ELSEVIER

Journal of Chromatography A, 833 (1999) 157–168

JOURNAL OF CHROMATOGRAPHY A

Enrichment of low-copy-number gene products by hydrophobic interaction chromatography

Michael Fountoulakis[a,*], Marie-Françoise Takács[b], Béla Takács[b]

[a]*F. Hoffmann-La Roche Ltd., Pharma Division, Preclinical Central Nervous System Research-Gene Technology, Building 93-444, 4070 Basel, Switzerland*

[b]*F. Hoffmann-La Roche Ltd., Pharma Division, Infectious Diseases, Basel, Switzerland*

Received 14 August 1998; received in revised form 27 October 1998; accepted 27 October 1998

Abstract

Enrichment of proteins in solution is the goal of a purification process and often a scientific challenge. We investigated the capacity of hydrophobic interaction chromatogaphy to enrich proteins, potential candidates for novel drug targets. The soluble protein fraction of *Haemophilus influenzae* was fractionated over a TSK Phenyl column and the proteins resolved were analyzed by two-dimensional gel electrophoresis and matrix-assisted laser desorption ionization mass spectrometry. Approximately 150 proteins, bound to the column, were identified, 30 for the first time. Most of the proteins enriched by hydrophobic interaction chromatography were represented by major spots, so that an enrichment of low-copy-number gene products was only partially achieved. The proteins enriched by this chromatographic approach belong to various protein classes, including enzymes, ribosomal proteins and proteins with as yet unknown functions. The results include two-dimensional maps and a list of the proteins enriched by hydrophobic interaction chromatography.

Keywords: Hydrophobic interaction chromatography; *Haemophilus influenzae*; Proteins

1. Introduction

The visualization of low-copy-number proteins, such as transcription factors, may be essential for the detection of novel drug targets. Since not all the proteins of an organism are expressed at the same level at a certain time point, the low-abundance ones cannot be readily visualized on a two-dimensional (2D) gel of the total extract. A specific enrichment of the low-copy-number gene products from larger volumes may then be necessary.

We have studied the capacity of various chromatographic steps to enrich soluble proteins of the bacterium *Haemophilus influenzae*. *H. influenzae* is a microorganism of pharmaceutical interest whose entire genome has been sequenced [1]. It comprises approximately 1740 open reading frames. The proteome of the microorganism has been analyzed by 2D gels and about 400 proteins have been identified so far [2–10].

The chromatographic steps we used in previous experiments were heparin chromatography and chromatofocusing [3,5,6,9]. Heparin shows an affinity for nucleic acid binding proteins, presumably by steric interactions. It also binds proteins on the basis of ionic interactions. Using this chromatographic ap-

*Corresponding author. e-mail: michael.fountoulakis@roche.com

0021-9673/99/$ – see front matter
PII: S0021-9673(98)00929-7

proach, we enriched several low-abundance proteins of the microorganism. Approximately 40% of the proteins bound to the column were nucleic acid binding proteins, in particular basic ribosomal proteins [3,5,9]. Chromatofocusing can efficiently separate complex protein mixtures on the basis of their charge. This resulted in the enrichment of many low-abundance proteins, mainly acidic ones, the majority were enzymes with various catalytic activities [6]. Recently, we used chromatography on hydroxyapatite which resulted in an efficient enrichment of many low-abundance proteins [11].

Here we applied hydrophobic interaction chromatography to enrich low-copy-number proteins of *H. influenzae* and identified most of the proteins that bound to the column. In hydrophobic interaction chromatography, proteins are separated on the basis of differences in hydrophobicity. The proteins are adsorbed onto an uncharged matrix carrying hydrophobic groups, in the presence of salts. Elution is achieved by lowering the salt concentration [12,13]. Because of the various number of hydrophobic sites carried by proteins, hydrophobic interaction chromatography can efficiently fractionate complex protein mixtures.

2. Experimental

2.1. Materials

Immobilized pH gradient (IPG) strips were purchased from Pharmacia Biotechnology (Uppsala, Sweden). Reagents for the preparation of the one- (1D) and two-dimensional (2D) polyacrylamide gels were from Bio-Rad (Hercules, CA, USA) and Serva (Heidelberg, Germany). The column matrix, TSK Phenyl 5-PW (30 μm) was purchased from TosoHaas (Stuttgart, Germany).

2.2. Protein extraction

H. influenzae cells, strain Rd KW20, were grown in a 10-l fermentor containing 2% fetal calf serum and 10 mg/l hemin as well as the necessary nutrients and salts [2]. The fermentation yielded approximately 70 g of wet biomass which was kept frozen at −70°C until use. A 16-g amount of wet cell paste was suspended in 32 ml of 50 m*M* Tris–HCl, pH 8.0, containing 1 m*M* $MgCl_2$, 100 U/ml aprotinin, 5 m*M* ε-aminocaproic acid and 0.2 m*M* dithiotreitol. Nucleic acids were hydrolyzed by the addition of 250 U/ml benzonase (Merck, Darmstadt, Germany). Cells were disrupted in a French press (SLM Instruments, Urbana, IL, USA) at $1.33 \cdot 10^8$ Pa. EDTA·Na_2 (pH 8.0) was added to the lysate to 5 m*M* final concentration. The mixture was centrifuged at 3000 *g* for 20 min to sentiment intact cells and cell debris. The supernatant was centrifuged further at 18 000 *g* for 60 min to remove cell envelopes. The supernatant from this last step was recovered and centrifuged at 150 000 *g* for 90 min to sediment cell membranes. From the last centrifugation step, 5 ml of the supernatant, containing about 90 mg of soluble cytoplasmic proteins, were dialyzed against 50 m*M* sodium phosphate, pH 7.0, containing 1 *M* ammonium sulfate and 1 *M* glycine.

2.3. Hydrophobic interaction chromatography

The dialyzate was centrifuged at 30 000 *g* for 15 min, the soluble protein fraction was filtered through a 0.22 μm pore-size membrane (Millex-GV from Millipore, Volketswil, Switzerland) and applied onto a 28 ml (140×16 mm) TSK Phenyl 5-PW column [14,15], equilibrated with 50 m*M* sodium phosphate buffer, pH 7.0, containing 1 *M* ammonium sulfate and 1 *M* glycine. The column was washed with three column volumes of loading buffer and the bound proteins were eluted with 20 column volumes of a linear gradient of the same buffer without ammonium sulfate at 1 ml/min. Fractions of 5 ml were collected and pooled according to the elution profile. Twenty pools were formed and each was concentrated to about 1 ml by centrifugation at 2000 *g* in a Millipore Ultrafree-15 device with Biomax-10 membrane. The concentrates were analyzed by 1D and 2D polyacrylamide gel electrophoresis (PAGE).

