Diffusion of highly charged intrinsically disordered proteins

Barbara Klepka, Agnieszka Stelmachowska, Anna Niedźwiecka
e-mail: annan@ifpan.edu.pl

Laboratory of Biological Physics, Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Introduction

Coral acid-rich proteins (CARP) of skeletal organic matrix (SOM) are responsible for biomineralisation. It is anticipated that their interactions with calcium cations could determine the crystal form of the skeleton, *i. e.* that the proteins control early stages of crystallisation by driving the amorphous calcium carbonite to calcite or aragonite crystal nuclei. Little is known about such proteins at the molecular level. Only four CARP SOM have been cloned and partially characterised.

Here we show expression, purification and hydrodynamic characterisation of two new CARPs SOM, an N-terminal fragment of secreted acidic protein 1A (B3EWZ0 in Uniprot) and aspartic and glutamic acid-rich protein (B7W112), discovered recently in *Acropora millepora* [1], a model organism to study environmental changes in the oceans due to global climate-warming. They are studied in two forms, with and without a SUMO-tag.

At issue is how to analyse diffusion of such biopolymers to observe changes in the Stokes radius upon interactions with calcium ions. The classical polymer approach is based on the relation of R_h *vs.* the number of units in the polymer [2]. Both denatured proteins in the random coil form [3] and proteins that have well defined, ordered 3D structures [4] can be described by this formalism but with different critical exponent. However, intrinsically disordered proteins (IDPs) have more intricate characteristics. A corrected model for IDPs was proposed [4] taking into account sequence–dependent compaction.

The goal of this study is to compare diffusion properties of our CARPs in relation to standard folded proteins by means of fluorescence correlation spectroscopy.

Coral Acid-Rich Proteins

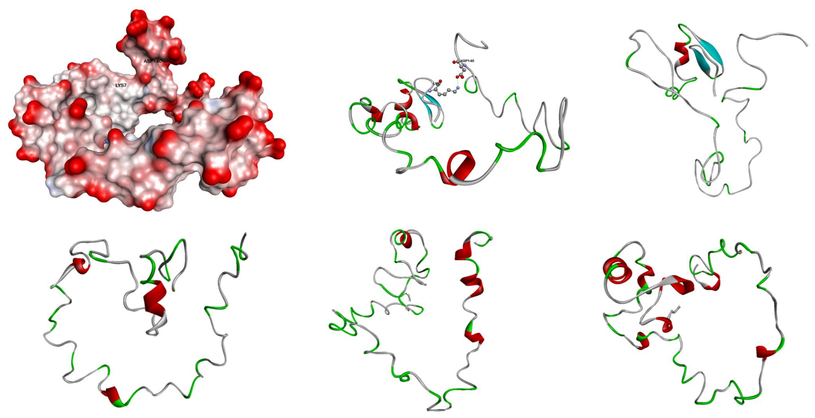

Figure 1. Five best example models of possible 3D structures of the *A. millepora* secreted acidic protein 1A predicted *ab initio* [8]. Residual secondary structural elements are visible as red α-helices, cyan β-strands and green β-turns. In the model 1., a salt bridge links Lys7 to Asp140. Upper left panel – solvent accessible protein surface of the model 1. is coloured according to the charge distribution.

The amino acid sequences, isoelectric points and net charges at pH 8 of the proteins are given below. Dark red: His-tag, blue: SUMO-tag. The folded part of the SUMO-tag is underlined.

Secreted acidic protein 1A

B3EWZ0 pI 3.39 Q @ pH 8.0 = -62.6

GLPLPLKNENAIVDGDGTSVVTTKEDASTIFERDPNPANQVSAMVTGVILDENGDPGESDESVENVDNDGEGGDKDDDKNGEDNDLDNKEHEEEKGDDDR
GDDEEEDDAEGDNDSNDNEGDDDDDDDSGDDDDVDESGADEDDDDDSGD

His-SUMO-B3EWZ0 pI 4.01 Q @ pH 8.0 = -57.8

MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPED
LDMEDNDIIEAHREQIGGGLPLPLKNENAIVDGDGTSVVTTKEDASTIFERDPNPANQVSAMVTGVILDENGDPGESDESVENVDNDGEGGDKDDDKNGE
DNDLDNKEHEEEKGDDDRGDDEEEDDAEGDNDSNDNEGDDDDDDDSGDDDDVDESGADEDDDDDSGD

