

Finja Krebs
 Christin Scheller
 Kristina Grove-Heike
 Lena Pohl
 Hermann Wätzig 

Institute of Medicinal and
 Pharmaceutical Chemistry,
 Technische Universität
 Braunschweig, Braunschweig,
 Germany

Received January 15, 2021

Revised January 29, 2021

Accepted February 1, 2021

Short Communication

Isoelectric point determination by imaged CIEF of commercially available SARS-CoV-2 proteins and the hACE2 receptor

Abstract In order to contribute to the scientific research on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), we have investigated the isoelectric points (pI) of several related proteins, which are commercially available: the receptor-binding domain (RBD) with His- and Fc-tag, the S1 subunit with His-tag, the S1/S2 subunits with His-tag and the human angiotensin-converting enzyme 2 (hACE2) with His-tag. First, the theoretical pI values, based on the amino acid (AA) sequences of the proteins, were calculated using the ProtParam tool from the Bioinformatics Resource Portal ExPASy. The proteins were then measured with the Maurice imaged CIEF system (native fluorescence detection), testing various measurement conditions, such as different ampholytes or ampholyte mixtures. Due to isoforms, we get sections with several peaks and not just one peak for each protein. The determined pI range for the RBD/Fc is 8.24–9.32 (theoretical pI: 8.55), for the RBD/His it is 7.36–9.88 (8.91) and for the S1/His it is 7.30–8.37 (7.80). The pI range of the S1/S2/His is 4.41–5.87 (no theoretical pI, AA sequence unknown) and for hACE2/His, the determined global range is 5.19–6.11 (5.60) for all experimental conditions chosen. All theoretically derived values were found within these ranges, usually close to the center. Therefore, we consider theoretical values as useful to make predictions about the isoelectric points of SARS-CoV-2 proteins. The experimental conditions had only a minor influence on the pI ranges obtained and mainly influenced the peak shapes.

Keywords:

CIEF / COVID-19 / hACE2 / Isoelectric point / SARS-CoV-2 proteins

DOI 10.1002/elps.202100015

The Coronavirus Disease 2019 (COVID-19) pandemic is currently shaking up the whole world. Since the beginning of the outbreak, research has been conducted into medicines and vaccines against the virus in the field of pharmaceutical research. For research purposes, there are several significant proteins of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on the market. Therefore, our intention was to contribute to the progress of research in this field with the means available to us. In this short communication, the results of the capillary isoelectric focusing (CIEF) measurements of different sections of the SARS-CoV-2 spike protein and the human angiotensin-converting enzyme 2 (hACE2) are presented.

SARS-CoV-2 uses the receptor-binding domain (RBD) of its surface glycoprotein (spike protein) to enter the human body via the hACE2 receptor [1]. The spike protein is com-

posed of two functional subunits: S1 and S2. While the S1 subunit binds to the hACE2 receptor, the S2 subunit is implicated in the merging of the viral and human cell membranes [2]. For more information on the physicochemical properties of SARS-CoV-2, please refer to the review of Scheller et al. [3].

In the CIEF, the isoelectric focusing takes place in a capillary, as a charge-based analysis via capillary electrophoresis. The CIEF method was implemented by Hjerten et al. in the 1980s and can be seen as an improved, new version of the conventional isoelectric focusing in slab gels [4,5]. The advantages of CIEF over IEF with slab gels are, for example, a smaller sample volume required, shorter analysis times, and higher sensitivity [6]. Using CIEF, a pH gradient is built up in the capillary. The ampholytic analytes migrate in this pH gradient and then remain in the pH zone that corresponds to their pI value [7]. It is also possible in CIEF to use immobilized pH gradients as applied in slab gels. This technique is still under development, but can bring many advantages [6, 8–10]. In this case, the imaged CIEF with carrier ampholytes was used, i.e., an on-line imaging detection system that does not require the mobilization of the analytes after focusing [11].

Since the isoelectric point of proteins is an important property, we sought to determine it using the Maurice (imaged) CIEF system from ProteinSimple, a Bio-Techne brand.