2.4. Two-dimensional polyacrylamide gel electrophoresis

The 2D gel analysis was performed as previously described [16]. In short, comparable protein sample amounts were applied onto immobilized pH 3–10 nonlinear and 6–11 linear gradient strips and isoelec-

tric focusing was performed at 5000 V for 48 h. For the second dimension, the proteins were separated on 9–16% linear gradient polyacrylamide gels at 40 mA/gel. The gels were stained with colloidal Coomassie blue (Novex, San Diego, CA, USA) for total protein detection and were destained with water. The gels were scanned in a Molecular Dynamics Personal Densitometer and the images were processed using ADOBE PHOTOSHOP and POWERPOINT software.

2.5. Matrix-assisted laser desorption ionization mass spectrometry

The MALDI-MS analysis was performed as reported [17]. The protein spots were in-gel digested. The gel pieces were destained with 50% acetonitrile in 0.1 *M* ammonium bicarbonate and dried in a vacuum speed evaporator. The dried gel fragments were reswollen in 3 μl of 3 m*M* Tris–HCl, pH 8.0, containing 1 μg of endoproteinase Lys-C (Wako, Neuss, Germany) and incubated at 37°C for 12 h. A 3-μl volume of 30% acetonitrile, containing 0.1% trifluoroacetic acid, was added. After sonication for 3 min, 1 μl of the peptide extract was applied onto 0.5 μl of air-dried matrix. The matrix solution was prepared by dissolving 15 mg nitrocellulose (Bio-Rad) and 20 mg α-cyano-4-hydroxycinnamic acid (Sigma, St. Louis, MO, USA) in 1 ml of acetone–isopropanol (1:1, v/v). The samples were analyzed on a time-of-flight mass spectrometer (PerSeptive Biosystems, Cambridge, MA, USA) equipped with a reflectron. An accelerating voltage of 20 kV was used. Calibration was external to the samples.

3. Results and discussion

3.1. Protein enrichment

In a series of approaches, we investigated the capacity of several chromatographic steps to enrich low-abundance soluble proteins from microorganisms. Protein enrichment may be useful in the identification of novel drug targets and in the design of purification protocols to isolate proteins of interest. Recent improvements in the daily sample analysis throughput by MS and the information resulting from the sequence of the genomes of various microorganisms make possible the identification of practically all proteins present in fractions collected from a chromatographic step. This knowledge enables the optimization of the various stages of a chromatographic process and may be a useful guide to scientists aiming to isolate such proteins from the same or other organisms. Use of master protein enrichment and purification steps results in increased efficiency by avoiding labor- and time-consuming 'trial and error' approaches [9]. Here we investigated protein enrichment by hydrophobic interaction chromatography. This approach was chosen because it has a high binding capacity, it is convenient to perform and a large number of hydrophobic proteins exert important biological functions.

3.2. Hydrophobic interaction chromatography

The total soluble protein fraction of *H. influenzae* was used for fractionation on a TSK Phenyl column. After sample application, the column was washed with equilibration buffer. Few proteins were detected in the flow-through and wash fractions. Bound proteins were eluted with a linear gradient of low-salt buffer. The collected fractions were pooled according to the elution profile (Fig. 1A) and the pools were first analyzed by sodium dodecyl sulfate (SDS)-PAGE (Fig. 1B; the numbers correspond to the pools, indicated in Fig. 1A). Twenty pools were formed. The first pools contained only a small number of proteins (Fig. 1B, pools 1–4). The subsequent pools contained a larger number of proteins (Fig. 1B, pools 5–20). The pools of the fractions collected from the hydrophobic interaction column were also analyzed by 2D-PAGE. The 2D-gels of selected pools are shown in Figs. 2 and 3.

3.3. Protein enrichment by hydrophobic interaction chromatography

Four major proteins which did not bind to the column were recovered in the flow through fraction in a high purity: hypothetical protein HI0148, DNA-binding protein HU-alpha (HI0430), periplasmic arginine binding protein (HI1179) and Isg locus hypothetical protein (HI1693) (Fig. 1B, FT; a 2D gel image is not shown).

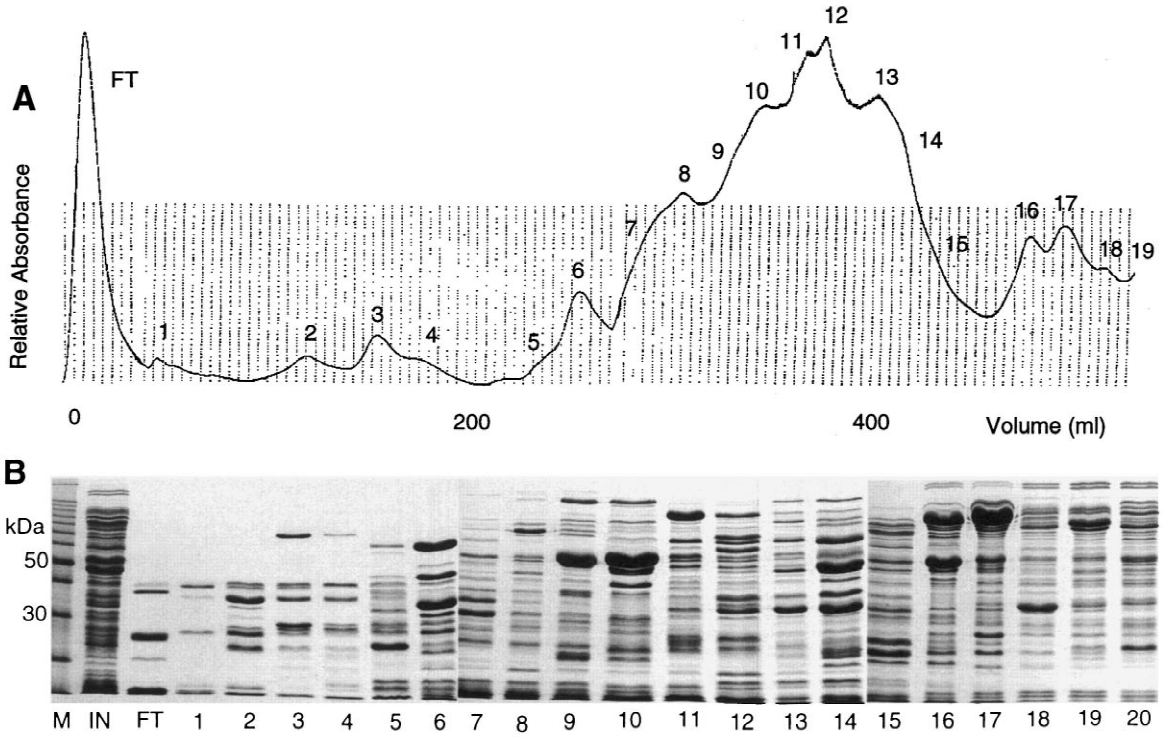

Fig. 1. Partial protein elution profile (A) and SDS-PAGE analysis (B) of the *H. influenzae* proteins eluted from the TSK Phenyl column. The soluble protein fraction was applied onto the column and eluted as stated under Section 2. (A) Parts of the elution profile which include protein peaks are shown. The numbers above the peaks indicate the pools formed by mixing the fractions included in the corresponding peaks (pool 20 is not shown). (B) Pools were analyzed by 15% SDS-gels under reducing conditions. The gels were stained with Coomassie blue. The numbers correspond to the numbers of the pools. M, Protein size markers of M_r 10 000 intervals, kDa=kilodalton; IN, input, material loaded onto the column; FT, flow through fraction.