Aspartic and glutamic acid-rich protein

B7W112 pI 3.94 Q @ pH 8.0 = 151.6

SPLRNRFNEDHDEFSKDDMARESFDTEEMYNAFLNRRDSSESQLEDHLLSHAKPLYDDFFPKDTSPDDDEDSYWLESRNDDGYDLAKRKRGYDDEEAYDD
FDEVDDRADDEGARDVDESDFEEDDKLPAEEESKNDMDEETFEDEPEEDKEEARFEEFAEDERADERDDDADFDFNDEEDEDEVDNKAESDIFTPEDFAG
VSDEAMDNFRDDNEEEYADESDDEAEEDSEETADDFEDDPEDESDETFRDEVEDESEENYQDDTEEGSEIKQNDETEEQPEKKFDADKEHEDAPEPLKEK
LSDESKARAEDESDKSEDAAKEIKEPEDAVEDFEDGAKVSEDEAELLDDEAELSDDEAELSKDEAEQSSDEAEKSEDKAEKSEDAEPELPKEKSEDEA
EKAEDAAGKESNDEGKKREDEAVKSKGIARDESEFAKAKKSNLALKRDENRPLAKGLRESAAHLRDFPSEKKSKDAAQGNIENELDYFKRNAFADSKDAE
PYEFDK

His-SUMO-B7W112 pI 4.09 Q @ pH 8.0 = -146.8

MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPED
LDMEDNDIIEAHREQIGGSPLRNRFNEDHDEFSKDDMARESFDTEEMYNAFLNRRDSSESQLEDHLLSHAKPLYDDFFPKDTSPDDDEDSYWLESRNDDG
YDLAKRKRGYDDEEAYDDFDEVDDRADDEGARDVDESDFEEDDKLPAEEESKNDMDEETFEDEPEEDKEEARFEEFAEDERADERDDDADFDFNDEEDED
EVDNKAESDIFTPEDFAGVSDEAMDNFRDDNEEEYADESDDEAEEDSEETADDFEDDPEDESDETFRDEVEDESEENYQDDTEEGSEIKQNDETEEQPEK
KFDADKEHEDAPEPLKEKLSDESKARAEDESDKSEDAAKEIKEPEDAVEDFEDGAKVSEDEAELLDDEAELSDDEAELSKDEAEQSSDEAEKSEDKAEKS
EDEAELSEDEAKQSEDEAEKAEDAAGKESNDEGKKREDEAVKSKGIARDESEFAKAKKSNLALKRDENRPLAKGLRESAAHLRDFPSEKKSKDAAQGNIE
NELDYFKRNAFADSKDAEPYEFDK

Spectroscopic methods

Labeling of proteins with fluorescent dye

All the proteins should be fluorescently labeled with a stable and photobleaching-resistant fluorescence dye before the measurements. Alexa Fluor® 488 (A488) was used in the study.

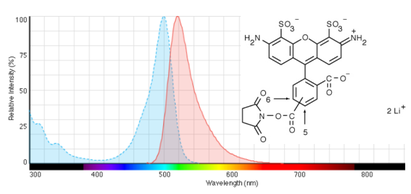

Figure 2. The NHS ester (succinimidyl ester) of A488 is a tool for conjugating the dye to the primary amines (R-NH_2) of proteins, *i. e.* the side chains of lysine and arginine residues and the N-terminus. The absorption and emission spectra are shown together with the chemical formula of the compound.

Fluorescence Correlation Spectroscopy

Diffusion constants can be calculated from the average residence time in the focal volume, τ_D, measured by FCS [5] and the structural parameter of the experimental setup, ω_0, which can be estimated based on the known D value for A488 [6]:

$$D = \frac{{\omega_0}^2}{4\tau_D}$$

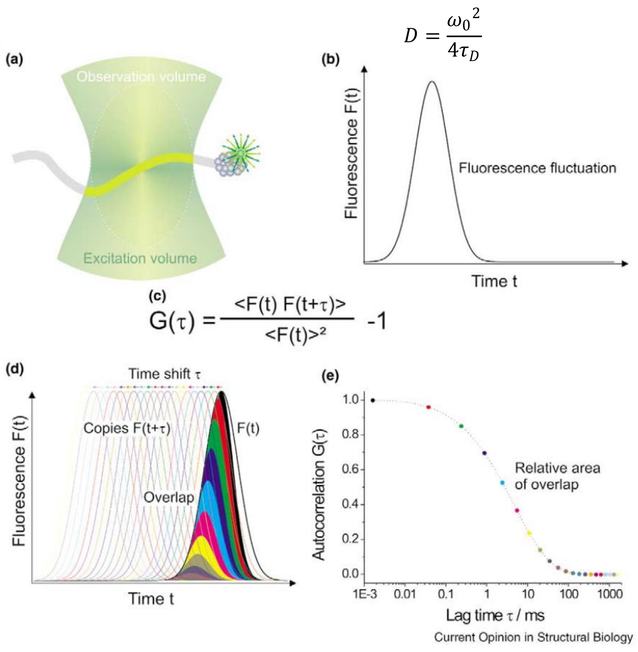

Figure 3. Graphical representation of fluorescence correlation spectroscopy. The figure is taken from [5]. Fluorescent molecules diffusing through the focal volume (a) give rise to fluctuations in the detected signal F(t) (b). These can be analyzed by FCS, which determines the self-similarity of the signal after a lag time τ. The mathematical expression for autocorrelation is given in (c) and demonstrated graphically in (d); a copy of the observed fluctuation F(t + τ) is shifted by a lag time τ relative to the original signal. Both are multiplied and the area under the resulting curve (*i.e.* the overlap integral) gives the value of the autocorrelation for τ (e).

Sample preparation

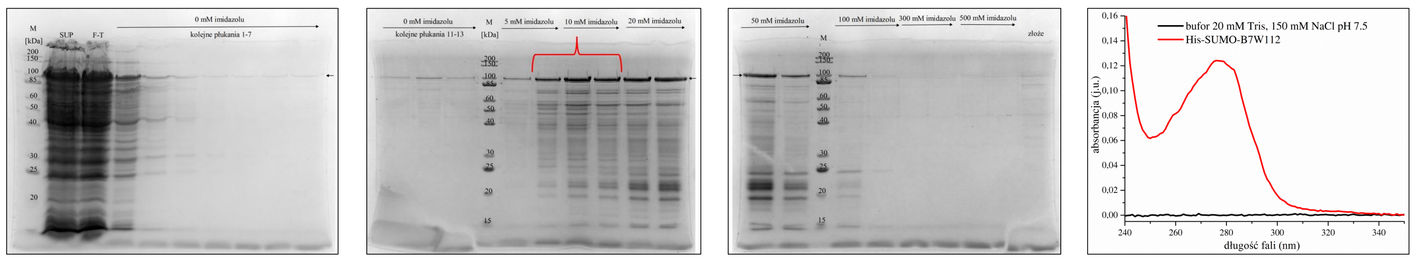

Figure 4. From left to right: the proteins in fusion with the SUMO-tag providing higher expression were purified by histidine – nickel affinity chromatography and analysed by the SDS PAGE. Best fractions (marked by the red frame) were merged. The concentration was determined from UV absorption spectrum.

Table 1. Parameters of proteins used in this study.

Protein	Molar mass [g/mol]	Molar extinction coefficient [$cm^{-1}\,M^{-1}$]	c_{prot} [μM]	c_{A488} [μM]	A488:prot
Standard proteins					
Lysozyme	14 313	37 720	5.2	8.2	2
α-Chymotrypsinogen A	25 666	50 273	21	21	1
Bovine Serum Albumin	66 562	41 863	9.5	18	2
Apoferritin (24-mer)	479 049	348 060	3.7	0.62	0.2
Coral acid-rich intrinsically disordered proteins					
B7W112	58 339	20 400	1.8	0.77	0.4
His-SUMO-B7W112	71 746	21 890	4.4	4.6	1
B3EWZ0	15 856	- *			
His-SUMO-B3EWZ0	29 263	1 490			