Correspondence: Prof. Dr. Hermann Wätzig, Institute of Medicinal and Pharmaceutical Chemistry, Technische Universität Braunschweig, Beethovenstraße 55, 38106 Braunschweig, Germany
Email: h.waetzig@tu-bs.de

Abbreviations: hACE2, human angiotensin-converting enzyme 2; RBD, receptor-binding domain; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

Table 1. Experimental conditions for the determination of the isoelectric points via Maurice CIEF of the SARS-CoV-2 proteins and the hACE2 receptor from R&D Systems; sample Load: 55 s, anolyte: 80 mM phosphoric acid in 0.1% methyl cellulose, catholyte: 100 mM NaOH in 0.1% methyl cellulose, V_{well} : 100 μL , 40% SimpleSol Protein Solubilizer (except for experiment f of hACE2/His), 0.35% methyl cellulose, 10 mM arginine (except for experiment f of hACE2/His); ρ = mass concentration of the respective protein in mg/mL

	RBD/Fc	RBD/His	S1/His	S1/S2/His	hACE2/His
Catalog number	#10499-CV	#10500-CV	#10522-CV	-	#933-ZV
ρ_{initial} [mg/mL]	2.100	1.860	0.350	0.470	0.468
Buffer	PBS pH 7.4	PBS pH 7.4	PBS pH 7.4	PBS pH 7.4	12.5 mM Tris, NaCl, ZnCl ₂ and glycerol
Theoretical pI	8.55	8.91	7.80	-	5.60
Separation	1 min 1500 V 7.5 min 3000 V	1 min 1500 V 7.5 min 3000 V	1 min 1500 V 8.5 min 3000 V	1 min 1500 V 11 min 3000 V	1 min 1500 V a-e: 20 min, f: 15 min 3000 V
Fluorescence exposure time	50 s	50 s	50 s	50 s	a: 5 s, b: 50 s, e: 30 s, f: 10 s
ρ_{final} [mg/mL]	0.036	0.035	0.040	0.080	a: 0.040 b: 0.016 e: 0.041 f: 0.098
Urea [mol/L]	-	-	-	-	e: 2 f: 7
Ampholyte	a/b: 3–10 (4%) c: 3–10 (1%), 8–10.5 (3%) d: 8–10.5 (4%)	a/b: 3–10 (4%) c: 3–10 (1%), 8–10.5 (3%)	a/b: 3–10 (4%) c: 3–10 (1%), 8–10.5 (3%) d: 8–10.5 (4%)	a: 3–10 (4%)	a/b/e: 3–10 (4%) f: 2.5–5 (3%), 5–8 (2%)
pI markers (1% each)	a: 3.38, 9.99 b/c/d: 7.05, 9.50	a: 3.38, 9.99 b/c: 7.05, 10.17	a: 3.38, 9.99 b/c/d: 7.05, 9.50	a: 3.38, 9.99	a: 3.38, 9.99 b/e/f: 4.05, 7.05

We used the SARS-CoV-2 proteins currently available on the market, namely the RBD with His- and Fc-tag, the S1 subunit with His-tag, and the S1/S2 subunits with His-tag.

As the virus binds to the hACE2 in the human body, this receptor was included in the study (all proteins were provided by R&D Systems, a Bio-Techne brand, Minneapolis, Minnesota, USA).

First, the theoretical pI values of the proteins were calculated using the ProtParam tool from the Bioinformatics Resource Portal ExPASy [12]. The basis for this was the given AA sequence of the corresponding protein, which can be found in the R&D Systems catalog [13] on the page of the respective protein (see catalog numbers in Table 1).

The SARS-CoV-2 RBD/Fc is derived from the surface glycoprotein (NCBI Reference Sequence: YP_009724390.1) Arg319-Phe541, which is linked to human IgG1 (Pro100-Lys330) via the sequence IEGRMD. The pI value calculated using this sequence is 8.55. The His-tagged SARS-CoV-2 RBD, with a theoretical pI of 8.91, is also the surface glycoprotein Arg319-Phe541, but with a C-terminal 6-His-tag. The S1/His subunit is also the surface glycoprotein, but the sequence Val16-Pro681, with a C-terminal 6-His-tag. A pI of 7.80 can be calculated from this sequence. The S1/S2/His protein is not yet commercially available and therefore not included in the R&D Systems catalog. For this reason, no theoretical pI value can be calculated as the exact AA sequence is unknown. The hACE2/His receptor consists of the AA sequence of the hACE2 protein (UniProtKB/Swiss-Prot: Q9BYF1) Gln18-Ser740, with a C-terminal 10-His-tag. The pI value of this sequence calculated via ExPASy is 5.60.

Subsequently, it is investigated to what extent the theoretical pI values agree with the experimentally measured ones.