Several bound proteins were efficiently enriched and are represented by strong bands or spots on the 1D- or 2D-gels, respectively. The major enriched and partially purified proteins were: in pool 2, D-galactose-binding periplasmic protein (HI0822), which was about 20% pure in pool 2 (Fig. 2A, the corresponding spot); in pool 5, inorganic pyrophosphatase (HI0124) (Fig. 1B, the strong band migrating at M_r about 25 000); in pool 6, phosphoenolpyruvate carboxykinase (HI0809) and glyceraldehyde-3-phosphate dehydrogenase (HI0001) (Fig. 2B, the corresponding spots); in pool 9, enolase (HI0932) which was approximately 20% pure (Fig. 2D); in pool 10, lipoamide dehydrogenase (HI1231) (not shown); in pool 11, methylenetetrahydropteroyltriglutamate methyltransferase (HI1702) (Fig. 3A); in pool 14, elongation factor Ts (HI0914) and hypothetical protein (HI1125) (Fig. 3C) and many others.

No clear separation of the peaks was obtained during elution (Fig. 1A), so that some proteins were present in more than two pools, such as phosphoenolpyruvate carboxykinase (HI0809), which was found in pools 6 to 8 (Fig. 2B and C). Other proteins were present in several pools such as the 50S ribosomal protein L10 (HI0640), which was present in all pools from 6 (Fig. 2B) to 14 (Fig. 3C). The basic proteins identified after separation on pH 6–11 IPG strips are indicated in Fig. 4. Many of them are also visible on the gels made from pH 3–10 IPG strips (Figs. 2 and 3). Several basic proteins were enriched by the hydrophobic interaction chromatography, mainly ribosomal proteins, L1 (HI0516), L4 (HI0778), L10 (HI0640), ribosome releasing factor

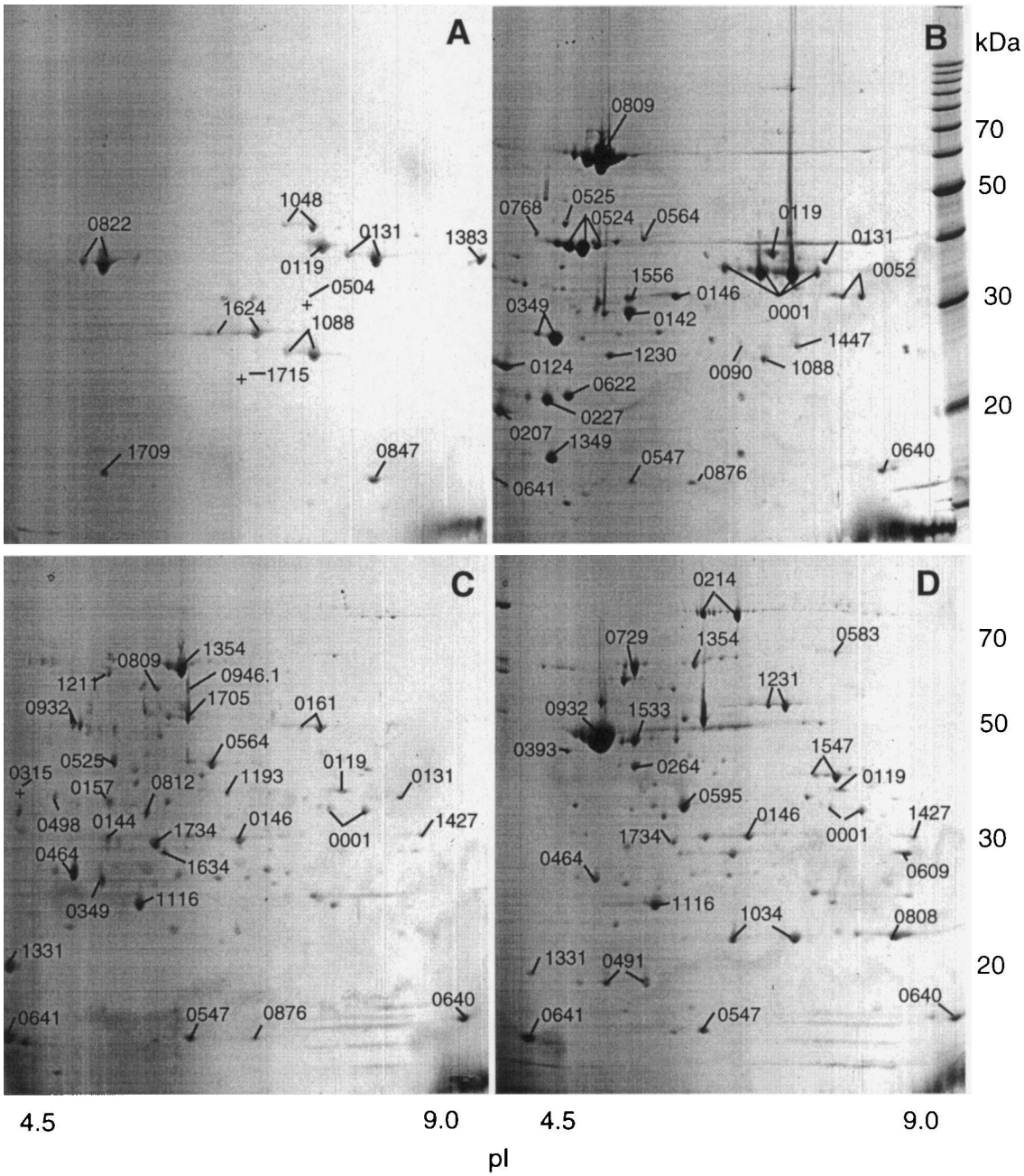

Fig. 2. Two-dimensional PAGE analysis of pools 2 (A), 6 (B), 8 (C) and 9 (D) of fractions collected from the TSK Phenyl column. The proteins were eluted with a descending ammonium sulfate gradient as described in Section 2. The samples were analyzed on 3–10 non-linear immobilized pH gradient strips, followed by 9–16% gradient SDS-gels. The gels were stained with colloidal Coomassie blue and destained with water. The destained gels were scanned in a molecular dynamics personal densitometer and the images were processed using ADOBE PHOTOSHOP and POWERPOINT software. The proteins were identified by MALDI-MS. The numbers next to the protein spots indicate the HI (*H. influenzae*) identification numbers [1]. The names of the corresponding proteins are given in Table 1. (B) At the right side of the gel the protein size markers are shown. Because not all 2D gels are shown, the approximate locations of certain protein spots, detected in other gels, are indicated.

