vinogradov S. N., and Kapp O. H., eds) pp. 245-249, Springer-Verlag New York Inc., New York
- 24. Shishikura, F., and Nakamura, M. (1996) Zool. Sci. 13, 849-856
- 25. Shishikura, F. (1996) Zool. Sci. 13, 551-558

The Evolution of Extracellular Hemoglobins of Annelids, Vestimentiferans, and Pogonophorans

Enrico Negrisolo, Alberto Pallavicini, Roberto Barbato, Sylvia Dewilde, Anna Ghiretti-Magaldi, Luc Moens and Gerolamo Lanfranchi

J. Biol. Chem. 2001, 276:26391-26397. doi: 10.1074/jbc.M100557200 originally published online April 6, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M100557200

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 20 references, 4 of which can be accessed free at http://www.jbc.org/content/276/28/26391.full.html#ref-list-1