For the Maurice CIEF system, the Maurice CIEF cartridges and the Maurice CIEF method development kit (including all reagents used, such as the anolyte, catholyte, ampholytes, pI markers, SimpleSol, etc.) were used (provided by ProteinSimple, a Bio-Techne brand, San Jose, California, USA). With native fluorescence at an excitation wavelength of 280 nm, emitted light at 320–450 nm was used for detection [14].

The measurements were performed under the experimental conditions shown in Table 1.

The preparation of the samples is now explained using the sample from experiment a of the RBD/Fc as an example. 8.5 μL of protein, dissolved in PBS buffer ($\rho_{\text{initial}} = 2.100 \text{ mg/mL}$), was mixed with 35 μL of 1% methyl cellulose, 40 μL SimpleSol, 4 μL ampholyte 3–10, 2 μL of 500 mM arginine, 1 μL each of pI markers 3.38 and 9.99 and 8.5 μL DI water, resulting in a total volume of 100 μL . Samples were then vortexed and centrifuged to be pipetted into a 96-well plate.

In order to get a general overview of the pI value of the respective protein, the wide range ampholyte with markers at the edges of the range was first selected (a).

In a second step, pI markers whose pI values were as close as possible to those of the respective protein were chosen (b).

Next step was to improve the resolution of the peaks by using ampholyte mixtures. For this purpose, wide range ampholytes and narrow range ampholytes were mixed in a ratio