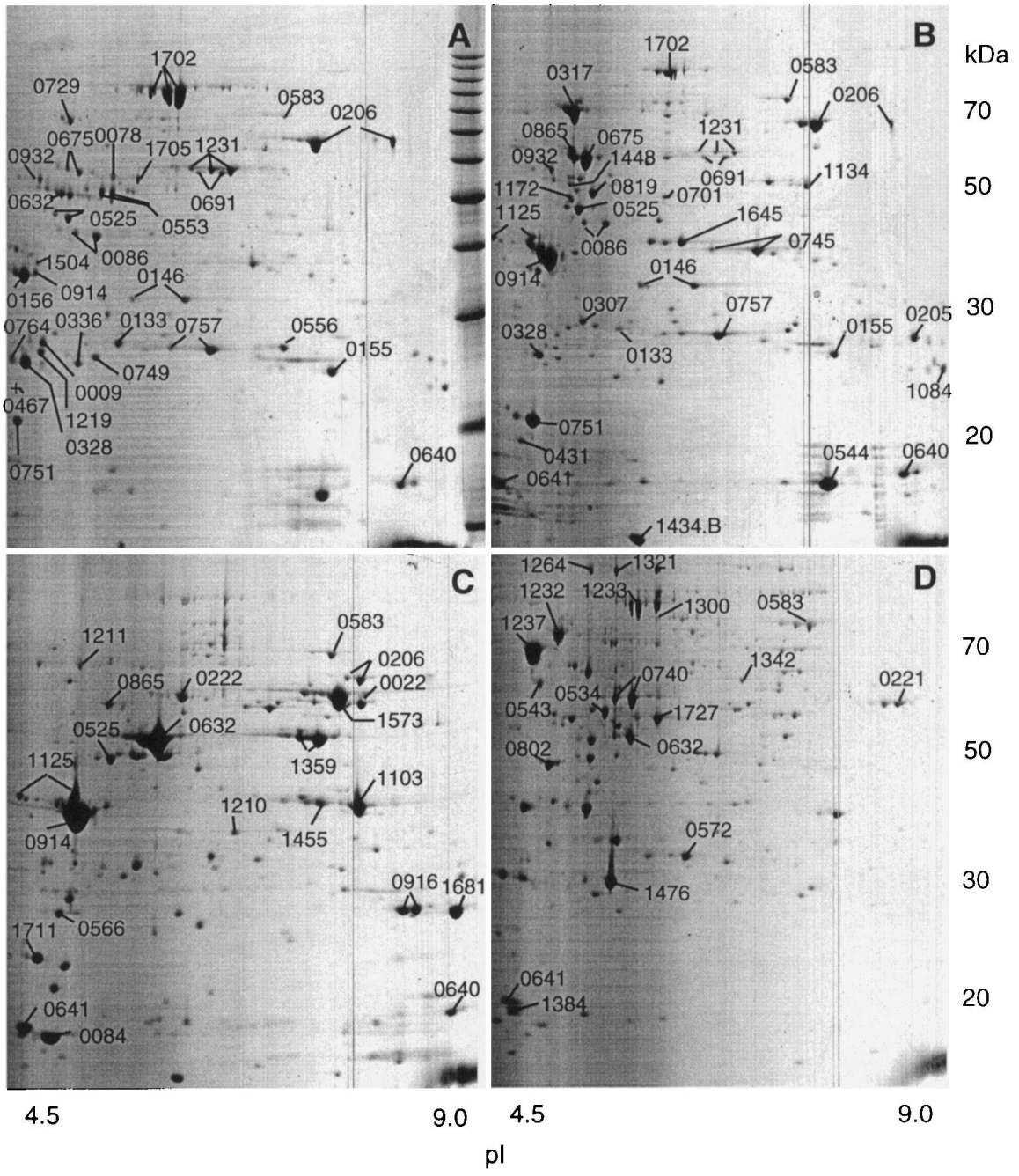

Fig. 3. Two-dimensional PAGE analysis of the pools 11 (A), 12 (B), 14 (C) and 20 (D) of fractions eluted from the TSK Phenyl column. The column was developed and the proteins were analyzed as stated in the legend to Fig. 2.