* B3EWZ0 protein sequence does not contain any aromatic residue

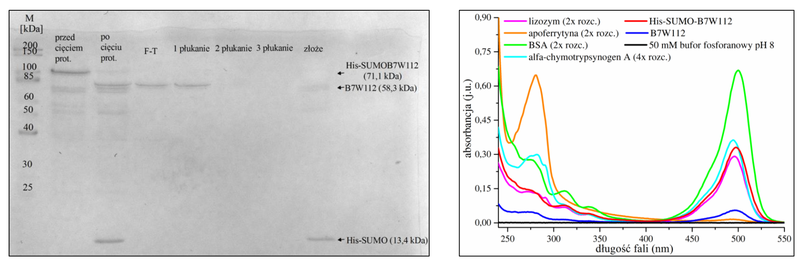

Figure 5. Left panel: the samples of SUMO-fused proteins were enzymatically digested by specific SUMO-protease, purified again by Ni-chromatography to remove SUMO, and analysed by the SDS PAGE. Each protein has been labeled by the NHS ester of the A488 dye (Fig. 2). Right panel: The fractions of the highest A488:protein stoichiometry were selected based on the UV absorption spectra for use in the FCS measurements.

Results

The residence times, τ_D, and then the diffusion coefficients for each protein were determined based on 150 measurements. The Stokes radii for the coral acid-rich proteins were calculated from the Stokes-Einstein equation, assuming that the T/η value was constant and equal for all sample droplets, and can be estimated from the measurements done for standard proteins of known R_h values [7]:

$$D = \frac{k_B T}{6\pi\eta R_h}$$

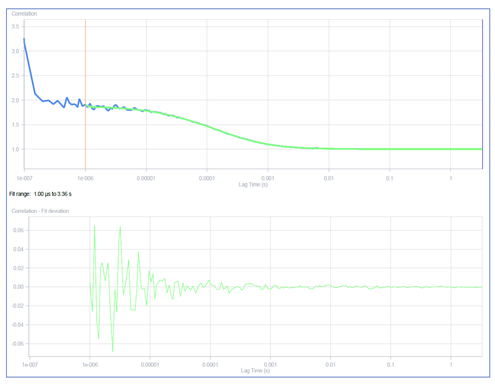

Figure 6. Example average data (blue) and autocorrelation curve (green, uncorrected) taking into account the presence of two diffusing species (the labeled protein and the free dye molecule) and the A488 triplet state.

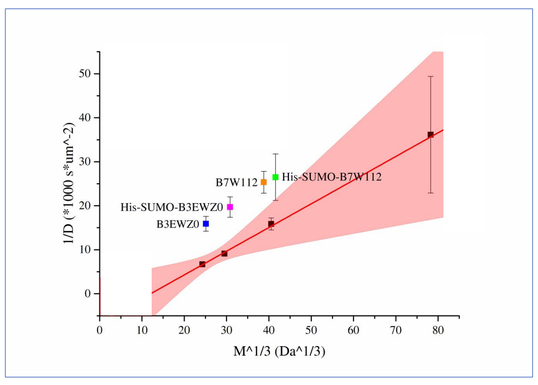

Figure 7. The FCS results. Linear dependence of 1/D vs. M^(1/3) for the standard proteins (black squares). The coral acis-rich intrinsically disordered proteins (colour squares) outlie the 99% CI of the standard line.

Table 2. Hydrodynamic parameters of the proteins.

	Calculated from polymer model [2,3,4]				Measured by FCS		
	R_h^f [Å]	R_h^d [Å]	R_h^{IDP} [Å]	$R_h^{IDP-corr}$ [Å]	τ_D [μs]	D [$\frac{\mu m^2}{s}$]	R_h [Å]
A488					30 ± 1	435*	
Protein							
Lysozyme	20	-	-	-	89 ± 4	150 ± 10	21 +
α-Chymotrypsinogen A	24	-	-	-	127 ± 4	109 ± 7	26 +
Bovine Serum Albumin	30	-	-	-	198 ± 14	63 ± 6	35 +
Apoferritin (24-mer)	53	-	-	-	480 ± 170	30 ± 10	66 +
B3EWZ0	**20**	**36**	**32**	**44**	**200 ± 20**	**63 ± 7**	**43 "**
His-SUMO-B3EWZ0	**24**	**50**	**43**	**51**	**250 ± 30**	**51 ± 6**	**53 "**
B7W112	**29**	**71**	**59**	**116**	**350 ± 30**	**39 ± 4**	**69 "**
His-SUMO-B7W112	**31**	**80**	**66**	**119**	**369 ± 70**	**38 ± 8**	**72 "**