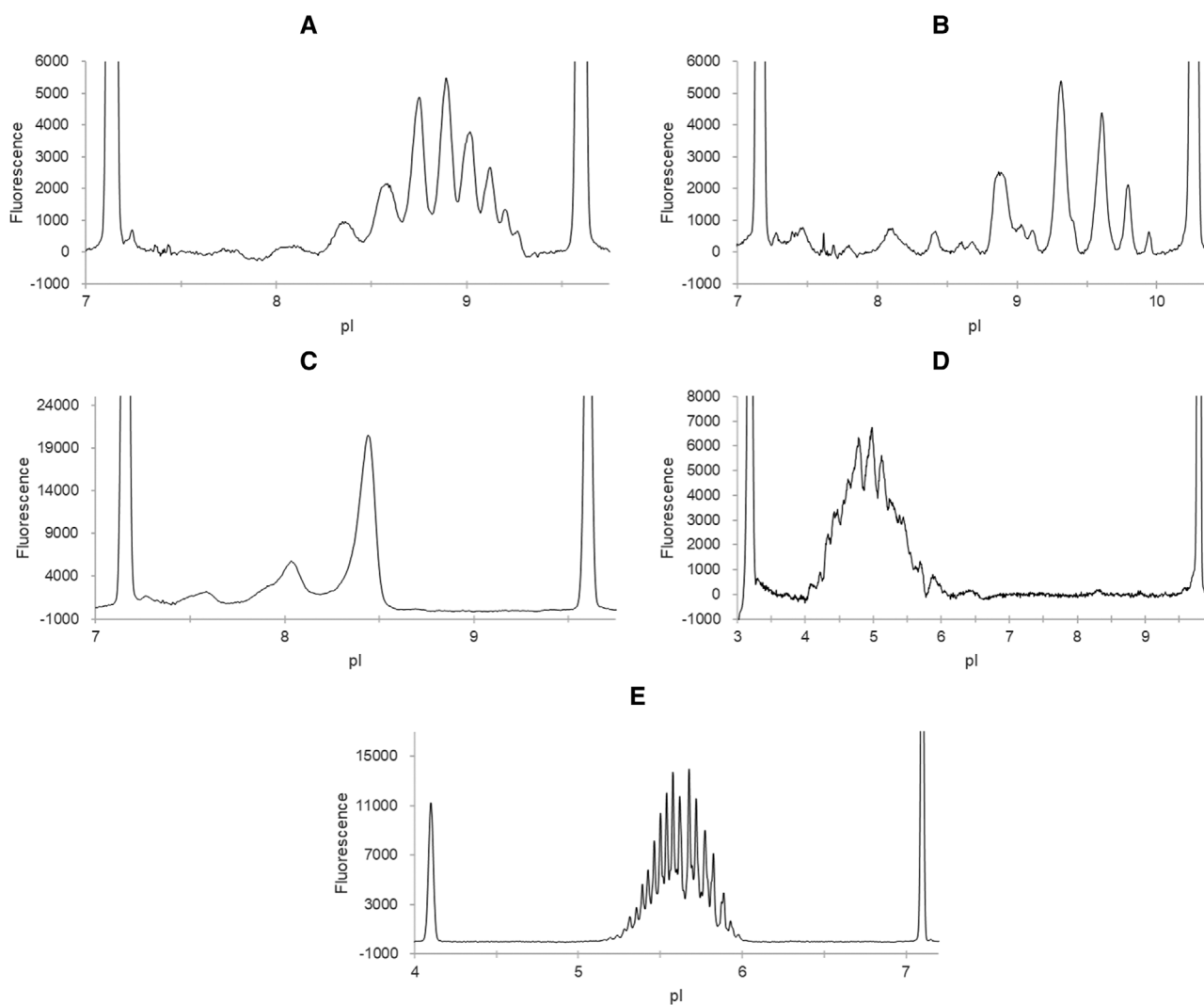


Figure 1. Electropherograms of the Maurice CIEF measurements. **A:** SARS-CoV-2 RBD/Fc experiment **b;** **B:** RBD/His experiment **b;** **C:** S1/His experiment **b;** **D:** S1/S2/His experiment **a;** **E:** hACE2/His experiment **f**. For better visualization, the y-axes were displayed only in the area of fluorescence of the sample peaks.

of 1:3 (c). The mixing ratio was chosen based on the work of Kahle et al. [15].

An additional approach was to only use narrow range ampholytes (d). Because the peak profile of the hACE2/His protein was not as reproducible, this protein was also measured with an addition of urea. For e, the two protein solubilizers SimpleSol and urea were used. Since the addition of SimpleSol and urea did not lead to a good peak shape either, SimpleSol was subsequently omitted and the concentration of urea was increased. In addition, several ampholyte mixtures were tested. Finally, a concentration of 7 M urea and a mixture of ampholytes 2.5–5 (3%) and 5–8 (2%) was used and thus well reproducible peaks were obtained (f).

Since the proteins are very expensive and only available in small quantities, further optimization attempts were not conducted. Furthermore, there were no repetitions of

non-evaluable measurements, as the proteins will be used for future affinity capillary electrophoresis (ACE) experiments, in which the binding behavior of the SARS-CoV-2 proteins to the hACE2 receptor will be investigated.

Figure 1 shows the electropherograms of each measured protein in the particular experiment where the protein yielded the most reproducible peaks with the best peak shape.

The electropherograms show, that the proteins produce very different peak profiles. In case of the SARS-CoV-2 RBD/Fc (Figure 1A), the individual isoforms can be seen very clearly. SARS-CoV-2 RBD/His (Figure 1B), on the other hand, shows very small peaks in the approximate range of pI 7 to 8.5 and then some larger ones in the range of pI 8.5 to 10.

The S1/His subunit (Figure 1C) produces three peaks, whose areas increase with pI, so the one with the highest pI is the largest at about 8.4.

Table 2. Results of the Maurice CIEF measurements of SARS-CoV-2 proteins and the hACE2 receptor (experiments a-f). Given are the mean values of the isoelectric points determined from a certain number of repetitions (n) and the relative standard deviation (RSD) in %