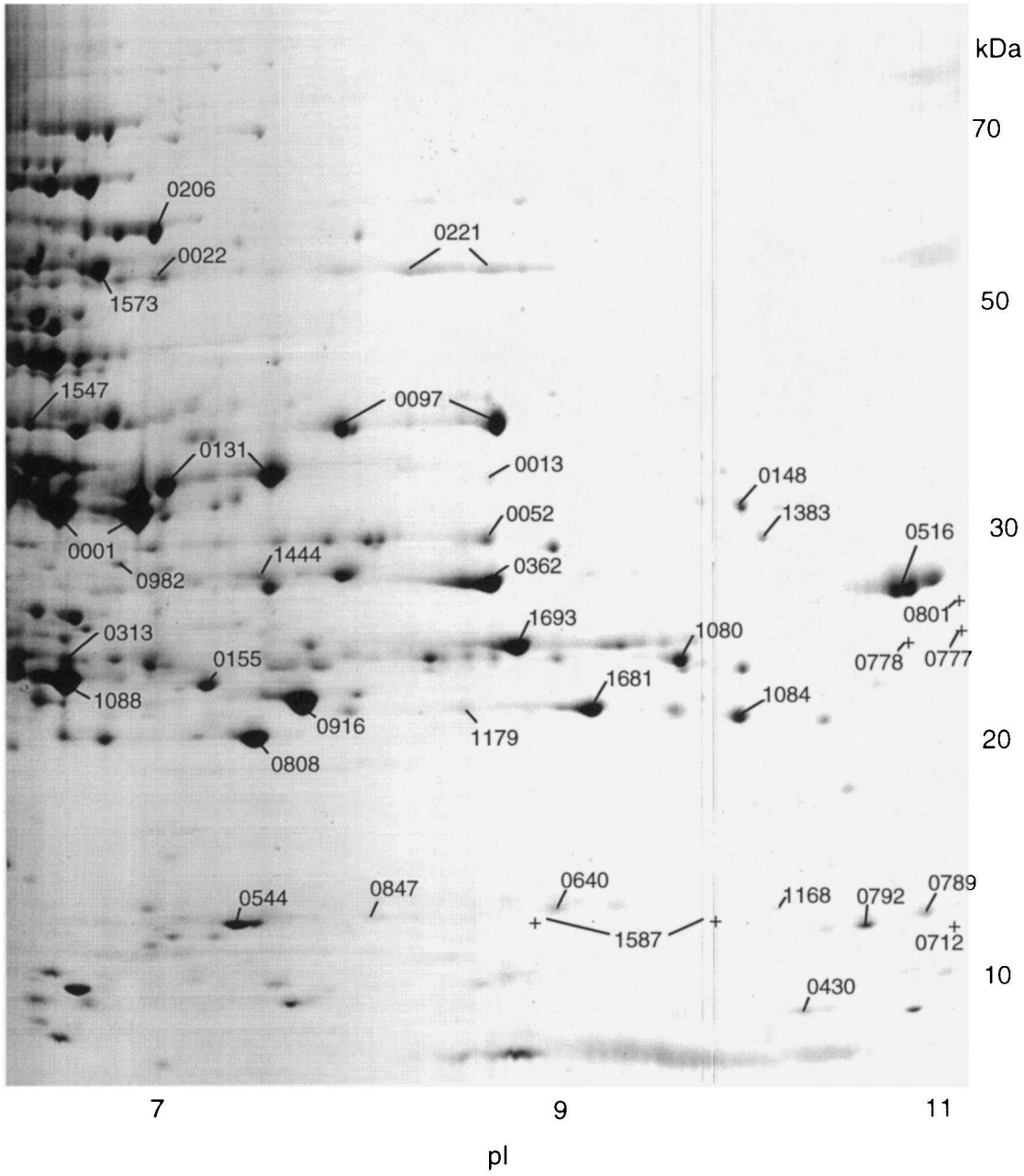

Fig. 4. Two-dimensional PAGE analysis of the basic proteins of *H. influenzae*. The proteins were analyzed on 6–11 IPG strips, followed by 9–16% SDS-gels. The spots representing proteins detected in pools from the TSK Phenyl column are characterized with their HI identification numbers [1]. The names are listed in Table 1. The spots representing the proteins found in the flow through fraction are shown as well (HI0148, HI0420, HI1179, HI1693).

(HI0808) and others (the particular 2D gel images of the various pools are not shown).

3.4. Protein analysis

The proteins bound to the TSK Phenyl column, present in the various pools 1–20, were identified from the 2D gels by MALDI-MS. The proteins spots in Figs. 2–4 are designated with the *H. influenzae* (HI) identification numbers [1] and their names are listed in Table 1. The superscript letter next to the name of certain proteins denotes that the particular protein has been enriched by the hydrophobic interaction chromatography step compared to its abundance in the starting material.

About 150 proteins were identified in the pools collected from the TSK Phenyl column (Table 1). A significant number of other proteins in the pools were represented by minor spots and were not identified. Table 1 includes 30 newly identified proteins. Some of them are not represented by visible spots on the 2D map comprising the total soluble proteins, therefore they could not have been identified if they were not enriched by hydrophobic interaction chromatography. Together with the proteins described here, the total number of identified proteins of *H. influenzae* increases to about 430 (Refs. [2–6,9,10] and this work). Comparison of the 2D gels of the pools collected from the column (partially shown in Figs. 2 and 3) with the gel of the soluble proteins of *H. influenzae*, i.e. the starting material loaded onto the TSK phenyl column, indicates that many of the enriched proteins were abundant proteins.

The proteins bound to TSK Phenyl column cannot be clustered into a single class. A large percentage of them are enzymes with various functions. Several ribosomal proteins are also included (Table 1). Table 1 includes many proteins that in the database are designated as hypothetical proteins or as predicted coding regions (unknown proteins). The detection of the hypothetical proteins in the proteome of *H. influenzae* provides the evidence that these proteins are really expressed, some of them in high copy number, as for example the proteins HI0131 and HI1681 (Fig. 4). Several proteins are represented by more than one spot, such as proteins HI0001, HI0097, HI0131 (Fig. 4) and many others. The reasons and the biological significance of the spot heterogeneity are not known at present. The heterogeneity may partially result from modifications during the performance of the 2D gel analysis, such as carbamylation, or from post-translational modifications, such as acylation and phosphorylation.

3.5. Comparison with other chromatographic approaches

Fractionation on the TSK Phenyl column resulted in an efficient enrichment of many abundant proteins. Nevertheless, chromatography on TSK Phenyl matrix did not result in protein purification to near homogeneity. Purification to near homogeneity was achieved for four proteins by applying chromatofocusing [6]. Except for a few exceptions, chromatofocusing did not enrich basic proteins, whereas both heparin chromatography and hydrophobic interaction chromatography resulted in efficient enrichment of large numbers of basic proteins, in particular ribosomal proteins [3,5,6,9].

Until now, three lists *of H. influenzae* proteins fractionated on heparin-Actigel [3,5,9], polybuffer exchanger [6] and TSK Phenyl (this work) have been prepared. A combination of two or more of these fractionation approaches should result in an efficient separation of complex protein mixtures into simpler fractions based on different binding principles. Each chromatographic approach adds a unique resolving power, which can be essential for the efficient analysis of the proteomes of higher organisms.

A comparison of the proteins included in the three lists, although not complete, shows that approximately 40% of the proteins enriched by hydrophobic interaction chromatography were also enriched by heparin chromatography and about the same percentage was enriched by chromatofocusing. Approximately 20% of the proteins listed in Table 1 were enriched by both heparin chromatography and chromatofocusing. We further observed that with the application of different chromatographic techniques for the enrichment of new proteins, we detected a constantly lower number of new proteins after each fractionation. This may mean that a large percentage of the rest of the proteins is not expressed in sufficient amounts under the cultivation conditions