* the value from [8]
\+ R_h values from Hydrodynamic Radius Converter [7]
" estimated values based on FCS results with the uncerntainty of ± 10-20%

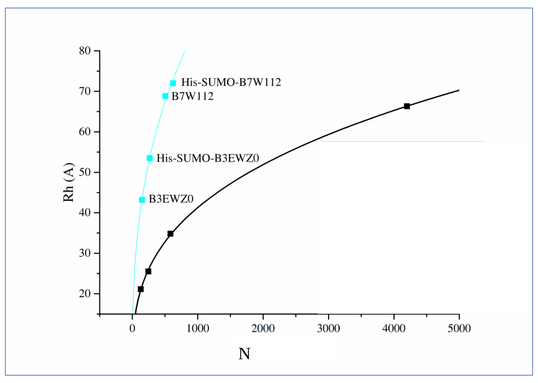

Figure 8. Estimated Stokes radii of the coral acid-rich proteins (cyan squares) and standard proteins (black squares) represented as a function of the biopolymer length. Both classes follow the power function $R_h = R_0 N^{\nu}$ but with different coefficients.

Conclusions

1. We present preliminary data obtained for hydrodynamic parameters of highly charged biopolymers, such as coral acid-rich intrinsically disordered proteins from *Acropora millepora*. These proteins has been cloned and characterised for the first time.
2. Contrary to the proteins that are folded in a well defined 3D structures and diffuse according to the Stokes-Einstein equation, the intrinsically disordered acidic proteins behaves hydrodynamically in a distinct manner, *i. e.* their Stokes radii are greater and the diffusion coefficients smaller that those for a corresponding ordered protein of the same molecular mass.
3. The corrected polymer model [4] is insufficient to characterise and explain the diffusion coefficient of the intrinsically disordered proteins of high molecular masses and very high like charges.
4. Based on the estimated R_h values, it can be concluded that all acidic protein studied herein diffuse as denatured random coils, which results most probably from their extreme electrostatic properties. The presence of the folded SUMO-fragment in the protein constructs has a negligible influence on the hydrodynamic properties, within the estimated experimental uncertainty.
5. Fluorescence correlation spectroscopy is a method of choice that enables us to discern the intrinsically disordered proteins from the ordered, structured proteins in a broad range of molecular masses, from the six-membered histidine tag up to 24-mer of apo-ferritine, which is unavailable for other experimental approaches.
6. These results form a promissing starting point for studies of interactions of the coral acid-rich intrinsically disordered proteins with calcium cations upon early stages of biomineralisation at the molecular level.

Literature

[1] Ramos-Silva, P. *et al.* The skeletal proteome of the coral *Acropora millepora*: The evolution of calcification by co-option and domain shuffling. **Mol. Biol. Evol.** (2013)
[2] Le Guillou, J. C. & Zinn-Justin, J. Critical exponents for the n-vector model in three dimensions from field theory. **Phys. Rev. Lett.** (1977)
[3] Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. **Proc. Natl. Acad. Sci. U. S. A.** (2004)
[4] Marsh, J. A. & Forman-Kay, J. D., Sequence determinants of compaction in intrinsically disordered proteins. **Biophys. J.** (2010)
[5] E. Haustein, P. Schwille, Single-molecule spectroscopic methods, **Curr Opin Struct Biol.** (2004) 14:531–540
[6] Petrášek, Z. & Schwille, P. Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy. **Biophys. J.** (2008)
[7] Tyn, M. T. & Gusek, T. W., Prediction of diffusion coefficients of proteins. **Biotechnol. Bioeng.** (1990)
[8] Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. **Proteins.** (2012) 80(7): 1715-35.

Acknowledgements

This work was supported by Polish NCN within the Sonata-Bis grant no. 2016/22/E/NZ1/00656. The studies were performed in the NanoFun laboratories co-financed by the European Regional Development Fund within the Innovation Economy Operational Programme POIG.02.02.00-00-025/09. We thank Magdalena Duszka and Remigiusz Worch for helpful assistance.