a		RBD/Fc (theoretical pI: 8.55)								n = 5	
mean	8.359	8.572	8.782	8.924	9.052	9.153	9.250	9.316			
RSD [%]	0.219	0.222	0.064	0.093	0.142	0.088	0.085	0.061			
b										n = 8	
mean	8.254	8.478	8.649	8.794	8.915	9.021	9.103	9.164			
RSD [%]	0.157	0.158	0.029	0.032	0.048	0.044	0.051	0.049			
c										n = 6	
mean	8.390	8.633	8.757	8.886							
RSD [%]	0.054	0.134	0.036	0.061							
d										n = 6	
mean	8.235	8.429	8.629	8.877							
RSD [%]	0.050	0.159	0.007	0.069							
a		RBD/His (theoretical pI: 8.91)								n = 7	
mean	7.412	7.738	8.025	8.338	8.799	9.229	9.503	9.691	9.833		
RSD [%]	0.114	0.137	0.201	0.203	0.081	0.110	0.084	0.073	0.072		
b										n = 8	
mean	7.357	7.685	7.991	8.314	8.766	9.217	9.509	9.695	9.849		
RSD [%]	0.135	0.117	0.104	0.117	0.212	0.085	0.061	0.047	0.131		
c										n = 6	
mean	7.925	8.199	8.440	8.733	9.086	9.316	9.477	9.702	9.875		
RSD [%]	0.060	0.307	0.098	0.125	0.106	0.058	0.070	0.075	0.056		
a		S1/His (theoretical pI: 7.80)									n = 8
mean	7.545	7.994	8.372								
RSD [%]	1.394	1.349	1.376								
b										n = 8	
mean	7.479	7.932	8.334								
RSD [%]	0.103	0.037	0.025								
c										n = 6	
mean	7.303	7.706	8.064								
RSD [%]	0.040	0.066	0.015								
d										n = 6	
mean	7.681	7.837	8.012	8.134							
RSD [%]	0.048	0.185	0.039	0.114							
a		S1/S2/His								n = 6	
mean	4.406	4.507	4.605	4.817	4.974	5.157	5.309	5.423	5.608	5.872	
RSD [%]	0.154	0.059	0.040	0.106	0.165	0.059	0.252	0.176	0.481	0.109	
a		hACE2/His (theoretical pI: 5.60)								n = 8	
mean	5.605	5.673	5.790	5.846	5.908	5.939	6.004	6.109			
RSD [%]	0.402	0.807	0.438	0.119	0.129	0.046	0.085	0.094			
b										n = 8	
mean	5.355	5.428	5.497	5.576	5.692						
RSD [%]	0.133	0.087	0.102	0.123	0.090						
e										n = 5	
mean	5.370	5.416	5.465	5.528	5.581	5.618	5.751				
RSD [%]	0.057	0.020	0.027	0.076	0.054	0.054	0.116				
f										n = 12	
mean	5.188	5.231	5.263	5.304	5.341	5.375	5.413	5.451	5.489	5.526	
RSD [%]	0.086	0.022	0.076	0.053	0.057	0.053	0.048	0.044	0.048	0.049	
mean	5.570	5.626	5.670	5.723	5.771	5.834	5.884	5.928			
RSD [%]	0.042	0.044	0.036	0.036	0.032	0.037	0.030	0.019			

With the S1/S2/His subunit (Figure 1D), it was challenging to get peaks at all. Using a very high concentration, a broad peak with some spikes is finally obtained.

After some optimization of the experimental conditions, the hACE2/His receptor (Figure 1E) shows a very reproducible peak profile, in which the different isoforms can be well recognized.

With the peak profiles of the individual proteins in mind, we now take a look at the determined isoelectric points of the proteins. Table 2 summarizes the results of the CIEF measurements of the five proteins.

With regard to the results, the RBD/Fc protein will be discussed first. The theoretically calculated pI value of this protein is 8.55. In all experiments (a–d) pI values in the range of 8.24 to 9.32 were obtained. The measured values are therefore in the range around the theoretically calculated pI, the distribution is probably due to isoforms.

Next, we come to the results of the SARS-CoV-2 RBD with His-tag. For the two measurements in ampholyte 3–10 (a, b) we have very similar pI values in the range between about 7.36 and 9.85. The pI values in measurement c, i.e., the measurement in the ampholyte mixture, lie between 7.93 and 9.88, which is probably due to the fact that the pI gradient in the ampholyte mixture at the edges of the narrow range ampholyte is not linear. The measurement with the narrow range ampholyte (d) could not be evaluated for this protein and was not repeated due to the small amount of material available. The theoretical pI of this protein is 8.91. In the measurements, this is in the range of the large peaks of the protein. Overall, the experimentally determined pI values can be found around the theoretically determined value in the range ± 1 pH unit for this protein.

Looking at the S1 subunit with His-tag, the pI range of the experiments a and b again hardly differs, so the pI values seem to behave nearly linear over the whole pI area. The determined range in these two experiments is about 7.48 to 8.37. The differences in comparison to the measurement in the ampholyte mixture (c) are probably due to the fact that the pI value of the protein is located at the border of the narrow range ampholyte (8–10.5) and pI values in this border area are probably somewhat distorted.

The pI values determined using the narrow range ampholyte in experiment d lie in the range from 7.68 to 8.13, i.e., the range is significantly smaller than in the previous measurements. The theoretically calculated pI of the S1 subunit with His-tag is 7.80 and thus approximately in the middle of the experimentally determined ranges.

When measuring the S1/S2 subunit with His-tag, the first problem was that no peaks could be found. The concentration of the protein was then continuously increased until peaks could finally be identified. These peaks, shown in the electropherogram in Figure 1D, could also be described as very broad and flat. Because of the high sample consumption due to the high concentration and the failed experiments before, only experiment a was performed. The experimentally determined pI values lie in a range from 4.41 to 5.87. Due to the unknown AA sequence, there is no

theoretically calculated pI value and therefore it is not possible to assess whether the values are in the same range.