Table 1. Proteins of *H. influenzae* bound to TSK Phenyl column

HI no.	Protein name	M_r		pI		MALDI-MS			Location (figure)
		Theor.	Observ.	Theor.	Observ.	Peptides		Sequence coverage	
						Matching	Total		
HI0001	Glyceraldehyde 3-phosphate dehydrogenase[a]	36 200	35 000	6.8	7	7	19	25	2–D, 4
HI0009	fdhE protein	34 658	30 000	4.7	4.7	3	14	14	3A
HI0013	GTP-binding protein	34 389	30 000	8.4	8.5	8	18	27	4
HI0022	Citrate lyase alpha-chain	54 556	50 000	6.9	7	3	17	10	3C, 4
HI0052	Hypothetical protein HI0052	37 592	33 000	9.6	8.8	3	10	9	2B, 4
HI0078	Cys-tRNA synthetase	52 779	55 000	5.5	5.5	3	17	6	3A
HI0084	Thioredoxin M	11 787	12 000	4.7	4.7	4	10	39	3C
HI0086	Cystathionine gamma synthase	41 954	43 000	5.2	5.2	5	15	3	3A, 3B
HI0090	Hypothetical protein HI0090	26 920	27 000	6.4	6.4	3	19	13	2B
HI0097	Major ferric iron binding protein[a]	32 137	32 000	9.3	8.9	5	20	33	4
HI0119	Adhesin B precursor	37 750	40 000	6.7	7	9	18	51	2A–D
HI0124	Inorganic pyrophosphatase[a]	19 827	25 000	4.5	4.5	4	20	36	2B
HI0131	*H. influenzae* predicted coding region HI0131	38 102	38 000	8.2	7.5	10	19	46	2A–C, 4
HI0133	Deoxycytidine triphosphate deaminase	21 660	28 000	5.4	5.4	3	18	15	3A, 3B
HI0142	N-Acetylneuraminate lyase	32 657	30 000	5.6	5.6	6	19	26	2B
HI0144	Glucose kinase	32 279	32 000	5	5	6	18	25	2C
HI0146	Hypothetical protein HI0146	36 489	32 000	6.8	6.2	6	19	20	2B–D, 3B
HI0148	Hypothetical protein HI0148[a]	40 823	35 000	9.8	9.8	13	19	52	4
HI0155	3-Ketoacyl-acyl carrier protein reductase	25 548	25 000	7.5	7.5	6	19	41	3A, 3B, 4
HI0156	Malonyl CoA-acyl carrier protein transacylase	33 631	35 000	4.8	4.8	4	18	26	3A
HI0157	β-Ketoacyl-acyl carrier protein synthase III	34 372	40 000	4.8	4.8	4	18	19	2C
HI0161	Glutathione reductase	49 697	51 000	6.5	7	8	19	26	2C
HI0205	*H. influenzae* predicted coding region HI0205	29 392	29 000	9.5	9.5	7	18	33	3B
HI0206	5′-Nucleotidase[a]	66 204	65 000	7.1	7	12	18	26	3A–C, 4
HI0207	Shikimic acid kinase I	20 279	21 000	4.6	4.6	8	19	67	2B
HI0214	Oligopeptidase A	78 367	80 000	5.5	6	2	9	12	2D
HI0221	Inosine-5′-monophosphate dehydrogenase	52 233	53 000	8.1	8.5	4	17	10	3D, 4
HI0222	GMP synthetase	58 547	60 000	5.7	5.7	3	18	5	3C
HI0227	Hypothetical protein HI0227	17 886	22 000	5	4.9	3	14	20	2B
HI0264	Heme-hemopexin-binding protein	41 071	45 000	5.1	5.2	5	19	23	2D
HI0307	Delta-1-pyrroline-5-carboxylate reductase	29 218	30 000	5.2	5.2	3	19	16	3B
HI0313	Holliday junction DNA helicase	22 589	22 000	6.5	6.5	3	15	17	4
HI0315	Hypothetical protein HI0315	26 595	35 000	4.4	4.5	4	18	24	2C
HI0317	Aspartyl-tRNA synthetase[a]	66 881	70 000	5.1	5.1	7	19	18	3B
HI0328	Elongation factor P	20 713	24 000	4.7	4.7	2	9	23	3A–B
HI0336	Hypothetical protein HI0336	21 791	24 000	5.2	5.2	4	13	26	3A
HI0349	Adenylate kinase[a]	23 550	28 000	5.1	5.1	7	20	44	2B–C
HI0362	Adhesin B precursor[a]	32 443	29 000	9.6	9	11	20	58	4
HI0393	Hypothetical protein HI0393	40 011	44 000	4.6	4.6	5	18	20	2D
HI0430	DNA-binding protein HU-alpha (HU-2)[a]	15 147	9000	11.1	11	4	11	42	4
HI0431	Hypothetical protein HI0431	22 934	21 000	4.6	4.6	4	9	17	3B
HI0464	Hypothetical protein HI0464	23 193	24 000	4.9	4.9	3	19	13	2C, 2D
HI0467	Hypothetical protein HI0467	33 306	23 000	4.8	4.8	3	14	8	3A
HI0491	*H. influenzae* predicted coding region HI0491	18 686	18 000	5.3	5.3	4	18	29	2D
HI0498	Spermidine/putrescine-binding periplasmic protein	35 236	37 000	4.8	4.8	5	18	16	2C
HI0504	Periplasmic ribose-binding protein	30 296	32 000	8.9	7	7	19	41	2A
HI0516	Ribosomal protein L1[a]	24 091	28 000	10.5	10.5	10	18	50	4
HI0524	Fructose-bisphosphate aldolase[a]	39 485	40 000	5.2	5.2	11	20	49	2B
HI0525	Phosphoglycerate kinase	41 071	42 000	5.1	5.1	7	18	30	2B–C, 3A–C
HI0534	Aspartase	52 020	51 000	5.2	5.2	10	20	34	3D
HI0543	Heat shock protein GroEL	57 654	57 000	4.7	4.7	5	17	12	3D
HI0544	Ribosomal protein L9	15 626	13 000	7	7.1	3	17	35	4
HI0547	Ribosomal protein S6	14 484	13 000	5.5	6	3	15	30	2B–D
HI0553	6-Phosphogluconate dehydrogenase	53 449	53 000	5.2	5.2	8	18	21	3A

Table 1. Continued

HI no.	Protein name	M_r		pI		MALDI-MS			Location (figure)
		Theor.	Observ.	Theor.	Observ.	Peptides		Sequence coverage	
						Matching	Total		
HI0556	Putative glucose 6-phosphate dehydrogenase isozyme	18 326	28 000	5.7	6.5	4	15	39	3A
HI0564	Asparagine synthetase A	37 522	41 000	5.7	5.7	6	18	22	2B–C
HI0566	Dod protein	25 709	28 000	5.6	5	3	18	18	3C
HI0572	Hypothetical protein HI0572	26 896	30 000	5.7	6	4	14	21	3D
HI0583	2′,3′-Cyclic-nucleotide 2′-phosphodiesterase[a]	72 889	72 000	7.2	7.2	14	18	28	2D, 3A–D
HI0595	Carbamate kinase	33 590	33 000	5.3	5.5	3	18	16	2D
HI0609	5,10-Methylenetetrahydrofolate dehydrogenase	30 612	28 000	7.6	8	5	18	28	2D
HI0622	Polypeptide deformylase	19 216	21 000	5.4	5.2	6	13	28	2B
HI0632	Elongation factor Tu	43 441	47 000	5.2	5.3	8	22	18	3C
HI0640	Ribosomal protein L10	17 809	15 000	9.2	9.2	4	19	39	2B–D, 3A–C, 4
HI0641	Ribosomal protein L7/L12	12 455	13 000	4.5	4.5	5	18	59	2B–D, 3B–D
HI0675	Peptidase D	53 026	58 000	5.2	5	8	17	33	3A, 3B
HI0691	Glycerol kinase	56 334	58 000	6.2	6.2	5	17	14	3A, 3B
HI0701	*H. influenzae* predicted coding region HI0701	38 694	46 000	5.9	5.9	3	19	10	3B
HI0712	Transferrin-binding protein 1(tbp1)	124 392	12 000	9.6	10.3	5	18	4	4
HI0729	Prolyl-tRNA synthetase	64 105	64 000	5.1	5.2	6	18	17	2D
HI0740	Hypothetical protein HI0740	52 763	53 000	5.5	5.5	9	15	27	3D
HI0745	L-Asparaginase II	37 360	39 000	7	7	8	19	35	3B
HI0749	LexA repressor	23 385	25 000	5.5	5.5	4	9	25	3A
HI0751	ToxR regulon	17 786	21 000	4.8	4.8	2	5	29	3A, 3B
HI0757	Phosphoglyceromutase	26 089	27 000	6.1	6.1	6	18	34	3A, 3B
HI0764	3,4-Dihydroxy-2-butanone 4-phosphate synthetase	23 467	26 000	4.7	4.7	3	14	26	3A
HI0777	Ribosomal protein L3	22 349	23 000	10.7	10.7	4	18	20	4
HI0778	Ribosomal protein L4	21 940	22 000	10.4	10.5	5	18	36	4
HI0789	Ribosomal protein L24[a]	11 285	13 000	10.5	10.6	2	7	25	4
HI0792	Ribosomal protein S8	14 034	12 000	10.5	10.5	5	17	51	4
HI0801	Ribosomal protein S4	23 566	25 000	10.9	10.9	4	18	20	4
HI0802	DNA-directed RNA polymerase alpha chain	36 618	41 000	4.8	4.8	3	17	13	3D
HI0808	Ribosome releasing factor[a]	20 716	20 000	7.7	7.7	3	9	26	2D, 4
HI0809	Phosphoenolpyruvate carboxykinase	59 595	60 000	5.3	5.2	15	20	36	2B, 2C
HI0812	Glucosephosphate uridyltransferase	32 543	31 000	5.2	5.4	4	18	14	2C
HI0819	Galactokinase	44 269	48 000	5.4	5.2	5	19	17	3B
HI0822	D-Galactose-binding periplasmic protein[a]	37 579	40 000	6.3	5	10	19	34	2A
HI0847	Hypothetical protein HI0847	12 854	13 000	7.3	8.2	6	19	79	2A, 4
HI0865	Glutamine synthetase	52 595	55 000	5.1	5.2	5	19	16	3C
HI0876	Nucleoside diphosphate kinase	16 096	16 000	5.9	6	4	15	40	2C
HI0914	Elongation factor EF-Ts[a]	30 282	38 000	5	4.9	3	18	17	3A–C
HI0916	Export factor homolog	21 704	28 000	9.4	8.2	8	18	51	3C, 4
HI0932	Enolase[a]	46 255	50 000	4.9	4.9	4	18	17	2D, 3A, 3B
HI0946.1	L-2,4-Diaminobutyrate decarboxylase	56 887	60 000	5.4	5.4	7	18	22	2C
HI0982	6-Phosphofructokinase	35 242	30 000	6.8	7	9	18	33	4
HI1034	*H. influenzae* predicted coding region HI1034	18 596	20 000	6.2	6.2	4	18	36	2D
HI1048	*H. influenzae* predicted coding region HI1048	41 347	45 000	7.2	7.2	7	20	26	2A
HI1080	Glutamine-binding periplasmic protein	28 319	25 000	9.8	9.8	8	20	46	4
HI1084	Hypothetical protein HI1084	24 494	22 000	10.3	10.3	6	18	30	4
HI1088	Superoxide dismutase	24 209	25 000	6.7	6.7	5	20	36	2A, 2B, 4
HI1103	Cysteine synthase	33 384	36 000	7	8	8	19	26	3C
HI1116	Deoxyribose aldolase	23 998	23 000	5.2	5.2	8	18	42	2C, 2D
HI1125	Hypothetical protein HI1125	35 054	40 000	4.9	4.9	8	17	39	3B, 3C
HI1134	UDP-murnac-pentapeptide synthetase	50 270	50 000	6.3	7	4	18	12	3B
HI1168	*H. influenzae* predicted coding region HI1168	13 858	14 000	10.2	10.2	3	18	26	4
HI1172	S-Adenosylmethionine synthetase 2	42 232	48 000	5.1	5.1	3	19	12	3B
HI1179	Periplasmic arginine-binding protein	26 119	20 000	9.1	8.9	8	20	47	4
HI1193	Branched-chain-amino-acid transaminase	38 218	40 000	5.9	6	6	19	26	2C

Table 1. Continued

HI no.	Protein name	M_r		p*I*		MALDI-MS			Location (figure)
		Theor.	Observ.	Theor.	Observ.	Peptides		Sequence coverage	
						Matching	Total		
HI1210	Malate dehydrogenase	32 693	33 000	6	6	5	19	33	3C
HI1211	Lysyl-tRNA synthetase	56 900	65 000	5	5	6	18	13	2C
HI1219	Cytidylate kinase	25 597	28 000	5.2	5	6	12	37	3A
HI1230	Adenine phosphoribosyltransferase	19 754	24 000	5.4	5.4	5	15	36	2B
HI1231	Lipoamide dehydrogenase[a]	51 520	55 000	5.9	6	6	18	17	2D, 3A, 3B
HI1232	Dihydrolipoamide acetyltransferase	59 431	62 000	5	4.8	3	19	8	3D
HI1233	Pyruvate dehydrogenase	99 239	80 000	5.5	5.5	6	16	8	3D
HI1237	Heat shock protein 70 000	68 296	68 000	4.6	4.6	8	19	25	3D
HI1264	DNA gyrase	97 928	100 000	5	5	5	16	6	3D
HI1300	Hypothetical protein HI1300	73 564	76 000	5.1	5.5	4	14	4	3D
HI1321	Exodeoxyribonuclease V	140 341	110 000	5.6	5.6	3	17	2	3D
HI1331	Transcription elongation factor	17 523	19 000	4.5	4.5	5	20	67	2C, 2D
HI1342	Hypothetical protein HI1342	52 605	58 000	5.3	6.5	3	17	9	3D
HI1349	Hypothetical protein HI1349	18 327	17 000	5.1	5.1	6	18	48	2B
HI1354	Glutaminyl-tRNA synthetase	64 551	64 000	5.5	5.5	5	18	13	2C, 2D
HI1359	ADP-glucose synthetase[a]	37 167	45 000	5.8	7	8	18	33	3C
HI1383	Periplasmic phosphate-binding protein	28 523	28 000	10	10	5	16	35	2A, 4
HI1384	Ferritin like protein	21 181	18 000	4.7	4.7	4	15	32	3D
HI1427	*H. influenzae* predicted coding region HI1427	30 931	30 000	8.6	8.9	6	17	30	2C, 2D
HI1444	5,10-Methylenetetrahydrofolate reductase	33 113	26 000	7.7	7.7	3	17	13	4
HI1448	Molybdopterin biosynthesis protein	44 567	50 000	5	5	3	6	9	3B
HI1455	Fimbrial transcription regulation repressor	40 180	40 000	6.9	6.9	2	3	5	3C
HI1476	Hypothetical protein HI1476	26 668	30 000	4.5	5.2				3D
HI1504	I protein	39 002	41 000	4.8	4.8	2	6	6	3A
HI1533	β-Ketoacyl-ACP synthase I[a]	43 021	45 000	5.1	5.2	5	19	10	2D
HI1547	DAHP synthetase	40 310	35 000	6.6	6.8	5	18	19	2D, 4
HI1556	Vancomycin resistance protein	34 842	34 000	6.2	5.8	6	19	33	2B
HI1573	Pyruvate kinase type II[a]	51 250	54 000	6.7	6.8	8	17	21	3C, 4
HI1587	DNA-binding protein H-NS	15 319	13 000	9.7	9.7	3	17	26	4
HI1634	Tetrahydrodipicolinate N-succinyltransferase	33 138	33 000	6.3	5.8	6	12	38	2C
HI1645	Fructose-1,6 bis phosphatase	37 077	40 000	5.5	5.5	6	21	20	3B
HI1681	*H. influenzae* predicted coding region HI1681	29 392	28 000	9.6	8.9	3	18	28	3C, 4
HI1693	Isg locus hypothetical protein	27 285	25 000	9.8	9	11	19	64	4
HI1702	Tetrahydropteroyltriglutamate methyltransferase[a]	85 470	85 000	5.7	5.8	10	19	21	3A, 3B
HI1705	Aminopeptidase A/I	53 951	55 000	5.5	5.6	8	17	21	2C, 3A
HI1709	*H. influenzae* predicted coding region HI1709	13 222	13 000	4.9	4.9	3	20	23	2A
HI1711	Glucose phosphotransferase enzyme III-glc	17 956	22 000	4.7	4.7	3	14	27	3C
HI1715	Hypothetical protein HI1715	21 300	22 000	6	6	4	19	36	2A
HI1727	Argininosuccinate synthetase	49 372	49 000	5.5	6	4	17	16	3D
HI1734	Short chain alcohol dehydrogenase homolog[a]	32 145	30 000	5.9	5.9	5	19	27	2C, 2D