The last protein investigated is the His-tagged hACE2 receptor, whose theoretically calculated pI value is 5.60. The pI range determined with the wide range ampholyte and the markers 3.38 and 9.99 is 5.61 to 6.11 (experiment a), but the peaks looked as if the protein was aggregated.

Subsequently, a lower concentration was chosen for measurements b, c, and d to avoid aggregation, whereby only experiment b could be evaluated. The pI range determined there is 5.36 to 5.69 and is in the range determined in experiment a, but smaller. Since the peaks still looked as if the protein was aggregated, a measurement with urea and SimpleSol as protein solubilizers was conducted afterwards (e). The reproducibility of the peak profiles was slightly improved and the values (5.37–5.75) hardly differ from those of measurement b. In order to obtain a reproducible peak shape and no aggregation, an attempt was made to optimize the method. In the following, SimpleSol was omitted and the concentration of urea was increased. In addition, several ampholyte mixtures were tried. In experiment f, with 7 M urea and ampholytes 2.5–5 (3%) and 5–8 (2%), peaks with recognizable isoforms of the receptor were finally obtained, which were very reproducible. The pI values in this experiment are between 5.19 and 5.93 and therefore in the range of the previously performed experiments. Thus, the aggregation did not have a strong effect on the pI value determination. Due to the good reproducibility, this method could be well used for quality control of the receptor, since even the smallest impurities would presumably have an effect on the peak shape.

In summary, the theoretical pI values calculated on the basis of the AA sequence are in the same range as those determined experimentally. It is not possible to determine an exact value experimentally for these proteins, since all electropherograms show several peaks due to isoforms. However, these ranges always coincide with the calculated values.

Therefore, the model calculations for these proteins are useful to make predictions about the isoelectric points of the proteins, if they are needed and no time and/or equipment is available for own experiments regarding the pI value determination. The experimental conditions (ampholyte, pI markers, concentration) do not have a particularly large influence on the determined pI value and merely the number and shape of peaks vary. By optimizing the experimental conditions, aggregation can be prevented and the separation of the individual isoforms can be made possible. With further optimization, quality control methods using CIEF can be developed for these proteins.

The Maurice CIEF, including all reagents, and the proteins were provided by ProteinSimple, a Bio-Techne Brand. We would like to thank Susanne Doerks, Carsten Lueck, Udo Burger, and Chris Heger for their tremendous support.

Open access funding enabled and organized by Projekt DEAL.

The authors have declared no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- [1] Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., Velesler, D., *Cell* 2020, *181*, 281–292.e6.
- [2] Matsuyama, S., Ujike, M., Morikawa, S., Tashiro, M., Taguchi, F., *Proc. Natl. Acad. Sci. USA* 2005, *102*, 12543–12547.
- [3] Scheller, C., Krebs, F., Minkner, R., Astner, I., Gil-Moles, M., Wätzig, H., *Electrophoresis* 2020, *41*, 1137–1151.
- [4] Hjerten, S., *J. Chromatogr. A* 1985, *346*, 265–270.
- [5] Hjerten, S., Liao, J.-L., Yao, K., *J. Chromatogr. A* 1987, *387*, 127–138.
- [6] Koshel, B. M., Wirth, M. J., *Proteomics* 2012, *12*, 2918–2926.
- [7] Righetti, P. G., Drysdale, J. W., *J. Chromatogr. A* 1974, *98*, 271–321.
- [8] Yang, C., Zhu, G., Zhang, L., Zhang, W., Zhang, Y., *Electrophoresis* 2004, *25*, 1729–1734.
- [9] Wang, T., Ma, J., Wu, S., Sun, L., Yuan, H., Zhang, L., Liang, Z., Zhang, Y., *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 2011, *879*, 804–810.
- [10] Liu, R., Pan, H., Zheng, Y., Hu, J., Cheddah, S., Wang, W., Wang, Y., Yan, C., *Electrophoresis* 2019, *40*, 1722–1730.
- [11] Wu, J., Pawliszyn, J., *Anal. Chem.* 1992, *64*, 2934–2941.
- [12] Swiss Institute of Bioinformatics (SIB), *ExPASy: ProtParam Tool*, <https://web.expasy.org/protparam/>, last viewed January 2021.
- [13] *R&D Systems*, <https://www.rndsystems.com/>, last viewed January 2021.
- [14] ProteinSimple, *Maurice CIEF specification*, <https://www.proteinsimple.com/maurice.html>, last viewed January 2021.
- [15] Kahle, J., Stein, M., Wätzig, H., *Electrophoresis* 2019, *40*, 2382–2389.