[a] Indicates that the protein was enriched compared to its abundance in the starting material.

The column was eluted with 50 m*M* sodium phosphate, pH 7.0, containing 1 *M* glycine. The fractions were pooled according to the elution profile (Fig. 1A), concentrated and analyzed by 2D PAGE (Figs. 2–4). The proteins from the 2D gels were identified by MALDI-MS. The theoretical and the approximately observed M_r and p*I* values of the proteins are indicated. The table includes the numbers of matching peptides found (matching) and of total possible peptides (total), that can be generated upon protein digestion with endoproteinase Lys-C. The column 'Sequence coverage' gives the percentage of the protein amino acid sequence covered by the sequence of the matching peptides. The sequence coverage is an indication of confidence of protein identification. The selected protein usually showed the highest sequence coverage. In the column 'Location', the figure where the corresponding protein spot can be found is indicated. The table represents a partial list of the bound proteins and includes proteins that are major components of the pools. Proteins, which were not found in the column fractions, although may be known, are not listed in this table.

HI, *H. influenzae* identification number, according to Fleischmann et al. [1].

applied and their enrichment could not be achieved by chromatographic means.

4. Conclusions

We fractionated the total soluble proteins of *H. influenzae* by hydrophobic interaction chromatography. The protein pools were analyzed by 2D gels and the proteins identified by MALDI-MS. Thirty new proteins were identified partially as low-abundance gene products. The results may be useful in the search for drug targets and in the design of master purification protocols for isolation of proteins of interest from *H. influenzae* and related microorganisms.

The search for new pharmaceuticals may require the isolation of low-abundance proteins or of proteins of unknown function. These proteins are referred to as hypothetical proteins or predicted coding regions and some of them might be potential drug targets. For the isolation of these proteins no literature information is available. Similar cases will certainly appear at an accelerated rate, since an increasing contribution of genomics to drug discovery is expected in the near future. A large number of hypothetical or unknown proteins are included in the lists of the three master protein enrichment steps, heparin chromatography, chromatofocusing and hydrophobic interaction chromatography, and this information may be useful in drug target search. Moreover, it represents a step toward protein purification automation. The analysis of the proteomes of higher organisms is anticipated to be more complex in comparison with the bacterial proteomes and will probably require a combination of many analytical approaches, comprising multiple chromatographic steps, 2D PAGE analysis and mass spectrometry.

Acknowledgements

We thank J.-F. Juranville for technical assistance and Drs. P. Berndt, C. Gray and H. Langen for helpful suggestions.

References

[1] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.-F. Tomb, B.A. Dougherty, J.M. Merrick, K. Kenney, G. Sutton, W. Fitz-Hugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.I. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, J.C. Venter, Science, 269 (1995) 496.
[2] H. Langen, C. Gray, D. Röder, J.-F. Juranville, B. Takács, M. Fountoulakis, Electrophoresis 18 (1997) 1184.
[3] M. Fountoulakis, H. Langen, S. Evers, C. Gray, B. Takács, Electrophoresis 18 (1997) 1193.
[4] M. Fountoulakis, J.-F. Juranville, P. Berndt, Electrophoresis 18 (1997) 2968.
[5] M. Fountoulakis, B. Takács, H. Langen, Electrophoresis 19 (1998) 761.
[6] M. Fountoulakis, H. Langen, C. Gray, B. Takács, J. Chromatogr. A 806 (1998) 279.
[7] A.J. Link, L.G. Hays, E.B. Carmack, J.R. Yates, Electrophoresis 18 (1997) 1314.
[8] P. Cash, E. Argo, P.R. Langford, J.S. Kroll, Electrophoresis 18 (1997) 1472.
[9] M. Fountoulakis, B. Takács, Protein Expression Purif. 14 (1998) 113.
[10] M. Fountoulakis, J.-F. Juranville, D. Röder, S. Evers, P. Berndt, H. Langen, Electrophoresis 19 (1998) 1819.
[11] M. Fountoulakis, M.-F. Takács, B. Takács, P. Berndt, H. Langen, in preparation.
[12] S. Påhlman, J. Rosengren, S. Hjertén, J. Chromatogr. 131 (1974) 99.
[13] N.C. Robinson, D. Wiginton, L. Talbert, Biochemistry 23 (1984) 6121.
[14] Y. Kato, T. Kitamura, T. Hashimoto, J. Chromatogr. 333 (1985) 202.
[15] Z. El Rassi, A.L. Lee, C. Horvath, in: J.A. Asenjo (Editor), Separation Process in Biotechnology, Marcel Dekker, New York, 1985, pp. 447–494.
[16] H. Langen, D. Röder, J.-F. Juranville, M. Fountoulakis, Electrophoresis 18 (1997) 2085.
[17] M. Fountoulakis, H. Langen, Anal. Biochem. 250 (1997) 153.